

Computational Social Sciences

Francesca D'Errico
Isabella Poggi
Alessandro Vinciarelli
Laura Vincze *Editors*

Conflict and Multimodal Communication

Social Research and Machine
Intelligence

 Springer

Computational Social Sciences

A series of authored and edited monographs that utilize quantitative and computational methods to model, analyze, and interpret large-scale social phenomena. Titles within the series contain methods and practices that test and develop theories of complex social processes through bottom-up modeling of social interactions. Of particular interest is the study of the co-evolution of modern communication technology and social behavior and norms, in connection with emerging issues such as trust, risk, security, and privacy in novel socio-technical environments.

Computational Social Sciences is explicitly transdisciplinary: quantitative methods from fields such as dynamical systems, artificial intelligence, network theory, agent-based modeling, and statistical mechanics are invoked and combined with state-of-the-art mining and analysis of large data sets to help us understand social agents, their interactions on and offline, and the effect of these interactions at the macro level. Topics include, but are not limited to social networks and media, dynamics of opinions, cultures and conflicts, socio-technical co-evolution, and social psychology. Computational Social Sciences will also publish monographs and selected edited contributions from specialized conferences and workshops specifically aimed at communicating new findings to a large transdisciplinary audience. A fundamental goal of the series is to provide a single forum within which commonalities and differences in the workings of this field may be discerned, hence leading to deeper insight and understanding.

Series Editors

Elisa Bertino
Purdue University, West Lafayette,
IN, USA

Jacob Foster
University of California, Los Angeles,
CA, USA

Nigel Gilbert
University of Surrey, Guildford, UK

Jennifer Golbeck
University of Maryland, College Park,
MD, USA

James A. Kitts
University of Massachusetts, Amherst,
MA, USA

Larry Liebovitch
Queens College, City University of
New York, Flushing, NY, USA

Sorin A. Matei
Purdue University, West Lafayette,
IN, USA

Anton Nijholt
University of Twente, Enschede,
The Netherlands

Robert Savit
University of Michigan, Ann Arbor,
MI, USA

Alessandro Vinciarelli
University of Glasgow, Scotland

More information about this series at <http://www.springer.com/series/11784>

Francesca D'Errico • Isabella Poggi
Alessandro Vinciarelli • Laura Vincze
Editors

Conflict and Multimodal Communication

Social Research and Machine Intelligence

 Springer

Editors

Francesca D'Errico
Uninettuno University
Roma, Italy

Isabella Poggi
Università Roma Tre
Roma, Italy

Alessandro Vinciarelli
Department of Computing Science
University of Glasgow
Glasgow, UK

Laura Vincze
Department of Education Sciences
Università di Macerata
Roma, Italy

Computational Social Sciences

ISBN 978-3-319-14080-3

ISBN 978-3-319-14081-0 (eBook)

DOI 10.1007/978-3-319-14081-0

Library of Congress Control Number: 2015930048

Springer Cham Heidelberg New York Dordrecht London

© Springer International Publishing Switzerland 2015

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made.

Printed on acid-free paper

Springer International Publishing AG Switzerland is part of Springer Science+Business Media (www.springer.com)

Preface

“Peace is a gift.” This is what Pope Francis said on June 6, 2014, talking about the Israeli-Palestinian conflict during an interview with journalists. It is certainly a way to give due weight to the state of well-being and harmony between people that belong to groups with different and often incompatible interests. However, despite being a gift, peace is a costly process made up of steps that hardly ever respond to the laws of gratuity.

Very often peace goes through delicate and complex passages within the management of a conflict, depending on both its phases (i.e., conflict settlement, conflict resolution and reconciliation; Kelman 2006) and its characteristics (intragroup, intergroup, ethnic, Tajfel 1986; Constructive and deconstructive, Deutsch et al. 2006; High vs. Low stakes conflict, symmetrical and asymmetrical; Giebels 2012), which suggest what presumably might be possible solutions and attempts at reconciliation (Nadler et al. 2008).

Within the first phase of the conflict (conflict settlement), recent research in social psychology has focused on the processes of understanding and constructive management (Deutsch et al. 2006; Pruitt et al. 2003), mainly determined by socioemotional aspects of the relationship between former enemies (Nadler 2008): important examples are some real experiences like the Truth and Reconciliation Committee (TRC) in South Africa, where perpetrators and victims face a risky “apology-forgiveness” cycle through which the former restores his sense of guilt and the latter his sense of agency and empowerment.

Within this framework other scholars pointed out that forgiveness can be facilitated if both groups develop a “common victim identity,” i.e., if they become aware that their very fighting is determined by third parties: this decreases the so-called competitive victimhood, the tendency to reclaim the status of being “more victim” than each other, thus making forgiveness easier (Schnabel et al. 2013; Noor et al. 2012).

Another important contribution on the understanding of how to overcome conflict is the communicative analysis carried out by Bar-Tal (2010): the “narratives” of war and peace within both former enemies—including also media activists—can

promote “healthy strategies” to go beyond contrapositions, starting from ordinary people (Leone et al. 2007).

Some authors have instead focused on the tools (i.e., interactive problem solving; Kelman 2002; instrumental route; Nadler 2008) that after conflict resolution can improve not only the objective conditions but also the relationships between the former enemies. Kelman (2002) indicates what are the best conditions for informal mediation among israelo-palestinian people who are politically relevant or active in their communities, in order to address the parties’ basic needs through interactive problem solving, and to gain an higher degree of trust and new responsiveness to others’ needs.

Nadler in the case of intergroup conflict, highlighting the sociopsychological aspects to take care of when former enemies start working on cooperative projects, follows the conditions of realistic conflicts of Sherif (1961): *common goals groups*, an *enduring activity*, *equal status between groups*, and *institutional support*. These practices are used mainly for intergroup conflicts, and the experimental studies carried out within this framework are mainly based on a single behavioral choice (i.e., to forgive or not, help, cooperate or not, exchange opposite opinions or point of views), but in some recent studies on negotiation processes the idea is emerging of observing the conflict dynamics (Giebels 2012) by means of verbal and nonverbal signals detection and taking care also of the situated contexts and of a central variable in cultural and social psychology: the difference between high and low context cultures.

In a longer-term perspective, Kelman (2010) shows how persuasion processes can even bring about a reconstruction of the former enemy’s identity, producing an “ethos of peace and reconciliation” (Bar-Tal 2010) characterized by shared narratives of the conflict, mutual agreement and respect for others or the other groups, and by patterns of cooperative interactions (Bar Simon Tov 2004).

The premise from which the studies of Peace psychology (Christie et al. 2001) move is the fact that peace is a state to be achieved, in the very same way as at the individual level it is the welfare attainable through nonviolent solution of conflict.

This state of harmony can be reached by overcoming conflict. If conflict is defined (Lewin 1935) as any situation in which, at the same time, forces oriented in opposite directions but of comparable intensity act, it may constitute not only a negative state but also a useful and sometimes necessary way to change and restore the balance of dominant and unbalanced forces or powers. In this sense also Moscovici (1976), when dealing with minorities and describing their antagonist and alternative position against a majority, explains this situation as a conflict that can lead to innovation and social change. The idea of conflict as a process of change and progress can be traced back to studies by Mugny and Doise (1978) who discover it at the intra-individual level in the “ideal” socio-cognitive conflict in children; but it is also in the words of the trade unionist Pierre Carniti “We should reevaluate conflict, since without conflicts there is no social justice,” quoted by Castelfranchi (Chap. 1), who finally claims “conflicts are the engine of change and possibly of progress.”

With this in mind, in the present work we tried to go beyond the dark side of conflict, not considering it simply as an obstacle to be overcome, but firstly as an object to be dissected in its philosophical, linguistic, and psychological aspects, both to promote conflict management as a means for change and social justice and to develop tools for detecting its signals more readily, in order to the prevention of destructive and intractable conflicts.

One of the lenses by which we observe and study different types of conflicts will be multimodal communication, as mentioned in the title of the book.

Communication in conflicts very often includes, besides verbal insults, harsh criticism or bad words, also threatening facial expressions, angry gaze, proud postures, defiant head poses, loud voice, and interruptions and overlapping conversational turns; in a word, a whole game of signals of dominance and persistency that in negotiation or reconciliation phases may turn into signals of appeasement, politeness, and acceptance. Taking all this in due consideration requires keeping track of research on multimodality, which has been flourishing in the last two decades, and moving new steps in it.

Since the end of the 1990s, research in psychology, pragmatics, semiotics, and artificial intelligence emphasized the importance of multimodal communication. Starting with Ekman and Friesen (1978) and their complex Facial Action Coding System (FACS), the range of body behaviors under analysis significantly widened and attention was dedicated to movements of the face and head. Facial expression was studied mainly in relation to emotion communication (Scherer 1981; Scherer and Ekman 1984; Scherer and Grandjean 2008), but also concerning their power of emotion moderation and trust in conflict and negotiation (Diamantini and Pietroni 2002). Head nods, shakes, and other head movements were analyzed as signals of agreement, disagreement, dominance, and submission (Rienks et al. 2010; Poggi et al. 2010; Paggio and Navarretta 2013; Rahayudi et al. 2014). Signals of covert conflict were also investigated, like the facial and bodily expressions of acidity (D'Errico and Poggi 2014). With regard to another cornerstone of multimodal communication, gestures, besides classic contributions on autonomous and coverbal gestures (Morris 1977; McNeill 1992; Kendon 2004), persuasive gestures in political communication were investigated (Streeck 2007; Poggi and Vincze 2009), but also attentive and original semiotic analyses of space and gestures were carried out by Waisman (2010) in the conflict situations between Arabs and Israelis. Other areas of research on multimodal communication connected to conflict were the analysis of political and judicial debates, where turn overlapping and interruptions are a cue to conflictive talk (Pesarin et al. 2010; Navarretta 2013) and facial and gestural behavior may be a weapon of discredit and derision against the opponent (Poggi et al. 2012; D'Errico and Poggi 2013).

From an applicative point of view, one of the main aims of gesture and face analysis is the implementation of insights coming from the study of these typically human behaviors into Embodied Conversational Agents (ECAs) (de Rosi et al. 2003), i.e., virtual characters similar to a human being and able to engage in face-to-face interaction with the user (Cassell et al. 2000; Niewiadomski et al. 2008; Solano Méndez and Reidsma 2011). Research in the field of social sciences started

to be more and more intertwined with research in computer sciences and with the beginning of the twenty-first century, we witness an increase in the collaborative projects between researchers coming from these two fields (e.g., HUMAINE 2004–2007; SEMAINE, SSPNet 2009–2013) where members from both communities collaborate to develop instruments aimed at the analysis and synthesis of multimodal communication.

The interdisciplinary aim of the book with the encounter of cognitive science, social psychology, linguistics, ethology, and philosophy becomes a necessity when you want to apply knowledge gained in the field of computer science: conflictual communication—verbal and nonverbal—in this area is studied in order to promote the use of intelligent machines that automatically measure and understand the escalation, promote conflict management, and, therefore, support the negotiation process.

The innovative collaboration between humanities and computer science therefore achieves two goals: on one side, a goal of *knowledge* that exploits machines as a test of research hypotheses, through the application of theories or psychosocial models by means of intelligent technologies or by simulations; but it also enables an “aware” application since social research can drive applicative studies within computer science taking into account individual and contextual variables within a well-defined theoretical framework.

This dual aim becomes possible, first of all, through a “methodological” dialog made possible by means of research on multimodal communication as a place of encounter between seemingly distant disciplines.

In particular within the theme of conflict, this interdisciplinary research demonstrates that it is possible to extract the individual social cues that can predict the conflicting context by means of automatic or semiautomatic turn organization analysis (*who talks, to whom, how much, the dynamics of exchange between speakers*; Pesarin et al. 2012) or by techniques of speech recognition, detection of visible activity like head pose, face or hand gestures, and signal processing of physiological data like heart rate or electrodermal response (Narayanan 2013).

This is witnessed by the workshop “*Conflict and Communication. Multimodal Social Signals of Conflict and Negotiation in Humans, Animals, and Machines*,” held in Rome, October 29–31, 2013 organized by the European Network SSPNet—Social Signal Processing Network (see <http://www.klewel.com/conferences/sspnet-roma-2013/>).

Some of these applications, grouped in the last part of the book, “Technologies for Conflict Detection and Simulation,” represent an attempt by the computing community to improve the prediction of conflict outcomes but also to offer to social science tools useful for the design of new experimental settings, the detection of conflict signals in real contexts, and applications for conflict prevention and management. In fact, besides detection the dialog between social and computing science also helps tune educational tools (see Cheong et al., this volume) that can give former enemies or young people the possibility to better manage conflict.

Thus humanities and social psychology, which throughout history have often been to the service of the basic goals of power (keeping power, increasing

power, and demonstrating power, Morgenthau 1972, Christie et al. 2001) through this collaboration with computer science, rather become a tool for the analysis, detection, prediction, and understanding of social cues by means of sensitive and “aware” technologies for the prevention of conflict.

Parts and Chapters of This Book

The first part (“Theoretical Approaches to Conflict”) provides definitions and ontologies of conflict and explores its development, both concerning the stages of conflict escalation and concerning the evolution of the emotional and social mechanisms that feed conflict and its resolution.

The part is opened by Castelfranchi’s description of ontologies and dynamics of conflict. In his definition, there is a conflict whenever two or more goals (desires, needs, intentions, plans, norms, duties, order, or interests) are incompatible if pursued or fulfilled at the same time and in the same world. Due to their being “coherent seeking” devices, humans are sensitive not only to conflict with others but also to those between their own beliefs and goals, and any choice implies conflict. Besides formally defining the reciprocal connections between epistemic and social conflicts, but also between cooperation and competition, Castelfranchi examines the routes to conflict resolution, analyzing the notion of compromise, arguing how cognition is relevant in solving social conflicts, and stressing the cognitive requirements for a true negotiation—reducing an external conflict in an internal one, thus representing the other’s goals in one’s own mind. Yet he posits the existence of irremediable conflicts, when the parties have non-renounceable goals or independent reasons for feeding their reciprocal contrast. Stressing the “physiological” and even constructive aspects of conflict, not necessarily due—different from some philosophical hypotheses—to the agents’ selfish attitude, he finally points to its functions of social (r)evolution, emancipation and empowerment, improvement of science, and to its being the very bulk of democracy.

Porello, Bottazzi, and Ferrario, from the field of knowledge representation, sticking to the framework of social choice theory, define the notions of social agentive group, social propositional attitude, and social conflict, viewed as contradiction between two propositional attitudes of the same type (for instance, not a belief vs. an intention, but two beliefs or two intentions) in the formal system that represents them. Their analysis allows representing situations such as the Condorcet’s Paradox and the discursive dilemma and proposing a taxonomy of conflicts in terms of the type of agents involved and of the type of propositional attitude at issue, which can be in future integrated into the foundational ontology DOLCE.

Allwood and Ahlsén, after reviewing possible taxonomies of conflicts and previous models of stages in their escalation, integrate their theoretical claims with empirical research on the multimodal communication in conflict, analyzing the stance and behavior exhibited by politicians during conflictual episodes in American, Swedish, Italian, and Greek political debates. Drawing on the combinations

of features of multimodal behavior that express combinations of affective-epistemic states, they single out six short-term stages of conflictive interaction: (1) a preconflict phase, characterized by calm stances; (2) an initial confrontative claim, with attack or challenge; (3) a response to accusation, with an irritated or angry stance, possibly including a counterattack or acts of derision; (4) a further escalation, with repeated attacks and counterattacks; (5) a climax, characterized by turn overlap and high vocal intensity; (6) a final stage of superiority of the winner and defeat of the loser. Yet, comparing the situation of political debate with other types of social activities, like quarrel with neighbors and argument in a work group, they conclude that the number and type of stages are to a large extent determined by the type of activity and type of conflict: for instance, the stages of challenge/attack, response, and escalation are common to most conflicts but are probably necessary only in political debates. In the same vein, some types of communicative acts typical of conflicts, like pretending outrage or triumphant look, are probably determined by the particular setting of political debates, where some acts are by definition addressed more to the audience than to the antagonist.

Giardini and Conte provide a detailed cognitive analysis of revenge, viewed as a counteraggression aimed to re-establish a balance of power between actors, disrupted by an initial aggression that is framed as a *social damage*, an intentional disruption of one's power. Their in-depth analysis of the individual and social implications of revenge and their exploration of the avenger's motivations, encompassing an overview of the "cultures of honor" in which this behavior is strictly regulated, allows deepening also alternative mechanisms for the restoration of social damage, like punishment and sanction, and the regulation of revenge aimed at preventing further conflicts. They finally examine the intriguing issue of the adaptive function of this mechanism that, notwithstanding its costly side issues, still goes on operating in social interaction.

Adornetti looks at conflict in the communicative context of nonhuman primates, and the aim of her paper is to prove the cooperative nature of such interactions. In so doing, she calls into question Tomasello's model of language origins that supports an individualistic and competitive nature of nonhuman primates as opposed to the altruistic nature of humans and discusses some recent experimental data on chimpanzee vocal communication that go against that model. The results of these experimental studies support the idea that communication between apes is cooperative too, allowing Adornetti to argue for a kind of "altruism of knowledge" also in apes.

Chiera's contribution introduces to the emotional aspects of conflict from a philosophical point of view: in a phylogenetic perspective she highlights how emotions may have an adaptive value in terms of cooperation and group cohesion and at the same time generate hostility towards the outgroup. Furthermore, she describes some dynamics and rituals of nonhuman animals that allow a better interpretation of the relationship between conflicts and integration.

In Part II, the role of argumentative strategies is addressed by Paglieri, Lescano, and Bonelli.

To argue or not to argue? This question tackled by Paglieri deals with the following: while argumentation theorists have so far focused on premise adequacy or on the argumentative moves that further or impede the resolution of a difference of opinions between participants, Paglieri investigates whether the very fact of engaging in argumentation is always the best option. Is engaging in argumentation a strategy to gain consensus and eradicate conflict or, on the contrary, could it make things worse? Before engaging in an argumentative process, humans usually take into consideration its possible outcome; if they do decide to stand up for their thesis and defend it with argumentation, they make a series of choices as far as *when*, *how*, and *for how long* to expose the “opponent” to argumentation. In contrast with the traditional lack of scholars’ interest for this step, Paglieri analyzes argumentation as a decision-making process and proposes a taxonomy of argumentative decisions. As he points out, conflict plays a key role in all these decisions: arguers carefully examine the odds of winning the argument and strategically choose what to say and how to say it, also taking into consideration the opponent and context appropriateness.

Paglieri brings light on the argumentative planning and decisions one has to put into action in order to further the resolution of the original disagreement in one’s favor, while paying attention at the same time not to succumb to the dangers of generating additional conflicts.

Lescano’s interest also dwells in argumentation, namely in the counterargumentation advanced by an opponent to refute the protagonist’s initial thesis. His analysis sheds light on the reframing strategy, a not so common refuting strategy that antagonists can adopt to oppose the protagonist’s standpoint. The reframing strategy challenges the traditional idea of a difference of opinion consisting of two opposing standpoints: P and non-P. It consists in restating part of the position defended by the protagonist while at the same time modifying the way it must be interpreted by the audience. It implies, on the one side, granting the standpoint to the protagonist, while, on the other side, correcting it by adding further content and leading therefore to a reinterpretation of the initial thesis. This strategy gives the impression that the initial standpoint is a *partial* view of the situation, while the second improved version (the reframed position) appears to be more general in scope. Lescano analyzes three conflicting discursive sequences in French (one from a political debate broadcast on television and two from Internet forums) and illustrates two versions of the reframing strategy, depending on which part of the opponent’s position is maintained (internal vs. external reframing).

Conflicts can emerge not only in the contexts of a difference of opinions but also when interlocutors fail to achieve a state of “*interpersonal convergence*,” i.e., to relate emotionally to the speaker and be attuned to his expressions of affect, both linguistically and paralinguistically. An analysis of the emergence and development of such conflictive interactions elicited by a lack of interpersonal convergence is provided by Bonelli. Still in the context of Internet forum discussions, Bonelli analyzes multiparty conflict talk with a special focus on the pragmatic resources and sequential strategies by which the users express their stance. In a thread extracted from an online forum discussion, Bonelli shows how the interlocutors’

disaffiliation to the initiator of the thread can be detected and measured by means of markers such as Caffi and Janney's emotive devices. As illustrated by Bonelli's analysis, while the initiator of the thread seeks agreement and elicits approach, his interlocutors criticize him and detach from his emotive stance, using devices of negative evaluation and distance, while at the same time strengthening their own emotive stance as that of a compact unitary group.

Bonelli proves the importance of mitigating devices and empathetic attunement in avoiding or de-escalating conflict and shows how a lack of such elements can lead to a reinforcement of the distance and disagreement between parties.

Part III analyzes the "Communication of Aggression and Aggressive Communication" with approaches papers ranging from social psychology to logic and linguistics.

Zamperini and Menegatto, within the areas of memory, reconciliation, and recovery from the wrong received, analyze a particular case of mistreatment and abuse: the action and communication of the police over the demonstrators during the Genoa G8 Summit in 2001. Through lexicographic analysis they single out, in the expressions of the police reported by the victims, the devices of delegitimization theorized by Bar-Tal. With their flash on the challenging area of therapeutic jurisprudence, they provide a vivid example of how the narrative reconstruction of the victims' truth within the cooperative work of a trial may have healthy functions helping victims to recover their dignity and identity.

Scardigno, Giancaspro, Manuti, and Mininni deal with verbal aggression in a classical type of conflict, football cheer, analyzing the attacks to rival teams by Drughi, the fans of the Italian football team Juventus. A diatextual analysis singles out their values (cohesion, illegality, pride), their shared implicit assumptions (sexism, racism, sacred, and the role of silence), and ingroup and outgroup identities: while attributing themselves features of omnipotence, magic, and myth creation, the Drughi, just as the police of Zamperini and Mengatto's chapter, use strategies of delegitimization towards the outgroup, defining them as animals, biological entities, objects, and demons.

Vogel proposes a formal semantics of the language of (im)politeness, also relying on accounts in pragmatics and social psychology. His view is coherent with Elias that politeness and impoliteness behaviors are manifestations of offence management, with offence rooted in disgust: politeness is an adaptive mechanism aimed at mitigating disgust, while impoliteness arises from the experience of disgust triggered by the target and the consequent desire of the speaker to share this view of the target with others. If approaches to politeness in terms of "facework" focus on agents, and those in terms of "relational work" on their relations, Vogel's analysis views (im)politeness as the "management of a fog of offence that might otherwise engulf the whole network of agents and their relations."

Poggi, D'Errico, and Vincze propose a model of insults in terms of a socio-cognitive framework and distinguish them from bad words, curses, and imprecations. While bad words are single words concerning tabooed contents, a curse is a communicative act wishing a bad event to a target and rejecting any further relation to him; an imprecation is a curse or an insult to an object one makes responsible

for an unlucky event. An insult is a communicative act that includes the target into an abasing category with the intent of offending him and spoiling his image and self-image. The linguistic form of insults is overviewed, showing the connection between their syntactic and pragmatic constraints and their social goals, and direct and indirect verbal and bodily insults from a corpus of debates on TV and social media are analyzed and classified in terms of a pragmatic typology.

Part IV deals with the management and multimodal expression of emotions in conflicts on TV and social media, from politics to police interviews, from interpersonal to group interaction. Bonacchi and Mela propose a dynamic model for the analysis of low stakes conflicts. As in Goffman's view, politeness is seen here as a ritual functional to prevent conflict and to maintain the people's face. Yet, as the authors stress, during interaction the interactants create their faces and the reciprocal expectations about each other's face, and their face need engage in a process of power allocation; verbal aggression and impoliteness are then aimed at gaining power, denying the other conversational rights. In a conflictive dialog during a reality show, the authors show how all modalities contribute to this struggle for interactional power: speech acts, facial expressions displaying emotions, type, amplitude and frequency of gestures displaying excitement and managing spatial allocation, proxemics gestures aimed at attack and defence, vocal pitch, pauses, turn-taking management with overlaps and interruptions, and backchannel signals of disagreement.

Marzano, Scardigno, and Mininni, in the line of attribution theory, analyze the combined roles of empathy, truthfulness, social desirability, and emotional impact on positive intergroup attitudes. Their study, by assessing the perceived truthfulness and empathy felt after reading a story, demonstrates that the type and level of empathy activated by a story may be predicted by the inferences and judgments related to the truthfulness of the story, the social desirability of an empathetic answer, and the induced emotional impact: perceived truthfulness influences the emotional impact of the story, which causes social desirability. In turn, social desirability results as the direct predictor of empathy, finally determining positive intergroup attitudes. The experimental procedure adopted in the study allows testing the causal role of empathy and its interaction with cognitive and emotional factors in giving rise to a "social construction" of the conflict resolution.

Bruijnes, Linssen, op den Akker, Theune, Wapperom, Broekema, and Heylen analyze police officers' behaviors during interviews, trying to single out the interpersonal stances and the mental states reciprocally attributed. Their analysis, relying on Leary's solid pattern, based on the crossed dimensions of dominance/submissiveness and willingness to cooperate with the listener, allows describing fragments of the corpus and capturing more or less conflictual scenarios. With the inclusion of constructs such as "face saving" and "politeness" and metaconcepts like "information" and "strategy," the work gives an account of the captured interactions and suggests possible solutions for the construction of a virtual suspect to be used for the training of police students. The chapter demonstrates how it is possible to apply solid theoretical approaches in real-life contexts and to go from behavioral analysis to technologies useful for conflict prevention.

An interesting “arena” of conflict encompassing both corpora of analysis and tools for the overcoming of conflict through rhetorical strategies is political communication. In such a context, Leone, Di Murro, and Serlupi exploit the concept of “pariah,” born in Arendt’s speculation, as an empowering communicative game to analyze some of Obama’s crucial speeches. Their analysis of multimodal communication and of the facial expressions of emotions of the leader shows that during the autobiographical narrative of Obama as a pariah, negative emotions are expressed but, coherently with persuasive communication, well regulated. The multimodal explorative analysis suggests that his parrhesiastic attitude—his veracity—is a powerful way of persuading the audience to accept a similar game concerning political difficulties.

The last part of the book, “Technologies for Conflict Detection and Simulation,” includes technological contributions aimed at automatic analysis and understanding of conflict in human-human interactions. Two contributions revolve around the automatic analysis of conflict level in political debates, a setting where conflict is particularly frequent given that the participants tend to pursue incompatible goals (e.g., if one participant acquires consensus, the other ones lose it). Caraty and Montacié show that the detection of interruptions plays a major role in the automatic measurement of conflict. Their approach is based on a detailed analysis of the SSPNet Conflict Corpus showing that the conflict level tends to be perceived as higher when the number of interruptions increases. Following this observation, the authors develop an interruption detection approach that allows them to predict the conflict level of short audio samples (30 s) extracted from political debates.

The contribution by Brueckner and Schuller proposes experiments over the same data as the work above and focuses on the adoption of deep neural networks, one of the most successful machine learning methodologies proposed in the last years. In particular, Brueckner and Schuller show that these algorithms achieve the highest performances proposed so far in the literature for both regression and classification tasks associated to the SSPNet Conflict Corpus.

Koutsombogera et al. target another scenario where conflict is frequent, namely conversations between call center operators and their customers. The overall goal of the work is to detect conflict in order to assess and possibly improve the quality of the services that call centers offer. The approach proposed in the chapter by Koutsombogera et al. relies on two main stages: the first is the detection of negative emotions likely to be elicited by conflict and the second is the analysis of turn-taking patterns likely to take place during conflictual interactions. The main conclusion of the authors is that the detection of conflict is a complex process that must include the detection of multiple cues.

The last contribution of this part (Cheong et al.) shows that serious games can help people to acquire conflict management skills. The authors describe experiments where young children (9–12) play a game where the participants are involved in conflicts of different intensity. This serves a double purpose: on the one hand,

it is possible to analyze the interplay between variables like gender, age, conflict resolution strategy type, cultural tendency, reported emotions and perception about the other players, and reported conflict intensity. On the other hand, the game allows one to better understand conflict dynamics and to elaborate effective strategies for conflict management.

Roma, Italy
 Roma, Italy
 Glasgow, UK
 Roma, Italy

Francesca D’Errico
 Isabella Poggi
 Alessandro Vinciarelli
 Laura Vincze

References

- Arendt H, Varnhagen R (1997) *The life of a Jewess*. Weissberg L (ed). Johns Hopkins University Press
- Bar Simon Tov Y (2004) *From conflict resolution to reconciliation*. Oxford University Press
- Bar-Tal D (2010) *Intergroup conflicts and their resolution: a social psychological perspective*. Front Soc Psychol. Psychology Press
- Bar-Tal D, Teichman Y (2005) *Stereotypes and prejudice in conflict: representations of Arabs in Israeli Jewish society*. Cambridge University Press, New York
- Caffi C, Janney RW (1994) Toward a pragmatics of emotive communication. *J Pragmatics* 22:325–373
- Cassell J (2000) More than just another pretty face: embodied conversational interface agents. *Communications of the ACM* 43(4):70–78
- Cassell J, Sullivan J, Prevost S, Churchill E (2000) *Embodied conversational agents*. MIT Press, Cambridge, MA
- Castelfranchi C (2015) The cognition of conflict: ontology, dynamics, and ideology. In: D’Errico F, Poggi I, Vinciarelli A, Vincze L (eds) *Conflict and multimodal communication: social research and machine intelligence*. Springer, New York, pp 3–32
- Christie DJ, Wagner RV, Winter DD (eds) (2001) *Peace, conflict, and violence: Peace psychology for the 21st century*. Prentice-Hall, Upper Saddle River, NJ. <http://academic.marion.ohio-state.edu/dchristie/Peace%20Psychology%20Book.html>
- D’Errico F, Poggi I (2014) Acidity the hidden face of conflictual and stressful situations. *Cogn Comput* 6(2)
- D’Errico F, Poggi I, Vincze L (2013) Discrediting body. A multimodal strategy to spoil the other’s image. In: Poggi I, D’Errico F et al (eds) *Multimodal communication in political speech shaping minds and social action*, vol 7688, Springer, LNAI., pp 181–206
- de Rosis F, Pelachaud C, Poggi I, Carofiglio V, De Carolis B (2003) From Greta’s mind to her face: modelling the dynamics of affective states in a conversational embodied agent. *Int J Human Comput Stud* 59(1)
- Deutsch M, Coleman PT, Marcus EC (eds) (2006) *The handbook of conflict resolution: theory and practice*. Jossey
- Diamantini D, Pietroni D (2002) *La fiducia nelle interazioni negoziali: uno studio empirico*. Cheli Enrico. La comunicazione come antidoto ai conflitti. Punto di Fuga Editore, Cagliari
- Ekman P, Friesen W (1978) *Facial action coding system: a technique for the measurement of facial movement*. Consulting Psychologists Press, Palo Alto
- Elias N (1969) *The civilizing process, vol I. The history of manners*, Blackwell, Oxford

- Giebels E, Ufkes EG, Van Erp KJMP (in press) Understanding high-stakes conflicts. In: Ashkanasy N, Ayoko OB, Jehn KA (eds) *The handbook of research in conflict management*. Edward Edgar Publishing, London
- Kelman HC (2002) Interactive problem solving: informal mediation by the scholar-practitioner. In: Bercovitch J (ed) *Studies in international mediation: essays in honor of Jeffrey Z. Rubin*. Palgrave Macmillan, New York, pp 167–193
- Kelman HC (2008) Reconciliation from a social-psychological perspective. In: Nadler A, Malloy T, Fisher JD (eds) *Social psychology of intergroup reconciliation*. Oxford University Press, New York, pp 15–32
- Kelman HC (2010) Conflict resolution and reconciliation: a social-psychological perspective on ending violent conflict between identity groups. *Landscape Violence* 1(1):Article 5
- Kendon A (2004) *Gesture. Visible action as utterance*. Cambridge University Press, Cambridge
- Leary T (1957) *Interpersonal diagnosis of personality: functional theory and methodology for personality evaluation*. Ronald Press, New York
- Leone G, D’Errico F (2007) Identità senza nemici. Una riflessione sui processi di riduzione dell’ostilità e di riconciliazione matura. In: Talamo A (a cura di) *La pluralità inevitabile. Identità in gioco nella vita quotidiana*. Apogeo, Roma, pp 71–103
- Lewin K (1935) *A dynamic theory of personality*. McGraw-Hill, New York
- McNeill D (1992) *Hand and mind*. University of Chicago Press, Chicago, IL
- Morgenthau H (1972) *Science: servant or master*. New American Library, New York
- Morris D (1977) *Manwatching*. Jonathan Cape, London
- Moscovici S (1976) *Social influence and social change*. Academic, London
- Mugny G, Doise W (1978) Socio-cognitive conflict and the structure of individual and collective performances. *Eur J Soc Psychol* 8:181–192
- Nadler A, Malloy T, Fisher DJ (2008) *Social psychology of intergroup reconciliation from violent conflict to peaceful co-existence*. Oxford University Press, Oxford
- Narayanan (2013) Presentation in the workshop “conflict and communication. Multimodal social signals of conflict and negotiation in humans, animals, and machines”. Rome, October 2013 www.sspnet.eu; <http://www.klewell.com/conferences/sspnet-roma-2013/>
- Navaretta C (2013) Predicting speech overlaps from speech tokens and co-occurring body behaviours in dyadic conversations. In: *Proceedings of the 15th ACM International Conference on Multimodal Interaction (ICMI 2013)*. Assoc Comput Machinery, pp 157–163
- Niewiadomski R, Ochs M, Pelachaud C (2008) Expressions of empathy in ECAs. *Lect Notes Comput Sci* 5208:37–44
- Noor M, Shnabel N, Halabi S, Nadler A (2012) When suffering begets suffering: the psychology of competitive victimhood between adversarial groups in violent conflicts. *Pers Soc Psychol Rev* 16:351–374
- Paggio P, Navaretta C (2013) Head movements, facial expressions and feedback in conversations – empirical evidence from Danish multimodal data. *J Multimodal User In* 7(1–2):29–37
- Pesarin A, Cristani M, Murino V, Vinciarelli A (2012) Conversation analysis at work: detection of conflict in competitive discussions through automatic turn-organization analysis. *Cogn Process* 13(2):533–540
- Poggi I (2007) *Mind, hands, face, and body. A goal and belief view of multimodal communication*. Weidler, Berlin
- Poggi I, D’Errico F (2010) Dominance in political debates. In: Salah AA et al (eds) *Human behavior understanding*, LNCS 6219. Springer, Heidelberg, pp 163–174
- Poggi I, Vincze L (2009) Gesture, gaze and persuasive strategies in political discourse. In: *Multimodal corpora*. *Lect Notes Comput Sci* 5509:73–92
- Poggi I, D’Errico F, Vincze L (2010) Types of nods. The polysemy of a social signal. In: Calzolari N et al (eds) *Proceedings of the seventh conference on international language resources and evaluation (LREC’10)*. Malta. May. European Language Resources Association (ELRA), pp 19–21. ISBN 2-9517408-6-7
- Poggi I, D’Errico F, Vincze L (2012a) Comments in words, face and body. *J Multimodal Users In*. doi:10.1007/s12193-012-0102-z

- Poggi I, D'Errico F, Vincze L (2012b) Ridiculization in public debates: making fun of the other as a discrediting move. In: Calzolari N et al (eds) Proceedings of the 8th conference on international language resources and evaluation (LREC'12). Istanbul 21–27 May. European Language Resources Association (ELRA), pp 44–50
- Pruitt et al (2003) Social conflict: escalation, stalemate, and settlement. McGraw-Hill Ser Soc Psychol
- Rahayudi B, Poppe RW, Heylen DKJ (2014) Twente debate corpus – a multimodal corpus for head movement analysis. In: Proceedings of the international conference on language resources and evaluation (LREC 2014), European Language Resources Association (ELRA), Luxembourg, pp 26–31
- Rienks RJ, Poppe RW, Heylen DKJ (2010) Differences in head orientation behavior for speakers and listeners: an experiment in a virtual environment. *ACM Trans Appl Percept* 7(1):2–14
- Scherer KR (1981) Speech and emotional states. In: Darby J (ed) *Speech evaluation in psychiatry*. Grune and Stratton, New York, pp 189–220
- Scherer KR, Ekman P (1984) *Approaches to emotion* (ed) Lawrence Erlbaum Associates, New Jersey
- Scherer KR, Grandjean D (2008) Inferences from facial expressions of emotion have many facets. *Cogn Emotion* 22:789–801
- Sherif M, Harvey OJ, White BJ, Hood WR, Sherif CW (1961) *Inter-group conflict and cooperation: the Robber's cave experiment*. University of Oklahoma Press, Norman, OK
- Shnabel N et al (2013) Overcoming competitive victimhood and facilitating forgiveness through re-categorization into a common victim or perpetrator identity. *J Exp Soc Psychol*
- Solano Méndez G, Reidsma D (2011) A BML based embodied conversational agent for a personality detection program. In: International conference on intelligent virtual agents (IVA (2011) Vilhjálmsson HH, Kopp S, Marsella S, Thórisson KR (eds) *Lect Notes Comput Sci* 6895. Springer, Berlin, pp 468–469
- Streeck J (2007) Gesture in political communication. A case study of the democratic presidential candidates during the 2004 primary campaign. *Lang Soc Interac*
- Tajfel H, Turner JC (1986) The social identity theory of intergroup behavior. In: Worchel S, Austin WG (eds) *The social psychology of intergroup relations*. Nelson-Hal, Chicago, IL
- Waisman OS (2010) *Body, language and meaning in conflict situations: a semiotic analysis of gesture-word mismatches in Israeli-Jewish and Arab discourse*. John Benjamins Publishing Company, Studies in functional and structural linguistics
- Watzlawick P, Bavelas JB, Jackson DD (1967) *Pragmatics of human communication. A study of interactional patterns, pathologies and paradoxes*. Norton & Co., New York

Contents

Part I Theoretical Approaches to Conflict

1	The Cognition of Conflict: Ontology, Dynamics, and Ideology	3
	Cristiano Castelfranchi	
2	Group Conflict as Social Contradiction	33
	Daniele Porello, Emanuele Bottazzi, and Roberta Ferrario	
3	On Stages of Conflict Escalation	53
	Jens Allwood and Elisabeth Ahlsén	
4	Revenge and Conflict: Social and Cognitive Aspects	71
	Francesca Giardini and Rosaria Conte	
5	Competition and Cooperation in Language Evolution: A Comparison Between Communication of Apes and Humans	91
	Ines Adornetti	
6	The Price of Being Social: The Role of Emotions in Feeding and Minimizing Conflicts	103
	Alessandra Chiera	

Part II Argumentation and Conflict

7	Arguments, Conflicts, and Decisions	117
	Fabio Paglieri	
8	Common Ground or Conceptual Reframing? A Study of the Common Elements in Conflicting Positions in French Interactions	137
	Alfredo M. Lescano	

9	Disaffiliation and Pragmatic Strategies of Emotive Communication in a Multiparty Online Conflict Talk	159
	Laura Bonelli	
 Part III Communication of Aggression and Aggressive Communication		
10	Giving Voice to Silence: A Study of State Violence in Bolzaneto Prison during the Genoa G8 Summit	185
	Adriano Zamperini and Marialuisa Menegatto	
11	The Rhetoric of Conflict Inside and Outside the Stadium: The Case Study of an Italian Football Cheer Group ...	207
	Rosa Scardigno, Maria Luisa Giancaspro, Amelia Manuti, and Giuseppe Mininni	
12	Some Puzzles of Politeness and Impoliteness Within a Formal Semantics of Offensive Language	223
	Carl Vogel	
13	Direct and Indirect Verbal and Bodily Insults and Other Forms of Aggressive Communication	243
	Isabella Poggi, Francesca D'Errico, and Laura Vincze	
 Part IV Emotions and Multimodal Communication in Conflict		
14	Multimodal Analysis of Low-Stakes Conflicts: A Proposal for a Dynamic Model	267
	Silvia Bonacchi and Mariusz Mela	
15	Rhetoric of Truthfulness in the Battle Between Social Attributions and Empathic Emotions	295
	F. Milena Marzano, R. Scardigno, and G. Mininni	
16	Social Behaviour in Police Interviews: Relating Data to Theories	317
	Merijn Bruijnes, Jeroen Linssen, Rieks op den Akker, Mariët Theune, Sjoerd Wapperom, Chris Broekema, and Dirk Heylen	
17	From Personalization to Parrhesia: A Multimodal Analysis of Autobiographical Recalls in Barack Obama's Political Speech	349
	Giovanna Leone, Francesca Di Murro, and Livia Serlupi Crescenzi	

Part V Technologies for Conflict Detection and Simulation

18 Detecting Speech Interruptions for Automatic Conflict Detection 377
Marie-José Caraty and Claude Montacé

**19 Be at Odds? Deep and Hierarchical Neural Networks
for Classification and Regression of Conflict in Speech** 403
Raymond Brueckner and Björn Schuller

20 Conflict Cues in Call Center Interactions 431
Maria Koutsombogera, Dimitrios Galanis,
Maria Teresa Riviello, Nikos Tseres, Sotiris Karabetsos,
Anna Esposito, and Harris Papageorgiou

**21 Serious Games for Teaching Conflict Resolution:
Modeling Conflict Dynamics** 449
Yun-Gyung Cheong, Rilla Khaled, Christoffer Holmgård,
and Georgios N. Yannakakis

Index 477

Part I
Theoretical Approaches to Conflict

Chapter 1

The Cognition of Conflict: Ontology, Dynamics, and Ideology

Cristiano Castelfranchi

1.1 Premise

What is a conflict? What is the relationship between actors' mental representations (e.g., beliefs, goals) and the conflicts between them? What is the relation between contradictions and conflicts and the need for mental coherence?

How to build a systematic ontology of conflicts taking into account objective and subjective types of conflict, the internal or individual and the external or social?¹

Are there objective conflicts that agents are unaware of? How do they work?

And what is the relation (if any) between individual/subjective conflicts (among my goals) and social conflicts? Do external conflicts require internalized/mentalized conflicts? How do we resolve personal or social conflicts?

These fundamental questions will form the basis of a discussion of three crucial issues:

- (a) *Grounding the theory and ontology of conflict in cognition.*
- (b) *Conflict and cooperation*, the two intertwining faces of sociality, where social action also aims at *changing* another person's mind and behavior, not just adjusting one's own behavior to circumstances or exploiting or blocking the others' behavior. Sociality is aimed at mind manipulation, influence, and power. Mind reading is aimed at cooperation, conflict, or both. The same holds for language/conversation and for argumentation.

¹A preliminary version of such an ontology can be found in Castelfranchi (1996).

C. Castelfranchi (✉)

ISTC – Institute of Cognitive Sciences and Technologies, National Research Council,
Via San Martino della Battaglia 44, Roma 00185, Italy
e-mail: cristiano.castelfranchi@istc.cnr.it

- (c) *Ideologies of conflict*: conflict (e.g., interpersonal, personal, social) is felt and treated as a bad thing or situation! However, we must also consider the usefulness of conflict. For example, competition is the best mechanism for resource allocation following Hayek's (1978) liberalism: no reasoning or planning can match it. Conflict is the essence of democracy.

1.2 Conflict Theory

For the theory of conflicts we need a double (but interrelated) foundation:

1. *Cognitive foundation*

Conflict theory must be based on the following considerations:

- An explicit and systematic theory of our specific *goals*, motivations that make something a means, a good, a resource for us.
- In fact empty theory of “maximizing utility” or of “maximizing pleasure” does not tell us so much about conflict; it cannot predict where we compete and where we cooperate. Conflict is due to specific/*content* goals and to specific *interferences* in a common world (“strategic” situations). When will interference generate conflict? And when will it generate cooperation or exchange?
- A theory of *beliefs*, since both internal and social conflicts are *based on our beliefs* about goals and their interference/dependence.

2. *Structural foundation*:

- A theory of a common world, of interference, of dependence, and power relations.

However, related to actors' beliefs, goals, and subjective powers, are the following issues:

1.2.1 *What is a conflict?*

A conflict consists of ²

²There is an even broader and a rather metaphoric notion in common usage, where “conflict” simply means any incompatibility between A and B in the same given context, for whatever reason. They are “in contrast” (objectively or subjectively: from the point of view of the perceiver); they cannot remain together; one tends to rule out the other, even from a simple aesthetic or perceptual point of view (like clashing colors or sounds, or like a light in front of me, and my goal of looking in a certain direction. To use the metaphor of conflict, what is needed is to look at the word (even physical force) from the perspective of “finalities.”

1. Two or more **goals** (e.g., *desires, needs, intentions, plans, norms, duties, orders*) or **interests** (potential goals and relative to goals) (Sect. 1.2.6)
2. that are incompatible
3. if carried out or pursued in the same/common world.

This is the notion we will use. However, the commonsense meaning is broader and can also be psychologically relevant. We may have conflicts not just between true goals but between goals (like intentions) and impulses, or there may be S-R reactions, automatic routines and habits, or conflicts with or between emotions (with their goals and impulses).

More generally, there is a possible conflict between the *two control systems* of our actions: explicit (reason-based) true “decisions” and other interfering mechanisms (either conscious or unconscious).³

Thus, a more general definition might be as follows:

A conflict exists when two mechanisms or processes or entities that control or activate/orient behavior would lead to incompatible conduct either from the executive point of view or in terms of their outcomes (for intrinsic, logical reasons or for extrinsic, practical ones).

1.2.2 Conflict (Cognitive) Ontology: Individual, Internal, or Intraagent Conflicts

Conflicts exist because we are *multipurpose systems*, not simply guided by stimuli or by a momentary impulse or “choosing” between several activated goals randomly or by design. Conflict is cognitively fundamental and behaviorally adaptive because it allows us to deal with several active goals at the same time, to discover contradictions, and to anticipate self-defeating behaviors; further, it makes it possible to exploit predictions, reasoning, problem solving, cognitive evaluations, and, thus, real formulate reason-based preferences and choices.

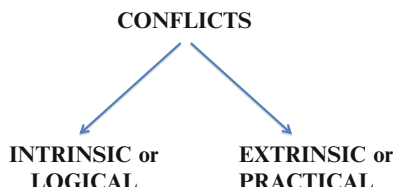
By grounding conflict theory on goal theory (Castelfranchi and Paglieri 2007) we can introduce the following categories.

I call *individual/private* or so-called *intrapersonal* conflicts those that arise between two goals in the same mind: X has both the goal (desire, intention, duty, need) P and the goal (desire, intention, duty, need) Q, but P prevents the realization of Q (and vice versa).

³See “dual system” or “processing” theories by Evans; Kahneman; Sloman; Stanovich and others. For a review, see Evans (2008).

1. Either X has (which does not mean “pursues”!) and formulates the goal P and the goal Not P; thus, they are *intrinsically* incompatible or logically *contradictory* (I would like my father to be alive and I would like him to be dead);
or X has the goal P and the goal Q, but P logically *implies* Not Q (I would like my father to be alive and I would like someone to kill my father).
2. Or P just “practically” entails that Not Q since it makes Q *practically impossible*:
They are incompatible because they require the same resource (e.g., money, time, place), which does not exist in sufficient quantities, i.e., it is scarce, or they require two incompatible practical conditions for their achievement.
(Of course, the practical case can be phrased as – and finally entail – a logical contradiction: if I pursue P, then, since pursuing P means not realizing/pursuing Q, this implies that Not Q: in a sense, in pursuing P I am pursuing Not Q) (Fig. 1.1).

Fig. 1.1 Two kinds of conflicts



3. *Conflicts can exist between two different kinds of goals*:
An intention can be in contrast with, for example, a set-aside *desire*; a desire can contradict a *duty*, or a duty (or desire or intention) a *felt need*.
Any goal can elicit conflicts: this is a property of “goalness,” not of desires or needs or pursued goals per se.
4. *There are “explicit” and “implicit” conflicts*:
Since we do NOT derive all the logical consequences of what we know or believe (no “closure”), which would be 99 % redundant and cumbersome, we just infer *what* and *when* we need it; and since, for this reason, *a large proportion of our “knowledge” is just implicit and more precisely “potential”* in our minds/brains, it follows that it is possible (frequent!) that we have the goal that P and the goal that Q, they are/would be in conflict since Q implies Not P, but we are not aware of this, we do not derive “Not P” from Q. Thus, *subjectively speaking* we have no conflict at all. We will realize this either by reasoning or, when pursuing both, and by discovering that one thwarts or harms the other.
5. Not only might there be objective conflicts between my goals that I do not realize, but the other converse might also hold. It is possible that subjectively I perceive as a conflict desiring or pursuing P and desiring/pursuing Q, though they are not incompatible at all. But subjectively what matters is that I *believe* them to be incompatible, for example, that I (wrongly) derive Not P from Q.

1.2.3 *Strength and Quality*

The nature, strength, and resolution of a conflict depend on the involved goals:

1. The strength of a conflict (more or less hard) derives from the *value V of the involved goals*.
2. The shorter the interval, meaning the smaller the difference between the V of the competitors' goals, the harder the decision: the goals are equivalent, and there is little basis for showing a preference for one over the other (especially if V is high).

These two dimensions are combined: the hardest conflict is where
 $V \text{ of } P = V \text{ of } Q$, with very high values.

3. Any subjective conflict requires “giving up” or “losing” something, some goal that must be sacrificed; the greater the value of that goal, the harder the decision. (We might even have a threshold of *unacceptable losses* that prevents us from taking any decision; we do not want to bear the responsibility of such a loss, independently of the gain).
4. Actually, choices must be between expectations, not just goals. Thus, what matters is not only the value of (goal) but also the ascribed probability or, better, the perceived possibility, the *strength* of the expectation. A lower-value goal can prevail over a high-value goal if it is perceived as much being more realistic, accessible, or likely.

Moreover, there are so-called avoidance goals and achievement/maintenance goals. For example, I have the goal that not P, to avoid the occurrence or continuation of P, versus the goal that P may remain or become true. And we have conflicts between two positive Gs, or between two negative goals, between two harms, losses, dangers (Lewin 1935).

There are also so-called ambivalent goal states: outcomes that are partly positive (realization of goals) and partly negative (frustration of goals). *Any (to be) pursued G should in fact be ambivalent since it necessarily implies some cost and risk*.

5. Also, the *qualitative* nature of the goal matters. If the conflicting goal is not a “motivating” one but merely a pleasant side effect, an additional benefit, its role is different: giving up is simpler. It does not just depend on the value of the goal.

There are, on the contrary, unrenounceable goals: for example, values, symbolic and identitarian goals, or the will of dogmatic authorities. No mediation is possible, it is either yes or no! All or nothing (Sect. 1.8.4).

1.2.4 *Conflicts Between Different Reflection Layers*

There are conflicts at different mental layers and between layers.

For example, a conflict between a given goal of mine and my goal of not having such a goal!

REFLECTION Conflict:

(Goal X P) and (Goal X (Not (Goal X P)))

or

(Goal X P) and (Goal X (Goal X (Not P)))

For example, I have a goal of exhibiting my own qualities, but I blame myself for my vanity; that is, I have the goal of not showing off to be admired.

METACOGNITIVE Conflict:

(Goal X (Goal X P)) and (Goal X (Not (Goal X P)))

is a conflict between two meta goals about my goals.

1.2.5 *Private/Individual Conflicts with Functions*

Our behavior may respond to two kinds of teleology: finalities, which are *internal*, driving goals [see the control theory model in Evans (2008)], versus *external* selective “functions,” which are biological or social (Castelfranchi 2001). Now, there might be conflicts between the internal goal of the subject and the “function” of his action/behavior; also, because we do not necessarily understand and thus intentionally pursue our biological or social functions.

An example of our conflict with a *biological function* could be the aim of having sex while avoiding procreation, which is the real goal (adaptive function, fitness advantage) of sex.

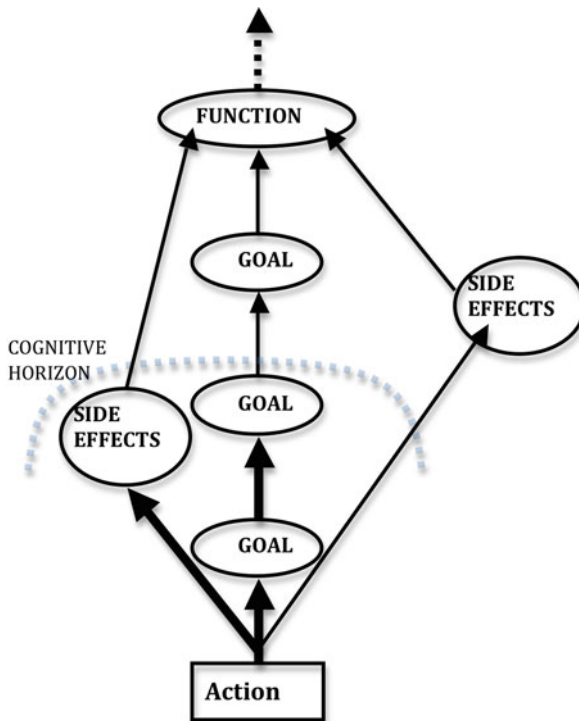
An example of a conflict between our goals and our *social functions* could be the goal that B be condemned while I am his defense attorney (Fig. 1.2).

In general, we carry out our social functions and roles (e.g., citizen, consumer, father, pedestrian, child) quite blindly, and this is not just due to the fact that they are carried out unconsciously or just based on reinforced learning or on mere *habitus* [in Bourdieu’s (1980) sense]. In fact, also our intentional and deliberate individual actions (evaluated on the basis of their visible and conceivable consequences) may “pursue” collective (bad) “ends” (Castelfranchi 2001).

If we, for example, realized our irrationality and understood how marketing induced so-called needs and deceived and manipulated us, we could not play very well our most crucial role in/for society, that of consumers! Our unawareness is very functional to society in that it supports markets and banks.

Functions install and maintain themselves in a *parasitical* relation to cognition: *thanks to and through agents’ mental representations but not as mental representations, i.e., without being known or at least intended*. Thus, it is possible (and frequent) that, following our personal motives, we play our roles in contradiction with the mission and collective utility of our social function.

Fig. 1.2 Goals vs Functions



1.2.6 Conflicts with One’s Own Interests

“Interest” (in one of its meanings) is a very crucial motivational notion; when we say “P would be *in* Y’s best interest,” this of course presupposes some goal G1 in Y’s mind, and P would like to realize or favor that goal⁴, but it is not already a goal of Y (instrumental to G1). Only if Y believes/knows that his best interest for G1 is P will he formulate the subgoal that P (G2).

Interests are *potential* goals of an agent. This possible ignorance of our own interests is what is really important in this notion.

In fact, since it is possible that Y does not understand which is in his best interest, he will not act in that direction; he will use bad or self-defeating strategies for his goals. Thus, it may be that while pursuing his goal G3 Y acts in *objective* conflict with some other goals of his, like G1.

For example, I do not realize that selling my stocks for a good price now is actually against my financial interest because they will be continue to increase in price. Or I do not realize that as a strike breaker I am going against my group or class interest and in the long term against my personal interests.

⁴While “contrary to Y’s best interest” means that P creates an obstacle or frustrates Y’s goals.

It is very relevant that another agent, X, might know what would be in the best interest of Y, though Y himself does not know. This may create a very important social relation, the “tutorial” relation: when X cares about Y’s welfare (goal achievement) and tries to encourage Y to pursue his own interest and not his current preferences. This creates a crucial and strange *social conflict* between X and Y, a “tutorial conflict,” not between the selfish goals of X and the selfish goals of Y, but between two goals of Y, one represented by X and the other by Y (Castelfranchi and Falcone 2000).

This conflict is typical of that between parents and children, doctors and patients, teachers and students, government and the people. The problem is this: how can X be sure that something is “for the good of Y” if Y does not agree that it is? Is X (consciously or unconsciously) following her own interests or conforming to the pressures of society and its customs and expectations?

The most dramatic case I know of where a conflict was the familiar one was between Vincent van Gogh and his brother Theo, who “lovingly” had Vincent placed in a mental hospital (of course, “for his own good”) and finally induced Vincent to commit suicide to liberate his poor, suffering family from the burden of his life as a tramp.

1.3 Conflict vs. Contradictions: The Mind as a Coherence-Seeking Device

There is a special relation between the contradiction between two epistemic/doxastic representations (like beliefs) and the conflict between two motivational representations in both cases the mind needs to arrive at coherence. Let us examine some aspects of this relation.

1.3.1 Coherence Seeking and Epistemic Conflicts

Propositionally conflicts are equivalent to explicit or implicit contradictions (P and NotP), but a contradiction, to become/generate a real conflict, must be between two motivational, not just doxastic, mental attitudes.

If I definitely believe that P and definitely believe that Not P, I’m in a contradiction, whereas if I would like that P and would like that Not P, I’m in a conflict. However, there are true “conflicts” in the epistemic domain.

Epistemic conflict. This is a conflict between, for example, two beliefs, inferences, or opinions.

More precisely, directly or indirectly *contradictory beliefs* become or create a “conflict” if and only if:

- we have/want to agree (goal) (*societal* level) or
- I must “decide” what to believe (goal).

The *epistemic conflict* might be considered a subtype of *metacognitive conflicts*: the conflict is in fact between the goal/decision of believing P and the goal of not believing P (or believing Not P). There is such a conflict between the belief that P and the belief that Not P only because there is a meta-goal of having coherent/consistent beliefs (see below).

Any new information must be integrated with our previous knowledge/beliefs in that context. Such information might even be rejected, not believed.

To be accepted, it must be *compatible* with previous knowledge, that is, at least noncontradictory. If there is an epistemic conflict, then I must *revise* my previous incompatible beliefs if the new information is much more credible (thanks to its sources). Hopefully it should not just be noncontradictory but supported by or supporting the other beliefs, in other words, well integrated: this is expressed by words like *in fact*, *since*, *obviously*, and so on.

However, what is the basis of the strength of possible *conflicting* beliefs? What determines the strength of the *opposers* of the new information? Clearly their certainty is due to their own sources and the reliability and convergence of those sources. It is also due to the degree of integration and reciprocal support among previous beliefs, but this is a source, too, since those beliefs constitute inferential links: I can “derive” (support) a given assumption from other assumptions.

This is in fact an additional reason for our good memory for the sources and origins of our beliefs. Sometimes, when we have a problem, we even try to explicitly retrieve the source of some knowledge item: *Who told me this? Where did I read this?*

This is very important since it means that in fact *usually a conflict between a previous belief and a new candidate belief represents a contest between their sources*: what are the most reliable sources, and how many are in favor of or opposed to the old and new beliefs?

As stated earlier, memorized or retrieved sources can be of any type: perception and direct experience, communication (something heard or read), or one’s own reasoning (inferences).

Given the preceding statements, *a knowledge conflict is in fact a contest over sources*; that is why arguing against or in favor of or about a source will constitute a major argument: “*But Encyclopedia Britannica is much more credible and serious . . .*” “*You cannot believe that; they are interested in selling a drug and are not a serious company.*” “*You can believe that, it is a medical network.*” “*You cannot believe that; they are just laypeople and emotionally involved in this disease . . .*”

1.3.2 *Conflict Dynamics or Processing*

An agent cannot – rationally (and consciously) – commit herself to two contradictory goals or beliefs. But there are stages of belief and of goal processing that are preliminary to any commitment.

Desires need not be either realistic or coherent, whereas intentions do.

The same holds for mere hypotheses with respect to true beliefs:

- Given a conflict between two impossible desires, choice and conflict resolution are not still required.
- If two conflicting desires must pass into the state of being an *intention* to be carried out, one must resolve the conflict and choose. In the postdecision stage, among current active intentions there cannot be conflicts.

Thus:

- (a) There is a conflict only at the same level/stage of processing.

It must be stressed that conflict is not only relative to contexts and beliefs (in a static perspective); (in a dynamic perspective) it is relative to the level of processing.

- (b) The conflict creates a “problem” to be solved (for example, by a choice) *only if* the goals (beliefs) are at or should reach a level of processing in which the agent is required to be coherent and to commit herself to a given belief or goal.

The mind is a coherence-seeking device. It is really remarkable to bring relating all these cases of contrast to just one and the same relation between the propositional content to some sort of general principle of “disharmony”:

1. Epistemic contradiction: between beliefs (Bel X P) and (Bel X or Y Not P).
2. Goal “conflict”: (Goal X P) and (Goal X or Y Not P).
3. Discrepancy: cybernetic mismatch in the control cycle between beliefs and goals: (Goal X P) and (Bel X Not P).

1.3.3 *Coherence in Goals and Plans*

Goal processing and intention formulation, besides introducing coherence among active goals, also make our actions and plans *intracoherent* (the actions of a plan do not undermine the overall architecture owing to conflicts) and *intercoherent* (plans that are simultaneously executed do not undermine each other) (Castelfranchi and Paglieri 2007) (Fig. 1.3).

Commitment (in intention formation) is aimed at *reducing future possible conflicts* and fostering stability, persistence, and coherence.

In a sense, the “strength of will” also has the function of managing internal conflicts: it serves to pit me socially against myself, to influence me to choose, to do, to persist.

In this sense, the mind is a coherence-seeking device: its processing also aims at this function.



Fig. 1.3 The mind as a coherence-seeking device

1.4 Conflict and Decision

Conflict does not mean that X is unable to choose or that choice is difficult. Conflict just means that X must choose, that s/he is considering two goals/alternatives that are (subjectively) incompatible. *Any choice presupposes a goal conflict.*

1.4.1 In Search of Conflicts

Conflicts are the presupposition of choice, in two senses. We must choose just when and because there is a conflict between our active goals; however, conflicts are also the condition for the choice, and we may actively search for them *in order* to give a differential value to the candidate goals and to build a preference, a difference of value.

If I have to choose between goals G1 and G2 but they do not have enough differential value, I search for differential consequences of the two scenarios and evaluate those consequences and additional outcomes against new goals that I activate from the bottom up. I build/imagine the pros and cons of G1 and of G2 in order to give them different values. However, the identified pros and cons, the “arguments” in favor of or against G1 and G2 that will become the goals settling my decision, must necessarily be incompatible, differential, that is, in conflict. If both G1 and G2 have the same additional effect (G3), this is not useful for choosing between G1 and G2; we need a goal G3 in addition to G1 and a goal G4 in addition to G2: a specific advantage or a specific cost of G1 and G2. That is, if I choose G1, then I also get G3, but I give up or avoid G4, and vice versa.

In sum, to choose between G1 and G2 (which might even be perfectly equivalent, for example, as alternative means for G0) I am actually choosing between G3 and G4!

Settling goals

Not all the goals motivating our actions are the goals we start from, the original “motives” of our activation, planning, deciding, and so forth, not only because some of the original goals (or parts of them) might be given up but because new goals necessarily enter into the decision not just as simple “positive” or “negative” predicted outcomes, but as being crucial for the choice.

In fact, we may start from a given motive G1 (going to Naples) and, by a *top-down* reasoning process, find/build its possible means and paths, but we may learn that we have some alternative means: we can either choose action A1 (go by train) or action A2 (go by car). How do we choose? We must start a reverse, *bottom-up* process by activating new goals that originally did not appeal to us; more precisely, we must construct a prediction scenario where we consider the possible further outcomes (in addition to the achievement of G1) of A1 and A2, and we must evaluate those outcomes against the *new* goals: for instance, G2 (reading during a trip) or G3 (arriving at a meeting without having to take buses or taxis). Those goals will serve to settle the matter of what it means to prefer (hence the label *settling goals*), since, as for G1 (going to Naples), A1 and A2 are equivalent; we will choose between those goals, not really between the two actions. Thus, the chosen goal (which makes a given action preferable) is a “necessary” outcome for it and is hence a “motivating” goal in a weak sense (it is certainly not “sufficient” in that the whole process is implied).

Preference between G1 and G2 can be reason-based, argued with pros and cons, or affective, based on the evocation of attractive or repulsive responses to two scenarios: a conflict between associated somatic markers (Damasio 1996). In that case, too, we can actively search for affective appraisal, evoking or imagining different affective experiences in the two cases.

1.4.2 Subjective or Psychological Conflict in the Strict Sense

Of course, subjectively speaking, there are different kinds as well as “strengths” of conflicts that make choosing a more or less difficult and demanding process. More precisely:

We can view conflict in a weak, potential, *broad* sense or in a *narrow* and specific sense of “being *in* conflict,” “experiencing conflict.”

Any decision in fact presupposes the perspective of more than one scenario, of some “alternative”: to do or not to do, or to do either A or B. We examine/consider pros and cons (at least the “costs”), or different ways to achieve the higher goal, and we must *choose*. However, this is just a *possible* conflict; it does not necessarily give rise to a “subjective experience of conflict.”

A felt psychological conflict is a special kind of internal conflict⁵:

- between *active goals*;
- among rather important goals (high value, with important consequences and concerns);
- involving a *difficult choice* resulting from a balanced comparison of the goals' values;
- producing some level of *anxiety*.

To be *in conflict* or to experience a conflict, I must have and perceive some difficulty that must be resolved; I feel “split” between the two perspectives, for the following possible reasons:

- Because there is too much *uncertainty* or lack of crucial information, and so I feel anxiety and worry about making the wrong decision;
- Or because the value of the “sacrificed” goal is too high; as stated earlier, any decision in a conflict implies a perceived “renunciation,” that is, a *loss*, a suffering;
- Or the perceived risk in case of failure is too high, in that potential harm is very high or its probability is high (subjective risk of failure).

I also worry about possible future “regrets,” repentance, and self-reproach, and I can add this to my mental accounting.

The top level of conflict is when my decision is suspended, blocked; I am unable to decide. Either I suspend my decision or I give up trying to decide, for example, by delegating the decision to somebody else or leaving it to chance.

After a decision is made, the situation (mental focus) is transformed: we set aside the previous alternatives and just focus on the chosen goal, which must dictate our actions and monitor the results, and on the implementative subgoals and executive action schemas. Only when we reach the outcome (especially if some surprise occurs owing to a mismatch with our expectations) may we reconsider the predicted costs and possible alternatives, with affective reactions of disappointment for the result or of regret for our bad decision. In a sense the conflict can reemerge, although we are impotent, but we can learn for the next time.

Moreover, following a choice (and to stabilize our behavior and mind) we tend to emphasize why our preference is the best one.⁶ In fact, if we realize that we are acting in a way that is inconsistent with our beliefs/opinions, when we notice such a cognitive-behavioral dissonance and inconsistency, we tend to rationalize our

⁵It might be interesting to note in passing that internal conflicts also apply to abstract agents like, for example, groups, organizations, or states in which conflicts among members can implement/generate an internal conflict at the level of the abstract agent. If there is an internal conflict in an abstract agent, there should be either an internal conflict in at least one of its members or an interagent conflict among some of its members.

⁶And after we have invested in that perspective, we increase the value of the goal [see sunk costs effects in, for example, Arkes and Ayton (1999)] to make it more stable.

behavior as a well-grounded choice, rather than change our behavior. Paradoxically, action adjusts cognition to its needs: instead of being the result of our beliefs and goals, our actions become their cause [Festinger's (1957) cognitive dissonance effect].

1.5 Social, External, or *Interagent* Conflicts: Their Cognitive and Structural Foundation

1.5.1 *Sociality and Prosociality*

“Sociality” is not a synonym for “cooperation” or “pro-social behavior”; it is frequently misused and misunderstood. Conflict is a basic form of sociality.

Competition and hostility are two of the sources of social relations, actions, and minds and are interesting for social or evolutionary theory. Cooperation is just one way of improving the performance of a cluster of systems living in a common world. Conflicts are not important just for conflict resolution or for (sometimes) improving groups' and organizations' results; they are important in and of themselves.

1.5.2 *Definition and Kinds*

I call social or interpersonal conflicts those conflicts that arise between two agents with their own goals; they do not necessarily occur between two “persons” (as persons): such conflicts can be between, for example, two groups or teams, two companies, two states, or two institutions, as well as between two “roles” or between two conventions or norms.

Any “entity” that is defined in terms of certain goals and functions in terms of goals and is aimed at realizing those goals can be in conflict with other analogous entities.

For a social conflict we need not just two incompatible goals but two *subjects* guided by those goals: the conflict is no longer between the goals; it is *also* between the subjects.

However, clearly, any conflict between subjects A and B presupposes a conflict between a goal of A and a goal of B. Conflicts are always between goals.

For social conflicts, too, it is crucial to make the distinction shown in Fig 1.1, that of intrinsic vs. extrinsic conflict; however, also very important is the intersection of other dimensions (Fig. 1.4):

Objective conflicts. A social conflict can just be there “objectively”: to have a conflict, it is not necessary that one or both subjects be aware of it. As stated earlier, the two subjects can compete with each other and thwart or harm each other

	OBJECTIVE	NON-OBJECTIVE
SUBJECTIVE / BELIEVED	KNOWN / CONSCIOUS	FALSE / UNGROUNDED
NON-BELIEVED	IGNORED / ACTED	

Fig. 1.4

without understanding that they are doing so or even intending to do so: we do not understand or aim at all the outcomes or conditions of our purposive behavior.

Moreover, we should distinguish between the following types of conflict:

Same-level conflicts:

(Goal X P) and (Goal Y (Not P))

Impinging conflicts (or *influence* conflicts):

(Goal X (Goal Y P)) and (Goal Y (Not P))

a goal concerning the mind of the other agent.

Same-level conflicts evolve into *influence conflicts*: some agents will try to change the minds of other agents.

Conflicts can be bilateral, reciprocal, or mutual, not only objectively but subjectively speaking.

By bilateral I mean that X against Y and Y against X have a subjective conflict (each believes himself to be in conflict with a goal of the other, not necessarily about the same goal).

Reciprocal is a shared knowledge/awareness (or just a shared belief) about a given conflict:

Goal X P and Know X ((Goal Y Q) and (Q > Not P))

and

Goal Y Q and Know Y ((Goal X P) and (P > Not Q))

Hence the notion of “noxious” goals is derived: X’s goal P comes to be perceived as noxious to Y’s goal Q, and vice versa.

Goal X (Not REALIZE Y Goal Y Q))

and

Goal Y (Not REALIZE X Goal X P))

In a mutual conflict I would add the following:

- Both X and Y know about the other’s respective beliefs and goals.
- They both have their respective goals because the other has such a noxious goal.

But, of course, the analysis and typology can be more sophisticated.

Symmetric and *asymmetric* conflicts also seem to exist. The incompatibility between the two states of the world (P and Q) does not seem to be bidirectional in all conflicts. As a case of asymmetric conflict, consider, for example, X's goal P of cleaning the window of some shopping center in order to be paid by the owner, and the goal Q of Y (a "black bloc" in a political demonstration) of breaking that window. If X pursues and realizes his goal P, this will not prevent Y from pursuing and realizing her goal Q, whereas if Y realizes her goal of breaking the window, X will not be in a position to achieve P. Of course, a certain time order is presupposed; in fact, the conflict is not necessarily between simultaneous, parallel actions. Also, a sequence of actions can create conflict if the results of an action destroy the conditions for the performance or the effect of a subsequent action. This relation is not necessarily bilateral or symmetric.⁷

1.5.3 Full Social Conflict: Hostility

A *full* social conflict arises when there is a subjective awareness of a competitive situation. Awareness, of course, may be *unilateral*, *bilateral*, or *mutual*.

Unilateral means that X is willing and ready to act against Y, while Y does not have a symmetric attitude toward X because, for example, Y is not aware of the conflict situation, she disagrees about the presence of a conflict (different beliefs), she does not have the ascribed goals, or she has different priorities and does not want to fight with X. As we saw, bilateral means that X sees a conflict with Y and Y sees a conflict with X. It may or may not be the same conflict; in this case, they converge on this representation, but they do not necessarily know the other's viewpoint or the fact that the other knows about their view. This would characterize a mutual conflict.

Subjective conflicts (the awareness of incompatible goals) lead to hostility or aggression, and then (if the new goal prevails) to a higher level of conflict: a fight.

A theory of conflict presupposes and requires a theory of *basic social attitudes*:

In *adoption* (1995) one agent adopts the goal of another agent (i.e., she pursues it as her own goal) because she believes that it is the other's goal and in order to make the other agent achieve it. In *hostility*, on the contrary, one agent has the goal that another agent does *not* fulfill or achieve some of his goals.

Generalized hostility is the opposite of benevolence, which represents an attitude of favoring others, the disposition to adopt others' goals, to help or exchange or cooperate.

⁷This can also hold at the individual level: a conflict between the goal of X that P and action A1 for P, and another goal and action of X: A2. It may be that if X performs A1 before A2, A2 cannot be successful; the plan is wrong. But if X performs A2 before A1, there is no problem; they are in the right order. The conflict is due to the temporal order (conflict in planning).

Hostility is a quite unavoidable consequence of social conflict: if agent X believes that there is a conflict with the goal of agent Y, she will not just have an opposite goal; she will have *the goal that the other does not achieve his goal*.

When each agent actively pursues his or her goal, trying to prevent the other from achieving his or her goal or to damage the other (aggressive move), there is a new level of conflict: *fighting*.

Y has the goal of preventing or thwarting the aggressive action/goal of X: her goal is the aggressive goal that X fails (defensive move), and so on.

1.5.4 Epistemic and Social Conflicts

Epistemic conflicts are likely to generate new social conflicts in yet another way:

Epistemic conflicts (even among cooperating agents):

(Goal X p) and (Goal Y p) *goal agreement*

(Bel X (q > p)) and (Bel Y (q > (Not p))): *belief conflict*

(Goal X q) and (Goal Y (Not q)): *conflict*

Belief conflicts can generate goal conflicts.

Conversely, epistemic conflicts can generate cooperation, goal agreement, among people with conflicting goals, for example, between enemies:

if

goal conflict (Goal X p) and (Goal Y (Not p))

belief conflict (Bel X (q > p)) and (Bel Y (q > Not p))

→ *goal agreement* (Goal X q) and (Goal Y q)

In this form of cooperation, one of the two agents is wrong and is acting in a self-defeating way.

Of course, goal social conflicts do not presuppose or necessarily imply epistemic social conflicts: agents can be in a social conflict while being in perfect belief agreement: they know the same things but they want different things.

Two agents in an epistemic social conflict are not necessarily in a goal social conflict. However, as we have seen, a simple epistemic conflict can generate a goal conflict.

1.5.5 Individual Selfishness and Social Conflicts

In our analysis, *social conflicts are due not to the selfish attitudes of social actors*, that is, to the fact that the actors caring only about their own goals and welfare. On the contrary this is an implicit premise of social conflict theories or an explicit claim as in Randall Collins' theory (Collins 1974).

Also, in that sense there is no opposition between prosocial attitudes (= cooperation) and conflicts. A conflict can consist of prosocial attitudes, not only because, for example, I can fight with Z in order to help or protect Y, but my coming into conflict with you can in itself be in your interest and for your benefit, as in the case of so-called tutorial conflicts (Sect. 1.2.6). Moreover, frequently a good collaboration based on a certain amount of autonomy requires some conflict in order to find the right solution or level of delegation (Castelfranchi and Falcone 2000).

The problem is not in being “selfish” but in pursuing some goal (being goal-directed), even an altruistic or prosocial one. I care about this and must achieve this; therefore, I can enter into a conflict with other people pursuing some incompatible goal (selfish or not). One should not confuse an autonomous actor, guided by his own internal goals, with a selfish one; pursuing one’s own goals does not mean pursuing them solely for one’s own benefit or pleasure.

It is very misleading to see social conflicts as being due only to the selfishness of individuals and perhaps cooperation as being due to prosocial attitudes or feelings. We may have cooperation, exchange, and collaboration among selfish actors for their private interests (as in Adam Smith’s theory of how markets function).

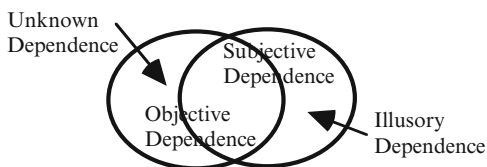
Even a basic conflict (between goals) is not due to the selfishness of goals (Bargh et al. 2008).⁸ Goals are not selfish and do not inherently and necessarily involve a competition of *omnes contra omnes*, as if behavior could be regulated by just one single goal. Goals can be contradictory or practically incompatible, but at the same time they can also be active and perfectly compatible (though independent and noninstrumental to each other) and pursued by the same action or plan. Moreover, *alliances between goals* can be formed to prevail over competitors. If I must choose between action A1 and action A2 or goal G1 and goal G2, I will try to look at possible outcomes, not just the original motives. I will look for additional advantages or costs and evaluate the different scenarios against additional goals that I create. The winner is the multigoal evaluated outcome whose combined value (the sum of the values of the realized goals minus the value of those that were thwarted or relinquished) is greater, that is, not a single goal but a set of goals versus another set of goals.

1.6 Conflict or Cooperation?

Conflict and cooperation, as stated earlier, are two complementary faces of sociality.

⁸For a criticism of this thesis see: https://www.academia.edu/5408854/Why_Goals_are_not_selfish_as_Barghs_and_colleagues_claim_

Fig. 1.5 The mind on dependence relations



1.6.1 Who Comes First?

Who is right, Hobbes or Durkheim? Is it “*bellum omnium contra omnes*” or *cohesion and solidarity*?

Conflict and cooperation have the *same foundation!*

The real structural basis and origin of sociality is dependence and power, but dependence and power *presuppose goals* and are “mentally grounded.” They depend on the minds of agents, not only because they derive from (different) goals and competences/skills, but because the agents’ *knowledge* about them is also crucial; that is the *cognitive emergence* of the conditions for conflict or cooperation.

Social dependence results from being in a common world, that is, from *interference* (my actions can facilitate or prevent your goal achievement, or vice versa).

X depends on Y as for a given action/resource (a) of Y relatively/for a given goal (that p) (Sichman et al. 1998).

Dependence is first of all an *objective social relation*: the combination of a lack of power of one agent (relative to one of her goals) and of the corresponding power of the other agent. But, of course, *subjective dependence* is also crucial (Fig. 1.5).

The dependence network *determines* and *predicts* partnership and coalition formation, cooperation, and exchange, as well as competition and conflicts; in addition, it affects the functional structure of organizations, rational and effective communication, negotiation power, and power over others, among other aspects.

Given our interdependence, we can compete and fight or we can exchange and cooperate. Both directions (*solidarity* and *homo homini lupus*) emerge spontaneously on a structural basis and are later orchestrated and organize social action and society. Both Hobbes and Durkheim are right.

If by war we simply mean the competition for scarce resources or incompatible outcomes, based on a common world, that is, on interference, there is no alternative and it is not our choice and decision. Interference and possible *competition* of all against all are objective, unavoidable, and emergent. They are intrinsic to the purposive nature (e.g., goals, outcomes, resources, conditions) of our behavior and to the different conditions we find ourselves in (e.g., skills, resources).

If X simply has the goal that P, but in order to achieve P she must use resource R or achieve or maintain the condition that Q, but in the same world (interference) Y exists, who has the goal that W (or that Not P, or W that implies Not P) or needs to exploit R, which is not enough for both, or needs to eliminate or avoid that Q, perhaps they will ignore each other, but they will actually fight one other, they will *compete* for R or for (Not) Q. Either X will win, that is, she will simply realize her goal,

or Y will. They objectively and unknowingly are interacting socially (in an ahostile way), but not subjectively: their actions are not social actions; they just have unwitting social effects. However, if there is selection or learning, then the *function* of that action is social (their success and adaptive character are due to the harm to the other).

Let us call competition that *unintended and objective* relation and situation where there is a potential unintended *bellum omnium contra omnes*.

Let us call on the contrary a fight or war a situation where the competition becomes explicit, and in the minds of the agents, the action becomes truly social, and harming the other is intended and even motivating. X knows about the competition, the objective negative interference of/with Y; thus, not only does she perform her actions calculating the harm to Y and adapting them to anticipate Y or to prevent Y's countermoves, but she might add to her plan specific actions against Y that would not be part of the plan just for P but are there in order to block or harm the other. This is a real *war* consisting of intentional (anti)social actions (hostility, as stated in Sect. 1.5.).

Note that the goal of harming the other is not an end per se, a motive, but is rather simply instrumental to P.⁹ Of course, there might be intrinsically hostile or aggressive goals (motives): X is motivated to kill or harm Y. Can one say that *homines homini lupi* in this sense? I do not think so, except perhaps against the *barbaroi* (out of diffidence or fear), the out-group people, or for terminalization (a subgoal that ultimately becomes an end goal) because of learning or culture: the goal of competing, previously instrumental, becomes an end in itself.

In this view, the main function of prosocial behavior or positive sociality is *the multiplication of the power* of the participating agents.¹⁰ Unlike Huberman and Hogg (1994), we do not assume that the greatest advantage of (cooperative) sociality is that it speeds up the search for solutions to *common* problems or leads to better solutions to those problems, but rather it *multiplies individual powers*: any agent, while remaining limited in its capabilities, skills, and resources, finds the number of goals it can pursue and achieve increased by virtue of its "use" of others' skills and resources. In a sense, any agent's limitations with regard to power and its differences from others in the kind of power it is endowed with become an advantage (Durkheim's perspective): although not omnipotent, *the agent is allowed to overcome its cognitive and practical limits through sociality*.

However, we can not only exploit others' powers by cooperation and exchange; we can try to obtain what we want/need by acting aggressively, harming others, or fighting with them. Moreover, we can not only derive benefits from others' powers and actions; we can be damaged or exploited by them, and we need to block/prevent their interference.

⁹Thus, perhaps *bellum omnium contra omnes* is not really synonymous with *homo homini lupus*.

¹⁰It seems that the less the level of individual self-sufficiency (the number of self-realizable goals out of the number of needs) the more sociality is useful and can multiply powers. (But the function is complex because we need agents with a high "power of" – skills, resources – and low "self-sufficiency"). In other words, the more individuals are dependent on each other, the more sociality multiplies their power. This is one reason why division of labor and specialization are so productive.

1.6.2 *Cooperation for War and War for Cooperation*

Not only are there two possible parallel directions in social emerging behavior from the structural layer of dependence (cooperation or competition), but they can coexist and emerge one from the other: the process is recursive and applies to new layers.

In fact, given the emergence of a cooperative or exchange social network in a given community, on top of these exchange relations we can develop a new, higher, level of competition and contests: competing to be partners with and to exclude the other. Given a market, we start to fight with each other in marketing, for our relational capital, for better positioning in the market: new goods (market positioning), new competition. On top of a network of competition and conflict we can build a new cooperative layer: given the intended fight, conflict, or war with Y, we might have common interests or the possibility of exchanging with Z against Y and building an “alliance” for war.

Some forms of conflict imply/require/presuppose cooperation. Some forms of cooperation require/presuppose conflict.

For example, economic competition (quite fierce and dishonest) actually implies cooperation at the *functional* level; those in conflict cooperate in providing, for example, the right selection, innovation, or value independently of the aims of the competing agents. There is no intentional cooperation, just the *invisible hand*.

But sometimes there is also intentional cooperation, at least regarding the rules or identifying and marginalizing dishonest competitors; this can be cooperation for innovation and knowledge exchange. For example, to compete in sports or games, we agree to play by certain rules, to follow a kind of script. Even in war, there can be some agreement (regarding, for example, prisoners or chemical weapons).

Of course, there is also cooperation without conflict, just adjustments, coordination, and some misunderstandings: fully shared goals and subgoals.

There are also conflicts without any shared rules, where everything is allowed and perhaps the adoption of the goal/expectation of the other is involuntary/false/apparent (as in an exchange of insults or the suffering of a sadist’s victim).

1.7 Subjective Conflict and Social Conflicts

Do External Conflicts Require Internalized/Mentalized Conflicts?

Sometimes, to be successful, an external conflict requires its internal formulation and the goal of beating the other.

1.7.1 *Mentalized or Internalized and Subjective Social Conflicts*

Suppose that at least on X's side the conflict with Y is explicit and conscious.

Goal X P and Know X ((Goal Y Q) and (Q > Not P))

Then, as stated earlier, the conflict can become intentional: *counteradoption*, hostility.

The problem is that if we are just acting out our conflict without awareness, without understanding the reciprocal opposition, we probably are not acting effectively: we will badly predict, anticipate the other's behavior and attempts to block us; we will not have the goal of harming him in general, not just by anticipating him.

"Mind reading" has been an essential foundation of human sociality; it is crucial not only for coordination and cooperation but also for competing and fighting. As stated earlier, sociality does not mean prosociality or cooperation. How can I compete with you if I cannot anticipate your actions; and how can I harm you if I do not know what you care about and how much you care about it?

The advantage of subjective/mentalized social conflicts is that I can plan (as a subplan) to harm my opponent in some way in order to prevent him from realizing his goal which is conflicts with my interests. For example, not only can I adjust my own running speed based on my opponent's in an attempt to finish first, but I can also try to trip him up.¹¹

Of course, as with private conflicts, subjective social conflicts can be illusory, based on wrong beliefs. Your and my goals may not really be incompatible, but since we think they are, we compete or fight with each other. Thus, what really matters is

Goal X P and Bel X ((Goal Y Q) and (Q > Not P)).

Of course, it makes a difference if it is true that (Goal Y Q) or not; we might harbor a sort of useless hostility toward one another and even display aggression and fight (Sect. 1.5.4).

As stated earlier, if our goal consists in ensuring another agent Y's does not realize his goal, that is hostility, and it involves being against or angry/upset with Y. There are different degrees and kinds (reasons) of hostility. For example, "hate" seems to imply that whatever goal Y might have, my goal is to prevent Y's goal from being realized or even that what Y already has been lost. I am a true enemy of Y.

Other forms of hostility might revolve around specific goals of Y (such as, for example, out of jealousy), specific competitions, or some unspecified goal (as in envy: I wish that something good or important to you goes wrong).

¹¹The disadvantage is that I can no longer harm others or pursue my selfish interests and desires without caring about others without, for example, valuing their needs or without feeling guilty.

A social conflict is particularly important if it is about motives or so-called terminal goals (like values), not just about subgoals or means. In fact, the means might be replaceable; one might find other ways of achieving a goal. Thus, means/goals are more negotiable. Of course, final goals, also terminal, can be renounced (globally or partially; see Sect. 1.8.3) for a greater value, although some of them are nonrenounceable and, thus, nonnegotiable.

1.7.2 Do External Conflicts Require Internalized/Mentalized Conflicts?

Sometimes, to be successful, an external conflict *requires* internal formulation and the goal of beating the other.

External conflicts might require some explicitness and the representation of that relation in the mind of the competitors; however, *rarely do social conflicts between two subjects require/create some internal, personal/private, psychological conflict.*

Only if X is aware of the conflict between her goal that p and the goal of Y (goal Y q) might she, as a consequence of the external conflict, generate an internal conflict.

For example, is in a tutorial relationship, cares or must care about Y's welfare; thus, she has or should have the goal that Y realizes his goals (including that q), but this goal is in conflict with her personal goal that p (since p and q are incompatible); or she might even formulate the goal that Y does not realize q. In either case, she now has an internal conflict between two of her own goals.

Not all social conflicts imply a strong subjective conflict, *of feeling in conflict*; they may just imply – as in all intentional actions – some decision. I can create a very fierce conflict against Y without feeling any conflict about it within myself. I'm fully determined, happy, convinced, and coherent.

1.8 Resolving Conflicts

1.8.1 Resolving Internal Conflicts

The elimination or resolution of an internal conflict can be accomplished in at least three ways:

- (a) Choose among the goals by “promoting” one of them to the next level of goal processing.
- (b) Kill or drop at least one of two competing goals.
- (c) Eliminate a critical belief if it causes an indirect conflict, for example, changing the world in order to make resources available.

For psychological conflicts:

- (d) Remove the conflict from one's awareness (attention or accessible memory) or eliminate the meta-belief that there is a conflict: a "self-deceptive" solution.

1.8.2 Resolving Social Conflicts

The elimination or resolution of social conflicts can be accomplished in one of only three ways:

- (a) Eliminate (kill) at least one of the two agents from that world through
- the migration or expulsion of an agent from the "common world"
 - physical suppression of that agent (killing and war).

- (b) Change the mind of one of the agents.

If one of the agents abandons or drops her goal, there will no longer be a conflict.

This goal dropping can be actively pursued by the other agent, which could try to (overtly or covertly) kill the other's goal.

This can be done in two different ways:

- (b1) By eliminating some support of or reason for the target goal: goals are dropped, for example, out of a lack of motivation (the end goal for which they are instrumental), because of a lack of possible plans or means, or by the invalidation of triggering beliefs. (Castelfranchi and Paglieri 2007).
- (b2) By conflict internalization; this case is, in my view, the most frequent and prototypical and serves as the basis of true negotiation.

Conflict internalization

Let us focus on conflict internalization. The general law of conflict resolution among agents is as follows:

1. An interagent conflict generates (is transformed into) an intraagent conflict in at least one of the opponents (internalization process);
2. Internal (mental) resolution of the intraagent conflict results in the resolution of the original interagent conflict.

Internalization can work via goal adoption:

X's influencing Y occurs via Y's understanding (normally communication) of X's goal about Y's mind, and the adoption of X's goal by Y.

Following internalization there is an *internal conflict* in Y:

(Goal Y (Not (Do Y a))) and (Goal Y (Do Y a))

Example:

Mother's goal: child does homework

Child's goal: watch TV (not do homework)

Mother: "if you finish your homework, I will take you to the movies"

Child's internal conflict: “do homework so I can go to the movies” or “watch TV (not do homework).”

If the mother is right and is skilled at reading her child's mind, and thus the competitor goal she activates (going to the movies) is for the child of greater value than doing homework, then the child will resolve his *internal* conflict in the right direction, and the *social* conflict with his mother will be automatically resolved; this is why the mother has induced a conflict *inside* her child.

1.8.3 *Compromises*

For a compromise, we need opposite goals that are “gradable” or “composite,” that is, achievable in part. So one agent can give up part of her goal if the other agent gives up part of his goal (the incompatible part or degree),

Or one must compensate: X gives away her goal G1 (or part of it) and receives, as a *compensation*, a reward, that is, the realization of another goal not in the original conflict.

Some sort of exchange:

X gives something to Y, and there is *reciprocation*.

A kind of solution that in fact dissolves the conflict involves

- changing the beliefs or conditions that make the two goals incompatible: if both goals can be achieved, there is no longer a conflict!

1.8.4 *Irresolvable Conflicts*

Not all conflicts can be resolved, not only practically (for utility or compromise) but *in principle*.

Some conflicts cannot be resolved:

- (a) Either because some goal of the actors is *nonrenounceable and nonnegotiable*; it cannot be reduced or replaced or compensated, like some value or principle (e.g., “This country is ‘our’ country, homeland, identity territory; we cannot renounce it!”; “God will it so; we cannot discuss or violate God’s will”),
- (b) Or because there are meta-goals regarding the conflict or *functions of the conflict that exist independently of the contents*.
 - It might be that X is looking for a conflict with Y because X feels that it is necessary or has a need to define herself in oppositional terms or because she must show hostility toward Y for some reason. What matters is being in conflict with the other.

- It might be that being in conflict is structural to the relation between X and Y; they are competitors, they must beat each other, X's goal is to prevail over Y, for example.

When conflict is an end and no longer a means, or when a conflict is due to nonrenounceable and nonnegotiable values, no resolution or exit is possible.

1.9 In Praise of Conflicts

Prejudice: *We have a conflict with conflicts!* We harbor an attitude or feeling of avoidance toward conflict.

For many people, the ideal psychological and social condition would be the absence of conflicts. This is a strange perception and conception.

Actually, conflicts are a fundamental “motor” or “leaven” of *individual* and *social* growth. This has been clear since Coser’s sociological theory of conflicts but does not affect our common sense. Conflicts are not necessarily a cause or a sign of instability, and they have several crucial functions, such as, for example, stimulating innovation and social change, reinforcing the unity and identity of a group against other groups, reinforcing a central authority, and clarifying diverge positions (Coser 1956).

Let us consider some cases and aspects of conflicts.

1.9.1 *The Functions of Conflicts*

Why do we need not only competition (both individually and societally) but some sort of (explicit) conflict and fighting?

We are even biologically and psychologically designed for conflicts; fighting (and beating, harming, or humiliating somebody) excites and motivates us, individually and collectively. So-called sports fill that role; but competition in politics, the arts (e.g., festivals), and beauty attract and excite people.

We should strive to avoid wars but not conflicts, competition, or even fights.

Since we need to be better than (some) others (in the same domain or community), we are automatically in conflict over that.

1.9.1.1 Social (r)evolution

Conflicts are the engine of change and possibly of progress, at both the individual level (e.g., adolescents’ or females’ emancipation conflicts against parents or traditional costumes), and the collective level (e.g., new rights, better work conditions, social welfare, and greater equality are due to social conflicts, possibly

with the aggregation of *collective subjects* or actors, including classes, groups, and movements) since any given change affects some established interest or power, some common sense or prejudice. To change, we must fight.

The individual fights in order to defend his rights or against abuses or to gain new rights. At the collective level, groups and classes associate and cooperate against some adversary group or class in order to acquire new powers and rights.

“We should reevaluate conflict, since without conflicts there is no social justice” (Pierre Carniti, Italian trade unionist).

1.9.1.2 Emancipation and Empowerment

There are powers that *in principle* cannot be given, where real empowerment is not a received, subordinated, passive process or position.

Some powers can only be *actively acquired* by an individual: for example, skills, control over emotions, self-confidence, sense of autonomy, the capability of reclaiming rights. Some powers must necessarily be grabbed, taken in opposition; they require fighting for a subtraction of power, for seizing power from and against Y.

This power *cannot* be given/granted because the meaning of the act would be self-defeating, contradictory. There would be another result.

For example, the power to violate norms, the power to stand up for oneself, to not be submissive and passive. This applies to emancipation movements for respect and dignity, against stigmas, antidiscrimination, for pride. And it also is the case of the physiological journey of adolescents “against” their parents. In fact, *conflicts can be educational*, aimed at growth and self-esteem; they educate one about conflict, autonomy, and rights. Parents, educators, mentors, and others are faced with the difficult task of sustaining and even eliciting conflicts (against paternalism and similar attitudes).

In general, *power over Y* and the *power to influence or manipulate Y* may in and of itself elicit Y’s opposition, a conflicting attitude, a tendency to exit from that situation or to rebel. This is a natural and not necessarily bad impulse; order is not always good in itself.

1.9.1.3 Science

Science and research are *based on conflict*: criticism, falsification, opposition, and fights over different theses, not (only) for personal motives, prestige, gain, or other reasons but as a rule of the game, that is, for a better collective result, for a sort of conflicting cooperation, for stronger and validated theories.

1.9.1.4 Conflicts: The Presupposition of Democracy

No conflicts, no democracy. Democracy is not only a “response” to conflicts and aimed at moderating them; it would and should be a way of encouraging, expanding, and resolving them.

Conflicts are not just to be governed, reduced, or reconciled; they should even be *promoted*, and this is in fact the role and function of specific forces and organizations, such as, for example, trade unions, parties, special interest groups, associations, and movements, the crucial stakeholders in a democracy who are responsible for the typical social, cultural, and economic progress of western countries in recent centuries and now of the rest of the world.

Of course, conflicts can be dangerous, leading us to combat, violence, or war. It is true that societies and groups need rules to govern them, to avoid degeneration. The centralized state was one of these solutions: the state monopolizes violence; private or group violence is forbidden.

Democracy is also a solution to conflicts; more precisely, as stated earlier, it presupposes and needs conflicts and represents a way to make them useful and progressive; it provides space (demonstrations, parties, parliaments, rallies), a voice, roles, and rules for expressing conflicts, not necessarily because conflicts are reconcilable; some agreement is possible through argumentation and persuasion.

Conflicts are not just conflicts of points of view or opinions or result from different conceptions, information, or ways of reasoning. There are conflicts of interests: if you realize your goal, I cannot realize my goal or lose something I have. So the problem has to do with conflicts between groups or classes or conflicts between private interests as opposed to common interests, the so-called commons and public goods.

Democracy is not just a forum for discussion; it has rules for prevailing over others, for changing society in favor of or against the interests of some group however with shared norms and values (constitution).

Social conflicts in fact *do not have* a verbal/cognitive or a so-called technical solution that is based on data and technical principles; they have political solutions; they are a matter of power and of *prevailing interests and compromises* (equilibrium, partitions/shares).

Conflicts are thus the motor and principle of democracy and of its possible effectiveness in changing society *in favor of subjugated subjects, disadvantaged classes and groups*, etc. Up with conflicts! *Dissent is the highest form of patriotism* (Thomas Jefferson).

When you find yourself on the side of the majority, it is time to stop and reconsider.

Mark Twain

1.10 Concluding Remarks

My aim in this chapter was to give some hints about the *complexity* of the issue and its *foundations*:

- Some of the basic aspects/notions of conflict theory requiring formalization,
- Relating the notion of conflict to goals (and to their processing) but disentangling it from selfishness,
- Objective and subjective aspects of social conflicts and their relationships,
- Why and how cognition is relevant in resolving social conflicts.

I also tried to draw attention to our presumed and rather unconscious prejudice against conflicts. Is or should human society be socially cohesive and demonstrate solidarity or cooperation (Durkheim)? Or is society a matter of *bellum omnium contra omnes*? We explained why conflicts are a necessary expression of one and the same foundation.

Conflicts are not just an obstacle, a breakdown, a malfunctioning of the mind, of interpersonal relations, of society, of political power. Conflicts are physiological and useful; they help to make society and individuals *dynamic* and even to reinforce stability (the functions of conflicts).

The problem is: *Which* conflicts are dysfunctional and *when* are they such?

References

- Arkes HR, Ayton P (1999) The sunk cost and concorde effects: are humans less rational than lower animals? *Psychol Bull* 125(5):591–600
- Bargh JA, Green M, Fitzsimons G (2008) The selfish goal: unintended consequences of intended goal pursuits. *Soc Cogn* 26(5):534–554, <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2599951/>
- Bourdieu P (1980) *Questions de sociologie*. Minuit, Paris
- Castelfranchi C (1996) Conflict ontology. In: *Proceedings of ECAI'97 workshop on conflicts*. Budapest
- Castelfranchi C, Falcone R (2000) Conflicts within and for collaboration. In: Tessier C, Chaudron L, Muller HJ (eds) *Conflicting agents: conflict management in multi-agent systems*. Kluwer, Dordrecht, pp 33–61
- Castelfranchi C (2001) The theory of social functions. Challenges for multi-agent-based social simulation and multi-agent learning. *J Cogn Syst Res* 2:5–38, Elsevier. <http://www.cogsci.rpi.edu/~rsun/si-mal/article1.pdf>
- Castelfranchi C, Paglieri F (2007) The role of beliefs in goal dynamics: prolegomena to a constructive theory of intentions. *Synthese* 155(2):237–263
- Collins R (1974) *Conflict sociology*. Academic, New York, 1974
- Conte R, Castelfranchi C (1995) *Cognitive and social action*. UCL press, London
- Coser L (1956) *The functions of social conflict*
- Damasio AR (1996) *Descartes' error*. London
- Evans JS (2008) Dual-processing accounts of reasoning, judgment, and social cognition. *Annu Rev Psychol* 59:255–278
- Festinger L (1957) *A theory of cognitive dissonance*. Row & Peterson, Evanston, IL

- Hayek FA (1978) Competition as a discovery procedure. In: Hayek FA (ed) *New studies in philosophy, politics, economics, and the history of ideas*. University of Chicago, Chicago, IL, pp 179–190
- Huberman BA, Hogg T (1994) *Communities of practice: performance and evolution*. Xerox Palo Alto Research Center, Palo Alto, CA, pp 1–23
- Lewin K (1935) *A dynamic theory of personality*. McGraw-Hill, New York
- Sichman JS, Conte R, Castelfranchi C, Demazeau Y (1998) A social reasoning mechanism based on dependence networks. In: Hunhs M, Singh M (eds) *Readings in agents*. Morgan Kaufmann, San Francisco, CA, pp 416–421

Chapter 2

Group Conflict as Social Contradiction

Daniele Porello, Emanuele Bottazzi, and Roberta Ferrario

2.1 Introduction

This paper provides a number of fundamental elements in order to develop an ontologically grounded classification of group conflicts. Understanding groups' behavior is a challenging task that involves a number of disciplines such as game theory, sociology, and behavioral sciences. We are here interested in the perspective provided by logic and computational disciplines, in particular our methodology is related to the analysis of the interaction of a number of heterogeneous agents and groups that have been developed in the multiagent community (Woolridge 2008). The model of agency that is presupposed in this approach is the (belief-desire-intention) BDI model that allows for developing a mathematical representation of individual actions, plans, goals, etc. in their relationship with other agents (Bratman 1987; Georgeff et al. 1999).

We shall introduce the methodology of *social choice theory (SCT)* in order to formally grasp the relationship between the beliefs, desires, intentions, preferences, goals of the individuals belonging to the group and the corresponding attitudes that we may want to ascribe to the group itself. SCT is a branch of welfare economics that emerged at the beginning of the past century and that studies how a collective choice can be derived from individual possibly conflicting choices, by means of fair aggregation procedures (Taylor 2005; Gaertner 2006). SCT has been successfully applied in economics, political science, and recently in computer science and AI. In particular, in the area of multiagent systems, SCT has provided the key concepts

D. Porello (✉) • E. Bottazzi • R. Ferrario
Institute of Cognitive Sciences and Technologies, CNR, Trento, Italy
e-mail: daniele.porello@loa.istc.cnr.it; emanuele.bottazzi@loa.istc.cnr.it;
roberta.ferrario@loa.istc.cnr.it

for understanding and defining notions such as group information, group choice, and group intention. (Brandt et al. 2013).

SCT views groups as collectives of individuals that may have different preferences, opinions, desires, who have to decide and agree on a single collective stance. An example of such a group is given by a parliament in which the representatives may express a number of divergent positions and who settle the possible disagreement by voting. Moreover, the members of the board of stakeholders of a corporation who decide possible courses of action can be analyzed by means of SCT. More generally, any assembly of individuals that agree on the procedure to settle disagreement can be studied by means of social choice theoretic methods. An important difference that we want to stress is that SCT takes a different perspective on groups with respect to Game Theory (Neumann and Morgenstern 1944; Osborne and Rubinstein 1994). SCT is interested in the behavior of the group as a single entity, whereas the focus of game theory is on the interaction of a number of self-interested agents. In this sense, the notion of group that SCT defines and investigates imposes a strict form of social cohesion. Although SCT presupposes at least the implicit agreement on the procedure to settle disagreement, due to the variety of aggregation procedures that can be defined and discussed, SCT methods can deal with a wide spectrum of groups, such as parliaments, organizations, corporations, assemblies, and associations. SCT can be considered as a general theory of the aggregation of *propositional attitudes* in the philosophical sense, i.e. beliefs, desires, intentions, preferences, judgments. Once we are capable of modelling such attitudes in a clear formal language, we can define and evaluate the proper aggregation procedures by means of SCT techniques (Dietrich and List 2009). Moreover, SCT has been applied to model groups' intentions, for example in Boella et al. (2011).

It is important to make the level of our analysis explicit: We are interested in knowledge representation and in particular we propose a formal and general methodology to represent conflicts. We assume that conflicts are always about something. Thus, we shall introduce a formal language to represent possible matters of conflicts, such as preferences, beliefs, judgments, desires, goals, and intentions. We shall then propose an abstract notion of conflict between matters by using the formal concept of *contradiction* between the formal representations of the matter of conflict. For example, a conflict of opinions is represented by the contradiction between a proposition A and a proposition *not* A , an actual conflict that may emerge between two agents concerning their opinions can be described by assuming that an agent is claiming A whereas the other agent is claiming *not* A .

The motivation for using SCT in analyzing conflict is that it allows for singling out a peculiar notion of conflict of groups. Although it may seem at first that the agreement forced by SCT on the procedure to settle possible conflicts is sufficient for guaranteeing that any conflict among the members of the group can be settled, quite surprisingly, this is not the case. There are situations such that although the individuals that are members of the group agree on the norms or procedures that settle possible conflicts between individuals, nevertheless the actions, beliefs, judgments, preferences of the group turn out to be in a peculiar situation of conflict,

namely they turn out to be contradictory. Situations like these actually occurred in the deliberative practice of the US Supreme Court (Kornhauser and Sager 1993), and it is important to understand what type of conflict is peculiar to those situations. In such cases, being members of the court, the judges accepted to solve possible divergences by voting by majority. Hence, the procedure to settle possible conflicts of opinions was clearly accepted by every member of the group. Nevertheless, the outcome of the procedure turned out to be in contrast with some very basic principles of rationality. This is the specific type of conflict that our model aims to capture.

The remainder of this paper is organized as follows. In Sect. 2.2, we shall introduce some basic elements of SCT and present the cases of group conflict that we want to treat. In Sect. 2.3, we focus on judgment aggregation, a recent area in SCT, and we present it as a general theory for aggregating propositional attitudes. Judgment aggregation will provide the formal basis for presenting the peculiar notion of group conflict that we are going to analyze. In Sect. 2.4, we present the elements of our analysis of groups and conflicts. Section 2.5 capitalizes on the conceptual methodology of the previous sections and presents a taxonomy of group conflicts. Our conceptual analysis can be considered a preliminary step towards the integration of a taxonomy of conflict into a foundational ontology such as DOLCE (Masolo et al. 2003). Section 2.6 concludes and points at some future applications. In particular, we believe that our approach is particularly useful if implemented in complex socio-technical systems (Emery and Trist 1960), as conflicts may show up between various types of heterogeneous information, possibly originating both from humans and artificial devices. The abstract level of representation that we pursue in this paper can therefore deal with information coming from heterogeneous sources, thus it can be applied to model the rich informational entanglement that characterizes socio-technical systems.

2.2 SCT: An Informal Presentation

The seminal result in SCT has been provided by Kenneth Arrow's investigation of paradoxes in *preference aggregation*, namely the problem of aggregating a number of individual conflicting preferences into a social preference. Suppose that three parties in a parliament (label them 1, 2, and 3) have conflicting preferences over three possible alternative policies: *a*: "promote workers' salaries", *b*: "decrease entrepreneurs' taxation", and *c*: "increase unemployment benefits". Suppose agents' preferences can be represented by the following rankings of the options. Mathematically, preferences are assumed to be linear orders, thus individual preferences are supposed to be *transitive*: if an agent prefers *x* to *y* and *y* to *z*, then she/he should prefer *x* to *z*; *irreflexive*: an agent does not prefer *x* over *x*; and *complete*:

for any pair of alternatives, agents know how to rank them, x is preferred to y or y is preferred to x .¹ Profiles are lists of the divergent points of view of the three individuals, as in the following example:

- 1: $a > b > c$
- 2: $b > a > c$
- 3: $a > c > b$

In the scenario above, the agents have conflicting preferences and there is no agreement on which is the best policy to be implemented. Since the policies are alternative, 1 and 3 would pursue a , whereas 2 would pursue b . The example is supposed to model a parliament, thus the possible conflicts have to be solved, as we assume that the parliament as a whole should pursue one of the alternative policies. Thus, we have to ask what is the preference of the group, namely the preference that we can ascribe to the parliament composed by 1, 2, and 3. However, at this point, we cannot ascribe a single preference to the group without assuming a rule to settle disagreement. Suppose now that the individuals agree on a procedure to settle their differences; for example, they agree on voting by majority on pairs of options. Thus, agents elect the collective option by pairwise comparisons of alternatives. In our example, a over b gets two votes (by 1 and 3), b over c gets two votes (by 1 and 2), and a over c gets three votes. The majority rule defines then a social preference $a > b > c$ that can be ascribed to the group as the group preference.

The famous Condorcet's paradox shows that it is not always the case that individual preferences can be aggregated into a collective preference. Take the following example.

- 1: $a > b > c$
- 2: $b > c > a$
- 3: $c > a > b$

Suppose agents again vote by majority on pairwise comparisons. In this case, a is preferred to b because of 1 and 3, b is preferred to c because of 1 and 2, thus, by transitivity, a has to be preferred to c . However, by majority also c is preferred to a . Thus, the social preference is not "rational", according to our definition of rationality, as it violates transitivity.

Kenneth Arrow's famous impossibility theorem states that Condorcet's paradoxes are not an unfortunate case of majority aggregation, rather they may occur for any aggregation procedure that respects some intuitive fairness constraint (Arrow 1963). In the next section, we shall discuss in more detail the formal treatment of the intuitions concerning fairness and we shall define a number of properties that provide normative desiderata for the aggregation procedure.

¹These conditions are to be taken in a normative way. They are not of course descriptively adequate, as several results in behavioral game theory show. However, the point of this approach is to show that even when individuals are fully rational, i.e. they conform to the rationality criteria that we have just introduced, the aggregation of their preferences is problematic.

A recent branch of SCT, *Judgment Aggregation* (JA) (List and Puppe 2009) studies the aggregation of logically connected propositions provided by heterogeneous agents into collective information. The difference with preference aggregation is that in this case agents argue and provide reasons for their choices instead of simply reporting their preferences. For example, take a committee composed by three members, who have to decide whether to implement a policy B : “we should increase workers’ salaries” and the considerations that may support such conclusion, such as A : “low salaries cause crisis” and the material implication $A \rightarrow B$: “if low salaries cause crisis, then we should increase workers’ salaries.” Now suppose members hold different opinions, as follows

	A	$A \rightarrow B$	B
1	Yes	Yes	Yes
2	Yes	No	No
3	No	Yes	Yes

In this case, the conflict may emerge from the fact that individuals have divergent opinions on what is the best thing to do and no shared rule to settle such conflicts of opinions. If one asks what is the opinion of the group, one may simply answer that, due to divergencies, there is no group opinion. However, this is the claim of our paper, any statement concerning collective information depends on the procedure that is assumed to settle disagreement. If we don’t assume any procedure to solve conflicts, we simply say that individual conflicts may possibly arise, but we leave them as they are. If individuals agree that *unanimity* is the rule to elect a collective opinion, in the example above, neither A , B nor $A \rightarrow B$ is elected as the opinion of the group. If the *majority* rule is used, then the collective opinion is given by A (voted by 1 and 2), $A \rightarrow B$ (voted by 1 and 3), and B (voted by 1 and 3).

Analogously to the case of Condorcet’s paradox in preference aggregation, situations of inconsistent aggregations of judgments have been individuated. These paradoxical situations have been labelled in the literature *doctrinal paradoxes* or *discursive dilemmas*. It is important to notice that such paradoxical situations actually occurred in the deliberative practice of the US Supreme Court (Kornhauser and Sager 1993). This problem has been perceived as a serious threat to the legitimacy of group deliberation and it has been considered a seminal result in the recent debate on the rationality of democratic decisions (Pettit 2001; List and Pettit 2002).

We show an example of such paradox by slightly modifying the previous example. Suppose agent 3 rejects B because she/he rejects the premise A .

	A	$A \rightarrow B$	B
1	Yes	Yes	Yes
2	Yes	No	No
3	No	Yes	No

By majority, the group accepts A , because of 1 and 2, and $A \rightarrow B$, because of 1 and 3, but it rejects B . Thus, the group collectively accepts the premises of *modus ponens* while rejecting the consequence. If we assume that rejecting a proposition is

equivalent to accepting its negation $\neg B$, then, even if individual opinions are each logically coherent, the collective set A , $A \rightarrow B$, and $\neg B$ is inconsistent.

Again, doctrinal paradoxes apply to any aggregation procedure that respects some basic fairness desiderata, this is the meaning of the theorem proven by Christian List and Philip Pettit (2002). It is important to stress once again that the case of doctrinal paradoxes, far from being a curious example envisaged by means of some thought experiment, has actually occurred in the deliberative practice of judicial courts. In particular, the paradox has been perceived as a serious threat to the legitimacy of the decision of the Court by the judges of the Court themselves. A contradictory outcome, in that case, amounts to providing an inconsistent sentence that can be contested by the defendant who is being charged on that ground. Thus, it is important to provide a conceptual characterization of what type of conflict the doctrinal paradox exhibits, as the problem of understanding to which agent the conflict can be ascribed is not of immediate solution.

Summing up the content of this section, we have seen how SCT allows for individuating and formalizing an important form of group conflict that applies in normative settings and that is the specific notion of conflict that we want to analyze in this paper.

2.3 A Model of Judgment Aggregation

We present the main elements of the formal approach of judgment aggregation (JA). The reason why we focus on JA is twofold: on the one hand, it has been taken to be more general than preference aggregation (List and Puppe 2009), on the other hand, it has been claimed that JA can provide a general theory of aggregation of propositional attitudes (Dietrich and List 2009). Therefore, JA provides the proper level of abstraction for our abstract model of types of conflict. The content of this section is based on List and Puppe (2009) and Endriss et al. (2012) and builds upon them.

Let P be a set of propositional variables that represent the contents of the matter under discussion by a number of agents. The language L_P is the set of propositional formulas built from P by using the usual logical connectives \neg , \wedge , \vee , \rightarrow , \leftrightarrow .

Definition 2.1. An **agenda** is a finite nonempty set $\Phi \subseteq L_P$ that is closed under (non-double) negations if $A \in \Phi$ then $\neg A \in \Phi$.

An agenda is the set of propositions that are evaluated by the agent in a given situation. In the examples of the previous section, the agenda is given by A , B , $A \rightarrow B$, $\neg A$, $\neg B$, $\neg(A \rightarrow B)$. The fact that, given a proposition, the agenda must contain also its negation aims to model the fact that agents may approve or reject a given matter. The rejection of a matter A is then modeled by an agent accepting $\neg A$. In this model, for the sake of simplicity, we do not present the case of abstention, however it is possible to account for such cases by slightly generalizing our framework.

We define individual judgment sets as follows.

Definition 2.2. A **judgment set** J on an agenda Φ is a subset of the agenda $J \subseteq \Phi$.

We call a judgment set J **complete** if $A \in J$ or $\neg A \in J$, for all formulas A in the agenda Φ , and **consistent** if there exists an assignment that makes all formulas in J true, namely we assume the notion of consistency that is familiar from propositional logic.

These constraints model a notion of rationality of individuals, i.e. individuals express judgment sets that are rational in the sense that they respect the rules of (classical) logic.²

Denote with $J(\Phi)$ the set of all complete consistent subsets of the agenda Φ , namely $J(\Phi)$ denotes the set of all possible rational judgment sets on the agenda Φ . Given a set $N = \{1, \dots, n\}$ of *individuals*, denote with $\mathbf{J} = (J_1, \dots, J_n)$ a *profile* of judgment sets, one for each individual. A profile is intuitively a list of all the judgments of the agents involved in the collective decision at issue. For example, the profile involved in the paradoxical example of the previous section is the following: $(\{A, A \rightarrow B, B\}, \{A, \neg(A \rightarrow B), \neg B\}, \{\neg A, A \rightarrow B, \neg B\})$.

We can now introduce the concept of aggregation procedure that is, mathematically, a function. The domain of the aggregation procedure is given by $J(\Phi)^n$, namely, the set of all possible profiles of individual judgments.

Definition 2.3. An **aggregation procedure** for agenda Φ and a set of n individuals is a function $F : J(\Phi)^n \rightarrow \mathcal{P}(\Phi)$.

An aggregation procedure maps any profile of individual judgment sets to a single collective judgment set (an element of the powerset of Φ). Given the definition of the domain of the aggregation procedure, the framework presupposes *individual rationality*: all individual judgment sets are complete and consistent. Note that we did not yet put any constraint on the collective judgment set, i.e. the result of aggregation, so that at this point the procedure may return an inconsistent set of judgments. This is motivated by our intention to study both consistent and inconsistent collective outcomes. For example, in the doctrinal paradox of the previous section, the majority rule maps the profile of individual judgments into an inconsistent set:

$$\begin{aligned} (\{A, A \rightarrow B, B\}, \{A, \neg(A \rightarrow B), \neg B\}, \{\neg A, A \rightarrow B, \neg B\}) &\mapsto \{A, A \rightarrow B, \neg B\} \\ &\mapsto \\ &\{A, A \rightarrow B, \neg B\} \end{aligned}$$

²Of course this may be a descriptively inadequate assumption. However, on the one hand, these requirements are to be understood in a normative way, e.g. we exclude that a representative would vote for a proposal A and a proposal $\neg A$ at the same time. Moreover, the agenda may contain very simple logical propositions: as we shall see, it is sufficient to assume very minimal reasoning capacity to get the paradoxical outcomes.

The consistency of the output of the aggregation is defined by the following properties. An aggregation procedure F , defined on an agenda Φ , is said to be **collectively rational** iff F is:

- **complete** if $F(\mathbf{J})$ is complete for every $\mathbf{J} \in J(\Phi)^n$;
- **consistent** if $F(\mathbf{J})$ is consistent for every $\mathbf{J} \in J(\Phi)^n$;

That is, collective rationality forces the outcome of the procedure to be rational in the same sense of the individual rationality. Of course, the case of doctrinal paradox violates collective rationality.

We now introduce a number of *axioms* that provide a mathematical counterpart of our intuition on what a fair aggregation procedure is. The following are the most important axioms for JA discussed in the literature (List and Pettit 2002; List and Puppe 2009):

- **Unanimity** (U): If $\phi \in J_i$ for all i , then $\phi \in F(\mathbf{J})$.
- **Anonymity** (A): For any profile \mathbf{J} and any permutation $\sigma : N \rightarrow N$ we have $F(J_1, \dots, J_n) = F(J_{\sigma(1)}, \dots, J_{\sigma(n)})$.
- **Neutrality** (N): For any ϕ, ψ in the agenda Φ and profile $\mathbf{J} \in J(\Phi)^n$, if for all i we have that $\phi \in J_i \Leftrightarrow \psi \in J_i$, then $\phi \in F(\mathbf{J}) \Leftrightarrow \psi \in F(\mathbf{J})$.
- **Independence** (I): For any ϕ in the agenda Φ and profiles \mathbf{J} and \mathbf{J}' in $J(\Phi)^n$, if $\phi \in J_i \Leftrightarrow \phi \in J'_i$ for all i , then $\phi \in F(\mathbf{J}) \Leftrightarrow \phi \in F(\mathbf{J}')$.
- **Systematicity** (S): For any ϕ, ψ in the agenda Φ and profiles \mathbf{J} and \mathbf{J}' in $J(\Phi)^n$, if $\phi \in J_i \Leftrightarrow \psi \in J'_i$ for all i , then $\phi \in F(\mathbf{J}) \Leftrightarrow \psi \in F(\mathbf{J}')$.

Unanimity entails that if all individuals accept a given judgment, then so should the collective. Anonymity states all individuals should be treated equally by the aggregation procedure. Neutrality is a symmetry requirement for propositions that entail that all the issues in the agenda have to be treated equally. Independence says that if a proposition is accepted by the same subgroup under two distinct profiles, then that proposition should be accepted either under both or under neither profile. These axioms express our intuition concerning the fairness of the procedure, for example, (A) forces the procedure not to discriminate between individuals. This fairness condition may be used to model the arguments of an agent for accepting to solve conflicts by means of such a procedure. Systematicity is simply the conjunction of Independence and Neutrality and has been introduced separately as it is the condition used to prove the impossibility theorem in judgment aggregation. The impossibility theorem of List and Pettit (2002) is stated as follows.

Theorem 1 (List and Pettit 2002). *There are agendas Φ such that there is no aggregation procedure $F : J(\Phi)^n \rightarrow \mathcal{P}(\Phi)$ that satisfies (A), (S), and collective rationality.*

In particular, for any aggregation procedure that satisfies (A) and (S), there is a profile of judgment sets that returns an inconsistent outcome. The majority rule, that we have seen in the examples of Sect. 2.2, satisfies (A) and (S); accordingly, the discursive dilemma shows a case of inconsistent aggregation.

Very simple agendas may trigger inconsistent outcomes, one example being the agenda of the doctrinal paradox that we have presented in Sect. 2.2 $\{A, A \rightarrow B, B, \neg A, \neg(A \rightarrow B), \neg B\}$. Technically, any agenda that contains a minimal inconsistent set of cardinality greater than 2, such as $\{A, A \rightarrow B, \neg B\}$, may trigger a paradox. Thus, an agenda with respect to which the majority rule always returns consistent outcomes is a very simple agenda that contains, for example, only unconnected pairs of propositional atoms and their negations. Hence, paradoxical outcomes are very likely to occur in any complex social decision.

The methodology of JA can be extended to treat many voting procedures and characterize whether they may return inconsistent outcomes. Moreover, since the notion of aggregation procedure is very abstract, one can in principle model more complex procedures or norms, such as those that define decision making in organizations and corporations.

2.4 Social Attitudes and Conflict as Contradiction

We have seen that JA provides a precise mathematical modeling of the relationship between individual judgments and collective judgments. The relationship is formalized by means of an aggregation procedure and several properties of such aggregation can be discussed and analyzed. Moreover, it is possible to characterize the situations that lead to inconsistent outcomes. In this section, we introduce three notions that ground our ontological analysis of conflicts, that is, the notion of *propositional attitude*, the notion of *conflict as contradiction*, and the notion of *social attitude*.

2.4.1 Propositional Attitudes in JA

Propositional attitudes have been widely discussed in the philosophical literature and, roughly speaking, they express a relationship between an agent i and a propositional content p . For example, an agent can believe, judge, desire, prefer, ought, ... p , where p represents the content of the attitude. To our end, since the point of view of this work is knowledge representation, propositional attitudes are important as they allow for distinguishing a sharable propositional content of an attitude from the agent to whom the attitude is ascribed. Thus, we view individual propositional attitudes as sentences that are publicly expressed and communicated to other agents. Moreover, by using propositional attitudes, we are assuming that the matter of conflict between two agents can be in principle described by a third person in a sharable way.

We can introduce a formal language to represent how agents can communicate and reason about their attitudes, by building upon the rich logical tradition in the representation of propositional attitudes. For example, beliefs can be represented in

epistemic modal logic (van Benthem 2011). Intentions can be modeled by using a number of techniques in multiagent systems. Moreover, ought sentences are widely studied in deontic logics. Preferences can be represented by means of a fragment of first order logic: we introduce predicates Pab that represent the information “ a is preferred to b .” The rationality constraints on preferences, i.e. transitivity, reflexivity, and completeness, can be expressed by means of first order formulas (Porello 2010).

Therefore, general propositional attitudes can be in principle taken into account in the framework of JA (Dietrich and List 2009). We briefly sketch how. It is enough to extend the logical language that is used to model individual attitudes. For example, if we want to deal with beliefs, we extend the agenda Φ that we have introduced in the previous section, by adding individual belief operators in epistemic modal logic $B_i A$, standing for “The agent i believes that A .”

Let \mathbf{A} be a type of propositional attitudes, we label $L_{\mathbf{A}}$ the logical system for representing the type of propositional attitudes \mathbf{A} . That is, $L_{\mathbf{A}}$ refers to the language to represent propositional attitudes \mathbf{A} and to the logical rules to reason about such attitudes, e.g. an axiomatic system for that logic. Accordingly, we define an agenda $\Phi_{\mathbf{A}}$ as a subset of the language of $L_{\mathbf{A}}$. In the previous section, we have defined the possible sets of individual judgments by means of $J(\Phi)$, namely we assumed that individual judgment sets are consistent and complete with respect to (classical) propositional logic. In the general case, it is possible to define judgment sets that are rational with respect to different logical systems (Porello 2013). We define $J_{\mathbf{A}}(\Phi_{\mathbf{A}})$ as the set of possible sets of attitudes that satisfy the rationality constraints that are specific to \mathbf{A} . For instance, in case \mathbf{A} are preferences, sets of preference attitudes have to respect transitivity. In case \mathbf{A} are beliefs, they should be consistent, in the sense that an agent is not supposed to believe A and $\neg A$ at the same time, therefore we exclude sets containing both $B_i A$ and $B_i \neg A$.

The general form of an aggregation procedure is a slight generalization of the one introduced in the previous section. An aggregation procedure is a function from profiles of individual attitudes to sets of collective attitudes: $F : J(\Phi_A)^n \rightarrow \mathcal{P}(\Phi_A)$. The notion of collective rationality again may change as we may add more specific constraints on the type of attitudes at issue. For example, in preference aggregation we add the constraints on preference orders. Since each one of these extensions includes propositional logic, the impossibility theorem shall hold for the larger fragment. Thus, it is at least in principle possible to extend the map of consistent/inconsistent aggregation to richer languages.

2.4.2 Conflict as Contradiction

Once we represent agents’ attitudes, we can introduce a general definition of the notion of conflict. The notion of conflict that we define is placed at the level of the representation of propositional attitudes.

Given two sets of attitudes A and A' of the same type A , we say that A is in *conflict* with A' iff the set $A \cup A'$ entails a *contradiction* in the formal system L_A that represents those attitudes. That is, the two sets of attitudes are inconsistent with respect to L_A . For instance, two conflicting judgment sets in the sense of the previous section are simply two sets of propositions that are inconsistent with respect to propositional logic, e.g. $\{A, B, C\}$ and $\{\neg A, B, C\}$. Moreover, two conflicting preferences are two sets of preferences that together entail a contradiction, such as $\{Pab, Pac, Pbc\}$ and $\{Pba, Pac, Pbc\}$: i.e. Pab and Pba entail by transitivity Paa , which contradicts irreflexivity. Conflicting preferences and goals entail that they cannot be satisfied at the same time.

Note that our notion of conflict applies to sets of attitudes of the *same* type A . Thus, we do not say, for example, that an intention is inconsistent with a belief. This is so because, in our view, a belief can contradict an intention, or an ought, only with respect to a reasoning system that includes both attitudes and makes the relationships between them explicit. Such a reasoning framework has to contain a principle that links the different types of propositional attitudes that are matter of discussion. An example of such a principle is (one version of) the means-end *principle of instrumental rationality*³: “if I intend to A and I believe that B is a sufficient means for achieving A , then I intend to B .” By means of such a principle, we can see how a belief may contradict an intention as follows: suppose I intend to A , my belief that B is a sufficient means to get A would be inconsistent with the fact that I do not intend to B . Situations of *acrasia* can be represented by a similar argument. Our approach is motivated by the fact that in general we do not want to be committed with a philosophically onerous claim that a belief per se can contradict a preference or a desire or an ought.

In our modeling, the notion of contradiction has the following intuitive interpretation: two inconsistent sets of attitudes cannot be satisfied at the same time, e.g. two conflicting preferences entail that either one or the other can be accepted. We can define the conflict between two agents by simply saying that agent i is in conflict with agent j if the set of attitudes of i A_i is inconsistent with the set of attitudes of j , A_j , namely $A_i \cup A_j$ is inconsistent with respect to the formal system L_A . This can be easily generalized to conflicts involving m agents: $A_1 \cup \dots \cup A_m$ is inconsistent with respect to L_A .

Note that our definition allows for an agent being in conflict with him/herself, in case she/he maintains a set of inconsistent attitudes; For example, if an agent i has a set of judgments such as $\{A, A \rightarrow B, \neg B\}$. We shall use this fact in the next paragraph. The abstract notion of conflict that we have defined can be instantiated in order to provide a representation of actual conflicts. For example, if we want to view a chess match as a situation of conflict between two agents, we can represent the conflicting aspect of the match by describing the agents' opposing goals of winning by beating the other.

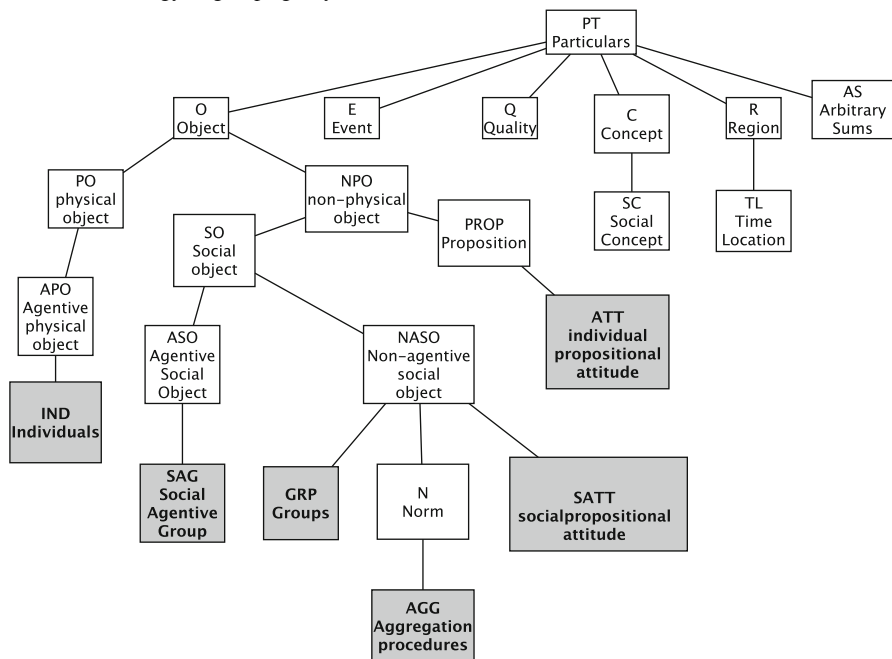
³For a discussion on the status of instrumental rationality, see Nozick (1993).

It is important to stress that, in order to talk about a contradiction, we need to make the reasoning system L_A explicit. For instance, the set of preferences $\{Pab, Pba\}$ is *not* inconsistent with respect to a reasoning system L_A that does not impose irreflexivity. Thus, in order to claim that some attitudes are inconsistent, the individuals have to agree on the reasoning framework that grounds the inconsistency claims.

The point is that any contradiction depends on the reasoning system that is adopted to evaluate the matter at issue. Imagine two agents that have apparently conflicting preferences but that do not share the common reasoning rules that define what a contradiction is. For example, the preferences of agent 1 and agent 2 may be incompatible from the point of view of agent 1 but not from the point of view of agent 2. Agent 1's most preferred option may be a , whereas agent 2 may have two equally most preferred options a and b . That is, 1 is reasoning according to the rules of preferences that we have presented before, namely she/he linearly orders alternatives, whereas 2 has a partial order on alternatives. In that case, 1 believes that the policy a has to be implemented, whereas 2 believes that both a and b have to be implemented. In such a case, the disagreement is on the nature of the alternatives and that is reflected on the rules that norm reasoning about such matters. Thus, the conflict is at a more abstract level: it is about the reasoning principle that norms the matter at issue. It is important to stress that also claiming principles is a form of propositional attitude, thus the conflict of principle is a type of conflict that fits the definition that we have presented, provided the agents agree on the reasoning framework that judges conflicting principles. By iterating this argument, we could imagine situations of an indeterminate *regressus*: in order to acknowledge that we are in conflict on a certain matter, we need to agree on the principles that establish such conflict, but if we are in conflict on such principles, we need other principles that establish the conflict about principles and so on. However, it is not clear whether such a situation can be classified as a conflict, namely it is not clear on what ground agents in such a scenario can claim to have conflicting attitudes. Although such a situation is theoretically possible and interesting to investigate, in this paper we want to focus on types of conflicts that are actually recognizable by the agents involved, and that require an agreement on what conflicting attitudes are. Hence, we shall not discuss this type of situations further. In this work we shall assume that the blame of inconsistency is shared among the individuals, namely that they agree on a common reasoning system that specifies what is a contradiction between sets of propositional attitudes, and we leave cases of asymmetric blame for future work.

2.4.3 *Social Agentive Groups and Social Contradiction*

We have seen that SCT defines how to aggregate the propositional attitudes of a number of possibly conflicting heterogenous agents into a single set of attitudes. In particular, SCT and JA provide a way to view the group as a *single agent* and to ascribe propositional attitudes to the group itself. We present some elements of an

Table 2.1 Ontology of group agency in DOLCE

ontological treatment of conflicts, by placing our treatment within the foundational ontology DOLCE. The categories that we use are summarized in Table 2.1 at the end of the paper, boldface categories are new wrt DOLCE. It is easy to define *individual attitudes* as propositional attitudes that are ascribed to an individual agent i . We introduce a relation $ASC(a, i)$ between propositional attitudes of a certain type and individuals. In order to define ascription, we need a category $ATT(x)$ for propositional attitudes and a category $IND(x)$ for individual agents: $ASC(a, i) \rightarrow ATT(a) \wedge IND(i)$. We shall also ascribe sets of attitudes A_i to individuals, we write $ASC(A_i, i)$ as a shorthand for $\bigwedge_j ASC(a_j, i)$, for all $a_j \in A_i$.

We are going to define the notion of *social propositional attitude* as a propositional attitude that is ascribed to the group of agents itself and not to any of the individuals belonging to the group. Of course, it does not seem meaningful to ascribe propositional attitudes to any set of individuals, or to any type of group. For instance, if we talk about the beliefs of 15 individuals randomly chosen from the phone book, we are simply talking about the sum of all individual beliefs, and not about a single belief that is ascribed to the group itself. We need to be careful when defining attitudes ascribed to groups since propositional attitudes are usually properly intended as ascribed to *agents*. Thus, we would make a category mistake in applying an attitude to something non-agentive, in the same way as we would make a category mistake in attributing beliefs or intentions to a time interval.

For example, take a strategic setting described by game theory, such as a market. We claim that it is a category mistake to ascribe attitudes to the outcome of the interaction of agents, e.g. “the market believes, decides, intends, . . . to p .” The point is that such ascription may be metaphorically effective, however it is not grounded in a definition of any agent who is entitled to carry the social or collective belief, decision, intention. Namely, the market is not constructed as an agent.

On the other hand, there are cases in which it is meaningful, and sometimes even necessary, to ascribe attitudes to parliaments, representative assemblies, corporations, organizations. For example, ascribing attitudes to the group is required, in case we want to ascribe responsibility to the group itself.

The point is that, whenever we want to ascribe attitudes to a group, we need to show that the group is some type of agent. We are going to define this specific type of group that we label *social agentive group*. We will show that this notion of social group is required in order to understand the type of conflict of SCT paradoxes.

A *social agentive group* depends on a set of individuals N and on an aggregation procedure in the sense of Sect. 2.3. The social agentive group is defined by those agents that agree to be subject to a particular aggregation procedure. The fact that such individuals acknowledge an aggregation procedure means simply that they agree on the rule to settle their conflicts. For example, the group of representatives in a parliament and the majority rule: a single representative may disagree with a collective decision, however she/he implicitly has to acknowledge it and be subject to the consequences of that decision. Note that any set of individuals and any of the aggregation procedures in the sense of Sect. 2.3 define a social agentive group. We view the agreement on the aggregation procedure as baptizing a new type of object, namely a new agent, the social agentive group, $SAG(g)$.⁴

We need to introduce the following categories. Let AGG be the class of aggregation procedures, GRP the class of groups (i.e., sets of individuals), IND the class of individual agents. We represent the membership of an individual i in a group N by means of the relation $MEMB(i, N)$. Moreover, we introduce $ACK(i, f)$ to represent the acknowledgment relation that holds between an individual i and an aggregation procedure f .⁵ Firstly, we define a social agentive group as a subclass of agentive social objects ASO defined in Masolo et al. (2004) and Bottazzi and Ferrario (2009). That is, a social agentive group is a social object that is assumed to have agency: $SAG(x) \rightarrow ASO(x)$.

⁴We are assuming that the social agentive group is a distinct object with respect to the group as a set of individuals. The reason is that we want to attribute to the social agentive group properties of a different kind with respect to those that we can attribute to the group. In this sense, the social agentive group is a *qua* object.

⁵Here we present the definitions in a semi-formal fashion. Our analysis can be incorporated in the ontological treatment of DOLCE (Masolo et al. 2003). Note that, although the definition seems to be in second order logic, it is possible to flatten the hierarchy of concepts by typing them. This is the so-called *reification* strategy of DOLCE. We leave a precise presentation of DOLCE for future work.

Moreover, the existence of a social agentive group depends on an aggregation procedure in the following sense. We assume as a necessary condition that an agentive group is correlated to a group of individuals as well as to an aggregation procedure.⁶

$$\text{SAG}(g) \rightarrow \exists f, \exists N \text{ AGG}(f) \wedge \text{GRP}(N) \wedge \forall i (\text{MEMB}(i, N) \rightarrow \text{ACK}(i, f)) \quad (2.1)$$

Definition (2.1) means that an agentive group g depends on a group of individuals N and an aggregation procedure f such that every individual in the group acknowledges f . Since the category of groups GRP and the category of agentive social objects ASO are disjoint, we are assuming that the set of individuals and the agentive social group are distinct objects of our ontology.

There are further conditions on social agentive groups, for example, given a social agentive group g , there is a unique aggregation procedure for g at a given time.

In order to simplify the presentation of social agentive groups and to focus on conflict, we abstract here from issues related to time and change (Porello et al. 2014).⁷

The acknowledgment relation is here intentionally designed to be abstract because it may be subject to different interpretations depending on the type of group and individual agents. For example, members of an organization subscribe the rules of the organization, employees sign the employment contract, representatives of the parliament are bound by oath to the constitution, and so on. The properties of aggregation procedures that we have introduced in Sect. 2.3 may be used in order to define under which conditions an agent is willing to accept an aggregation procedure, in less institutionalized cases; for example, an anonymous aggregation procedure can be accepted on the ground that it ensures a form of impartiality.

Here we assume, according to our previous analysis, that the acknowledgment of the aggregation procedure entails the individual agreement on the reasoning framework L_A that is used to judge conflicts. This is because an aggregation procedure is defined on a specific input which takes propositional attitudes that

⁶For a precise ontological treatment of the agency of groups, we refer to Porello et al. (2014).

⁷For example, we may discuss whether a social agentive group remains the same by adding or removing members of the set of individuals or by reforming the aggregation procedure. For this reason, we did not put the unicity constraint on N and f in Definition (2.1). Moreover, by viewing social agentive group with respect to time, the acknowledgment relation has to be parametrized wrt times as well. One application of a time-dependent acknowledgment relation is that, in order to reform the aggregation procedure at a certain moment, a new acknowledgment may be required. However, at a time slice, the group and the procedure are supposed to be unique. This is motivated by the simple observation that if we were to allow for two different aggregation procedures at a given time, with possibly divergent outcomes, the attitudes of the social agentive group would always be indeterminate.

are rational according to L_A . Hence, in order to accept an aggregation procedure, agents have to accept that only propositional attitudes that are rational wrt L_A can be submitted and that amounts to endorsing L_A .

We can now define social attitudes (SATT) as propositional attitudes that are ascribed to the social agentive group. Our definition does not entail that the social attitude is ascribed to any of the individuals of the group, although the attitude of the group may coincide with the attitude that is ascribed to some of its members. Since a social agentive group is defined by an aggregation procedure at a time, we can define the relation of *dependence* of the social agentive group on the aggregation procedure and denote it by $\text{DEP}(g, f, t)$. Moreover, we define the dependence of the social agentive group on the set of individuals at a given time by $\text{DEP}(g, N, t)$.

A *social attitude* is a propositional attitude (a) that is obtained by means of the aggregation procedure f . By using the notation of Sect. 2.3, $a \in f(A_1, \dots, A_n)$, meaning that a belongs to the output of the aggregation procedure when given as input the profile of individual attitudes A_1, \dots, A_n . An exhaustive ontological treatment of $a \in f(A_1, \dots, A_n)$ entails, for example, that the individual propositional attitudes A_j are ascribed to individual j .

$$\text{SATT}(a) \rightarrow \exists x \exists t (\text{SAG}(x) \wedge \text{DEP}(x, f, t) \wedge \text{DEP}(x, N, t) \wedge a \in f(A_1, \dots, A_n)) \quad (2.2)$$

Definition (2.2) means that a social attitude depends on the group and the aggregation procedure that define the social agentive group at a given time. Thus, we can now legitimate the ascription of a social attitude to the social agentive group by slightly modifying our previous definition of ascription: $\text{ASC}(x, y) \rightarrow (\text{ATT}(x) \wedge \text{IND}(y)) \vee (\text{SATT}(x) \wedge \text{SAG}(y))$. Note that, by Definition (2.2), a social attitude is necessarily ascribed to some social agentive group. This is motivated by the fact that we want to exclude that taking, for example, the beliefs of a number of randomly chosen individuals and aggregating them by majority is sufficient to define a social attitude. A propositional attitude needs to be ascribed to an agent, whereas an arbitrary number of individuals does not count in general as a single agent.

We can finally introduce the notion of *social contradiction* in order to analyze the paradoxical outcomes of SCT. Firstly, we introduce a relation for making the notion of contradictory set of attitudes explicit in our ontology. We identify sets of attitudes with conjunctions of formulas and we express that the formula $a_1 \wedge \dots \wedge a_m$ is inconsistent with respect to the reasoning principles of L_A by means of the relation $\text{CTR}(a_1 \wedge \dots \wedge a_m, L_A)$. According to our previous analysis, the notion of contradiction has to depend on the reasoning system that is adopted.

For example, $\text{CTR}(Pab \wedge Pba, \{\text{irreflexivity, transitivity, completeness}\})$ holds, whereas $\text{CTR}(Pab \wedge Pba, \{\text{transitivity}\})$ does not. A *social contradiction* is just an inconsistent set of social attitudes. This definition entails that there exists a social agentive group who maintains those inconsistent attitudes.

We can now stress the difference between the notion of social agentive group defined by means of SCT and other notions of groups that may be treated, for

example, by means of game theory. The fundamental difference is that SCT allows to view the group as an agent, namely as a social agentive group, whereas game theory does not provide agency to the group itself. The notion of social contradiction defines the contradiction of the social agentive group with itself viewed as a single agent.

This analysis of social contradiction precisely represents situations such as the Condorcet's Paradox and the discursive dilemma. Note that, without the concepts that we have introduced, it is hard to identify what type of conflict cases like that exhibit. Social contradictions are not conflicts between individuals that belong to the group, since the group is defined by the agreement on the procedure that settles individual possible conflicts. Social contradictions are not conflicts between different groups, as in the paradoxical case only one group is involved. Moreover, social contradictions do not apply to general groups of individuals, they are specific to social agentive groups. It comes with no surprise that a number of individuals may have conflicting attitudes and that there is no way to solve their conflicts. The point of social contradiction is that, although individuals agree on the rule to settle conflicts, this peculiar type of conflict can still occur. Therefore, the type of conflict of the SCT paradoxes is a specific type of conflict that applies only to groups insofar as they are viewed as social agentive groups and that is not reducible to any individual conflict. The non-reducibility of social contradictions to individual conflicts can be argued by simply noticing that we cannot say which conflict between individuals is responsible of the social contradiction. For example, by reducing the social contradiction to conflicts between individuals, we would not be able to distinguish the opposition between the majority and the minority in a paradoxical case and the opposition between the majority and the minority in a coherent and unproblematic majority voting. It can be argued that it is the procedure that is responsible for the paradoxical outcome, e.g. the majority rule. However, the majority is reliable in many other cases and SCT results show that the only procedures that ensure consistency are the dictatorships of some individual. Therefore, social contradictions are something we have to live with, as they may occur in any possible actual solution of individual conflicts that ascribes agency to the group. Without the notion of social agentive group, we could not ascribe propositional attitudes to the group itself, and we could only interpret SCT paradoxes as conflicts between individuals. Thus, our specific treatment of conflict as social contradiction is needed as social contradictions are non-reducible to other forms of conflict.

2.5 A Taxonomy of Conflicts

We present a taxonomy of conflicts along the conceptual analysis that we have outlined. We distinguish types of conflicts that depend on two parameters: the type of agents involved (individual agents or social agentive groups) and the matter of conflict (namely, the type of propositional attitude at issue).

	Agents	About (propositional attitudes)	Type
I	IND: i vs j	Beliefs, desires, judgments,...	Contradiction
II	GRP: $\exists i_1, \dots, i_m$ in G , i_1 vs ... vs i_m	Beliefs, desires, judgments,...	Contradiction
III	SAG: sag vs sag'	Beliefs, desires, judgments,...	Contradiction
IV	SAG: sag vs sag	Beliefs, desires, judgments,...	Social contradiction

(I) classifies conflicts between individuals (including the conflict of an individual with him/herself) that may be about any propositional attitude. (II) classifies conflicts within groups that are reducible to conflicts among members of the group. In this case, the group is not viewed as a social agentive group and the conflict within the group can be reduced to conflicts between subsets of individuals. As an example, take an auction in which a number of agents make their bids for getting a certain item and only one of them can win the item. (III) classifies conflicts between different groups each of them viewed as social agentive groups, for example two different parliaments of different states voting two incompatible policies. Finally, (IV) classifies the case of social contradictions that are exemplified by SCT paradoxes. From the point of view of our ontological analysis, (III) can be reduced to (I): namely conflict between two different social agentive groups can be modeled as conflict between different individual agents, that is, it can be modeled by using the notion of contradiction between propositional attitudes of two different agents. Moreover, our modeling shows that the type of conflict that is defined in (II) is actually a conflict between individuals: again, it can be modeled by means of the notion of contradiction between a number of individual attitudes. The notion of social contradiction is required only to model the conflict of the social agentive group with itself, namely the group conflict that is non-reducible to any conflict between any member of the group.

2.6 Conclusion and Future Work

We have developed the first conceptual elements to provide an ontological analysis of group conflicts. We have used the methodology of SCT in order to mathematically understand a number of types of conflicts and to define the concept of social contradiction. We have introduced some fundamental elements of an ontological analysis of conflicts by spelling out the required concepts of propositional attitude, conflict as contradiction, and social agentive group. In particular, we have argued that the concept of social agentive group is necessary in order to understand the type of group conflict that is involved in social paradoxes. We plan to provide a

fine-grained ontological representation of aggregation procedures that would enable modeling the dynamics of group formation and change, possibly motivated by conflicts, besides allowing us to distinguish between types of groups in terms of the properties of the aggregation procedure that is endorsed. A close examination of the norms that specifically apply to groups is then compelling (Aldewereld et al. 2014). The next step is to integrate our analysis within the general framework of a foundational ontology such as DOLCE. Complex aggregation procedures can be applied to treat the rich internal structure of organizations (Boella et al. 2004; Bottazzi and Ferrario 2009), for example by defining the notion of sub-organization and by formalizing the relationship between the different modules. To that extent, we have started developing a module for ascribing agency to groups and organization in Porello et al. (2014). This leads towards a generalization of our model to provide an understanding of the ascription of agency to complex social systems and socio-technical systems and to apply our treatment of conflicts in such complex social constructions. In particular, modeling socio-technical systems requires to integrate information coming from heterogeneous agents, human and artificial, and it is important to deploy conceptual tools, such as those that we have discussed, that provide a precise description of the concept of aggregate information. We have presented a number of applications of the methodology of SCT to model systemic information in socio-technical systems in Porello et al. (2014), we plan to integrate that analysis with the present investigation of conflict and social contradictions in order to grasp situation of crisis in socio-technical systems.

Acknowledgements D. Porello and R. Ferrario are supported by the VisCoSo project, financed by the Autonomous Province of Trento, “Team 2011” funding programme. E. Bottazzi is supported by the STACCO project, financed by the Autonomous Province of Trento, “Postdoc 2011” funding programme.

References

- Aldewereld H, Dignum V, Vasconcelos W (2014) We ought to; they do; blame the management! In: Coordination, organizations, institutions, and norms in agent systems IX. Springer, Berlin, pp 195–210
- Arrow K (1963) Social choice and individual values. Cowles Foundation for Research in Economics at Yale University, Monograph 12. Yale University Press, New Haven
- Boella G, Lesmo L, Damiano R (2004) On the ontological status of plans and norms. *Artif Intell Law* 12(4):317–357
- Boella G, Pigozzi G, Slavkovic M, van der Torre L (2011) Group intention is social choice with commitment. In: Proceedings of the 6th international conference on coordination, organizations, institutions, and norms in agent systems, COIN@AAMAS’10. Springer, Berlin/Heidelberg, pp 152–171
- Bottazzi E, Ferrario R (2009) Preliminaries to a DOLCE ontology of organizations. *Int J Bus Process Integr Manag* 4(4):225–238 [Special Issue on Vocabularies, Ontologies and Business Rules for Enterprise Modeling]

- Brandt F, Conitzer V, Endriss U (2013) Computational social choice. In: Weiss G (ed) Multiagent systems. MIT Press, Cambridge
- Bratman ME (1987) Intention, plans, and practical reason. Harvard University Press, Cambridge
- Dietrich F, List C (2009) The aggregation of propositional attitudes: towards a general theory. Technical report
- Emery FE, Trist E (1960) Socio-technical systems. In: Management sciences: models and techniques, vol 2. Pergamon, New York, pp 83–97
- Endriss U, Grandi U, Porello D (2012) Complexity of judgment aggregation. *J Artif Intell Res* 45:481–514
- Gaertner W (2006) A primer in social choice theory. Oxford University Press, Oxford
- Georgeff MP, Pell B, Pollack ME, Tambe M, Wooldridge M (1999) The belief-desire-intention model of agency. In: Proceedings of the 5th international workshop on intelligent agents V, agent theories, architectures, and languages, ATAL '98. Springer, London, pp 1–10
- Kornhauser LA, Sager LG (1993) The one and the many: adjudication in collegial courts. *Calif Law Rev* 81(1):1–59
- List C, Pettit P (2002) Aggregating sets of judgments: an impossibility result. *Econ Philos* 18:89–110
- List C, Puppe C (2009) Judgment aggregation: a survey. In: Handbook of rational and social choice. Oxford University Press, Oxford
- Masolo C, Borgo S, Gangemi A, Guarino N, Oltramari A (2003) Wonderweb deliverable d18. Technical report, CNR
- Masolo C, Vieu L, Bottazzi E, Catenacci C, Ferrario R, Gangemi A, Guarino N (2004) Social roles and their descriptions. In: Proceedings of the 6th international conference on the principles of knowledge representation and reasoning (KR-2004), pp 267–277
- Neumann JV, Morgenstern O (1944) Theory of games and economic behavior. Princeton University Press, Princeton
- Nozick R (1993) The nature of rationality. Princeton University Press, Princeton
- Osborne MJ, Rubinstein A (1994) A course in game theory. MIT Press Books. The MIT Press, Cambridge
- Pettit P (2001) Deliberative democracy and the discursive dilemma. *Philos Issues* 11(1):268–299
- Porello D (2010) Ranking judgments in arrow's setting. *Synthese* 173(2):199–210
- Porello D (2013) A proof-theoretical view of collective rationality. In: IJCAI 2013, Proceedings of the 23rd international joint conference on artificial intelligence, Beijing, 3–9 August 2013
- Porello D, Bottazzi E, Ferrario R (2014) The ontology of group agency. In: Garbacz P, Kutz O (eds) Formal ontology in information systems - Proceedings of the eighth international conference, FOIS 2014, Rio de Janeiro, 22–25 September 2014. Frontiers in artificial intelligence and applications, vol 267. IOS Press, pp 183–196
- Porello D, Setti F, Ferrario R, Cristani M (2014) Multiagent socio-technical systems: an ontological approach. In: Coordination, organizations, institutions, and norms in agent systems IX. Springer, Berlin, pp 42–62
- Taylor AD (2005) Social choice and the mathematics of manipulation. Cambridge University Press, Cambridge
- van Benthem J (2011) Logical dynamics of information and interaction. Cambridge University Press, Cambridge
- Wooldridge M (2008) Introduction to multiagent systems. Wiley, New York

Chapter 3

On Stages of Conflict Escalation

Jens Allwood and Elisabeth Ahlsén

3.1 Introduction

An issue in the theory of conflict is whether there are stages (steps, phases, or levels—the terminology varies) in conflict escalation (and de-escalation). If so, how many are there and what are their identifying characteristics?

A prerequisite for identifying stages in conflict is a definition of what a conflict is. In this paper, we take the following definition (cf. Allwood 1992) as our point of departure:

Conflict: A and B are in conflict = A and/or B believe they have incompatible interests and/or perform negative actions against each other.

3.2 Taxonomies of Conflict

There are a number of aspects that can be considered in characterizing and classifying conflicts. Some possible taxonomies of conflict are:

1. The number of participants. Is it a two-party (bilateral) or three-party (trilateral) conflict, or are many parties involved (multilateral conflict)?
2. The degree of interactivity: Is it a one-way or a two-way conflict?
3. The degree of overtness: Is it an overt or a covert conflict?

An *overt* conflict occurs when two agents are in overt conflict, if they both experience grounds for conflictual action against each other and as a result take

J. Allwood (✉) • E. Ahlsén
Department of Applied IT, SCCIIL Interdisciplinary Center, University of Gothenburg,
41296 Gothenburg, Sweden
e-mail: jens@ling.gu.se

such action. The experienced grounds for conflict can, but need not, correspond to any actual grounds for conflict.

A *covert* conflict can either be an actual two-party conflict which is concealed from another interested third party or a case where conflictual action is taken by one agent against another agent, who is unaware of the action, but who would, if the action were discovered, experience it as conflict generating and take countermeasures.

4. The distribution of power between the conflicting parties: Is it a symmetric (equal power) or asymmetric (unequal power) conflict?
5. The type of activity, organization, and topic which is involved in the conflict: Is it a salary/wage conflict, a courtroom trial, bargaining in a marketplace, a political conflict, a peace negotiation, a dowry negotiation, a divorce negotiation, or a family conflict (e.g., parent-child about pocket money, staying out at night, homework, husband-wife about house cleaning, etc.)?
6. What modalities are applicable—alethic, deontic, and epistemic? Is the conflict manifest vs. latent; actual vs. potential, possible, actual, and necessary; permitted vs. obligatory; or conceivable vs. certain?

A related distinction is that between normative and descriptive aspects of conflict. A *normative* perspective deals with the question of how conflicts should be pursued in different activities. A *descriptive* perspective studies how conflicts are actually pursued in different activities and organization. A *possible potential* perspective, finally, asks how a conflict can/could be pursued.

7. The type of medium of communication involved in the conflict: Is it face-to-face, telephone, written (letter, e-mail, etc.), chat, videoconference, or other Internet-based synchronous communication?

These taxonomic features can be used to classify both long-term conflicts over a period of time and short-term conflicts as in a short conflict episode or particular instance of a conflict.

3.3 Responding to Conflictual Communication

There are several options for reacting and responding to conflictual communicative action.

The main options are: (1) acceptance of other's claim, (2) rejection, (3) avoidance, and (4) prevention of conflict.

The manner in which conflict is initiated and pursued through communication and the responses to and management of this communication can be the basis for identifying possible stages or steps in conflict escalation and de-escalation. In the following, we will present five suggested models of stages in conflict and then turn to a specific type of conflict (televised political debate), where we will try to identify potential stages, in order to see to what extent the five models are applicable. Finally, we will, on the basis of our analysis, compare political debates with other types of conflictual communication.

3.4 Suggested Models of Stages of Conflict

Different authors have suggested different numbers of stages and different ways of characterizing them, e.g., Friedrich Glasl (1997) suggests nine steps of conflict, Douglas Noll (2000) suggests five phases, and Eric Brahm (2003) suggests eight phases. Some authors do not suggest a definite number of stages; rather, they give lists of possible stages. Examples of this are the book *Everyone Can Win* by Cornelius et al. (1997) and the book *Interpersonal Conflict Escalation Levels* by Hocker and Wilmot (1991). See Table 3.1, below, for a summary of the stages suggested in Glasl (1997), Noll (2000), Brahm (2003), Cornelius et al. (1997), and Hocker and Wilmot (1991).

If we compare the different models, we can see that all the models of conflict escalation, except Brahm's, end quite dramatically with full-blown conflicts, involving mutual "annihilation" (Glasl), "regression" (Noll), possible "violence" (Cornelius et al.), and "deadly combat" (Hocker and Wilmot). Only Brahm provides a less pessimistic view, going from "stalemate" (step 5), via "de-escalation" and "settlement/resolution," to "post-conflict" and, finally, "peace and reconciliation." Most of the models are, thus, models only of conflict escalation and do not include the possibility of de-escalation.

The differences in the number of stages and in the labeling of the stages indicate that the different authors have somewhat different types of conflict in focus, and that most of them are models of conflict of a long-term, very serious type of conflict. At least three of them (Glasl, Cornelius et al., and Hocker and Wilmot) contain escalation that involves moving from words to action, from verbal threats to trying to hurt another person physically. This type of escalation is not typical for most everyday conflictual communicative interactions that often mainly contain argumentation, discussion, and perhaps quarrel.

However, some of the stages in all of the models can, to some extent, be applied to more short-term, nonphysical types of conflict, but, as we have seen, most of them primarily have a focus on more long-term conflicts, being applicable to conflicts with more of a long-term perspective than conversations, including also conflicts between groups and nations, leading to very serious confrontations like suicide bombings or war.

One way to capture the difference between different types of conflict is to consider the nature of the social activity they develop in. In general, different social activities can contain different types of conflicts, connected with different stages of conflict development. The differences between activities and conflicts may, in turn, require an assumption of different conflict stages for the most satisfying analysis in a theoretical model. Finding a suitable model of steps or stages of conflict may therefore be dependent on identifying the type of social activity where the conflict is occurring. In many cases, also a subtype of that type of activity may be what is required to understand a particular type of conflict. In a long-term conflict, this can, for example, mean identifying a set of steps or stages of conflict in spoken interaction (taking place during one particular interaction), and then in a

Table 3.1 Examples of models of stages, steps, or levels of conflict escalation

Glasl	Noll	Comelius et al.	Hocker and Wilmot	Brahm
<p>0: <i>Dialogue</i></p> <p>1: <i>Discussion</i>—hardening positions</p> <p>2: <i>Debate</i>—polarization</p> <p>3: <i>Running over the other</i>—own goals</p> <p>4: <i>Harassment</i>—scurrilous images</p> <p>5: <i>Loss of face</i></p> <p>6: <i>Strategical threats</i></p> <p>7: <i>Painful attacks</i>—cause damage</p> <p>8: <i>Elimination</i>—attacking “nerve center”</p> <p>9: <i>Together down the abyss</i>—annihilation</p>	<p>1. Part of normal, everyday life. Even good relationships have <i>moments of conflict</i></p> <p>2. The parties fluctuate between cooperation and competition</p> <p>3. <i>Concrete action</i>—no common solution</p> <p>4. <i>Cognitive function regresses</i>—know but do not consider each other’s perspectives</p> <p>5. Progressive regression</p>	<p>1. <i>Uncomfortableness</i>: an inner, intuitive feeling that something is going wrong</p> <p>2. <i>Incidents</i>: irritation</p> <p>3. <i>Misunderstanding</i>: communication is deficient</p> <p>4. <i>Tension</i> negative attitudes. Consciously or unconsciously people hurt each other</p> <p>5. <i>Crisis</i>: repressed emotions release. Violence can appear</p>	<p>1: A problem to be solved</p> <p>2: A difference</p> <p>3: Confrontation</p> <p>4: Fight and/or flight</p> <p>5: Deadly combat</p>	<p>1. No conflict</p> <p>2. Latent conflict</p> <p>3. Emergence</p> <p>4. Escalation</p> <p>5. (Hurting) Stalemate</p> <p>6. De-escalation</p> <p>7. Settlement/ resolution</p> <p>8. Post-conflict</p> <p>9. Peace and reconciliation</p>

further analysis of the conflict, other specifying stages of conflict may be required in the interactions that are connected with the conflict. Examples of conflict that might involve slightly different stages with regard to communication are a trial in court, a political debate, a family quarrel, an argument in a work team, etc. The considerations above, therefore, lead us to propose an activity-based approach in order to identify typical or possible steps of conflict in the communicative spoken interaction of different social activities.

3.5 An Activity-Based Approach to Interpreting and Describing Stages of Conflict

We thus suggest that there is not only one correct answer to the issue of how many stages of conflict escalation there are and what these stages are. Rather, we think that the number and types of stages must be related to the type of conflict we are concerned with. Therefore, different types of conflict may typically show different numbers and stages with different properties.

We will illustrate and support this claim below by an analysis of the number and types of stages found in short conflict episodes, occurring between politicians in televised political debates from different countries (Germany, Italy, Greece, and the USA). The debates involve different types of conflict episodes, characterized by more or less aggressive, accusing, scornful, derisive, ironic, triumphant, defiant, resigned, etc. stances and behavior.

An analysis of the “social signals” involved in these stances, i.e., the multimodal expressions occurring at different moments in the conflict episodes has yielded a set of clusters of behavior, which can be used for identifying possible stages, steps, or phases in the different types of episodes.

In our analysis, we focus on the stances and behavior exhibited by the politicians, rather than on, for example, the long-term consequences, which are the focus of several of the models we have described above, for example, in Glasl’s nine-step model. This difference in perspective we think illustrates how different types of conflict also enable a focus on different conflict affordances in the data and in this way may give rise to different models of conflict escalation, suitable for different purposes.

3.6 Method

3.6.1 *Material*

In order to analyze and illustrate stages of conflict in televised political debates, we have used a corpus consisting of four political debates occurring in three different countries, Germany, Italy, and the USA:

1. A German debate on whether it was correct to support rebels in Libya with military interventions (German debate “Enthaltung ist keine Haltung,” that is, “Abstention is no position”)
2. A German debate, “Atomkrieger” (“Nuclear wars”), where the health and moral implications of using nuclear energy are discussed among the participants of the debate
3. An Italian debate “Giuliano Pisapia vs. Letizia Moratti,” which is an election debate of the two main candidates running for the position of Mayor of Milan (2011)
4. “Republican Debate October 18, 2011” or “Perry vs. Romney”—two candidates running in the primary elections of the US Republican Party—a debate concerning the nomination of the party’s candidate for running for the US presidency

3.6.2 Analysis

For transcribing the videos, we used the Gothenburg transcription standard and the modified standard orthography (MSO6) (Nivre 2000, 2004), while annotations of the videos were done using ANVIL (Kipp 2001). For vocal features, we used PRAAT (Boersma and Weenink 2013).

The analysis was based on combinations of features of behavior expressing combinations of affective-epistemic states (cf. Allwood et al. 2012), occurring in different stages of conflict episodes in the political debates. These stages will be discussed below in relation to (1) the exhibited behavior of the involved partners (Sect. 3.7) and (2) the different taxonomies of conflict mentioned above (Sect. 3.8).

3.7 Stages of Conflict in Televised Political Debates

The interpretation of conflict in terms of stages is, as discussed above, not straightforward. However, based on the corpus of televised political debates, a number of stages can be proposed for this particular activity.

3.7.1 Stage 1: Early Phase—Pre-conflict/Latent Conflict

This phase is characterized by overtly fairly “neutral” and calm stances. One party talks, making claims, which may contain arguments, that the other party can find offensive. The purpose of the activity is a political debate between persons that can be assumed to be antagonists so it is typically characterized by initial latent conflict. Among the five models of conflict stages, described above, only Brahm’s model recognizes this stage.

Fig. 3.1 Lafontaine starts his contribution (Debate 1)



Lafontaine has just been asked by the TV host what he thinks about the NATO attacks against Libya and starts his answer by gazing at the TV host, leaning against the back of his chair (Fig. 3.1).

3.7.2 *Stage 2: Initial (Confrontative)* *Claim + Challenge/Attack*

In this phase, a participant attacks or challenges the previous or present main speaker, adopting an accusing stance, typically with one hand forward and the index finger raised. The attacker is provocative, sometimes sarcastic and sometimes interrupting the main speaker.

Attacks of this type also reoccur in the following phases from both sides. Among the five models of conflict stages, Glasl's "discussion" and "debate" stages are related to this stage, as are Hocker and Wilmot's "confrontation" and Brahm's "emergence." As we can see, the different models are on different levels of abstraction and focus on different aspects of the interaction.



Fig. 3.2 Kienzle attacks Lafontaine (Debate 1)

After around 30 s, Kienzle tries to interrupt Lafontaine accusing him of abandoning the Libyan rebels. Kienzle leans his upper torso forward and points his index finger at Lafontaine (Fig. 3.2).

Kienzle: “Wenn ich Sie richtig verstehe . . . Wenn ich Sie richtig verstehe . . . a-la, jetzt, kein Wahlkampfreden, kein Wahlkampfreden.” (“*If I get your point . . . If I get your point . . . a-la. No electoral propaganda now. No electoral propaganda.*”)

3.7.3 Stage 3: Response to Accusation

A challenge is usually met by a response. The stance of the responding party is often annoyed, irritated, or even angry. The response can take different alternative forms. It can, for example, be a smile, trying to make the attack (or the attacker) seem ridiculous, irrelevant, or unimportant. Very often, however, the response is a direct counterattack, which can concern the content of the attack (Fig. 3.3a, b above) and/or the right to speak (claiming the floor back). The speaker can also show exaggerated surprise or shock at the attacker’s utterance or impoliteness in interrupting (Fig. 3.4). Finally, the attacked speaker can simply override the attacker by just continuing his/her speech and ignoring the attack (Fig. 3.5).

In relation to the five models of conflict stages, Glasl’s “debate,” Hocker and Wilmot’s “confrontation,” and possibly Brahm’s “escalation” are relevant, if we allow for the fact that the stages in their original form probably in all cases were to be seen as stages in more long-term conflicts than the ones we are considering.



Fig. 3.3 Moratti responding: irritated (a) and also accusing (counterattack) (b) (Debate 3)

Moratti (Fig. 3.3a): “la commissione antimafia in consiglio comunale non avrebbe avuto competenze/noi abbiamo chiesto al prefetto e sulla base di quello che la prefettura ci ha indicato abbiamo preso una decisione” (“*the anti-mafia commission in Milan would have had no powers/we asked the prefect and based on what he told us we took our decision*”)

Moratti (Fig. 3.3b): “credo che l{0}avvocato pisapia queste cose dovrebbe saperle” (“*i think lawyer pisapia should know these things*”)

Fig. 3.4 Roth (*woman second from the left*) responding with shocked surprise/outrage, posing a question as counterattack (Debate 2)



Roth: “Ah! Es ist nicht eine Aufgabe einer Kirche die ethische Begründung für eine Technologie in Frage zu stellen, die nicht beherrschbar ist?!“ (“*Ah! It is not the duty of a Church to question the ethical justification of a technology, which is not controllable?!*”)

Fig. 3.5 Lafontaine overriding the attacker, keeping the floor (Debate 1)



Lafontaine: das ist kein wahlkampfreden das ist eine frage... warum wo + warum... es war... es... (*this is no electoral propaganda this is a question... why wh + why... it was... it...*)

Kienzle then interrupts again and accuses Lafontaine of not answering his question, but instead giving a propaganda speech, his voice raised and his hand raised, pointing his index finger (“keine Wahlkampfreden” “*no electoral propaganda,*” repeated). Kienzle’s contribution overlaps with Lafontaine’s but Lafontaine keeps his turn. He produces this part of his argument raising his voice, moving his upper torso forward in Kienzle’s direction while holding his head upward.

3.7.4 Stage 4: Further Escalation of Conflict

This phase contains continued and often repeated attacks and counterattacks, usually with increasing intensity. Affective-epistemic stances are angry and accusing with behavioral features such as sarcasm or shouting while overlapping other speakers, leaning forward with hand forward, often with the forefinger raised. Considering the five models of conflict stages, Glasl's "debate," Hocker and Wilmot's "confrontation," and Brahm's "escalation" stages are still relevant which reinforce and illustrate that these stages are less temporally fine grained than the stages we are suggesting.

Fig. 3.6 Herles responding to the counterattack from Roth above with anger and sarcasm (Debate 2)



Herles: "Da wird eine Technologie zum absolut Bösen erklärt! Weiche Satan!" ("Then a technology is declared as absolutely evil! Be gone Satan!"/shouting (Fig. 3.6)

Fig. 3.7 Kienzle and Lafontaine arguing about the right to speak (Debate 1)



Lafontaine, irritated, raises his hand and counterattacks Kienzle's (this is not electoral propaganda). Contributions are overlapping all the time. Lafontaine, then, annoyed reminds his interlocutor of good manners: "Herr Kienzle, wenn Sie höflich sind, lassen Sie mich den satz zu ende führen, dann kommen Sie eher dran (*"Mister Kienzle, if you are polite and let me finish my sentence your turn will come sooner"*)." Lafontaine continues, now more vehemently, showing both passionate engagement and anger. After only a few seconds, Kienzle interrupts him again, repeating his accusation (Fig. 3.7).

3.7.5 Stage 5: Climax

The climax in a conflict can contain both parties shouting, leaning forward, and speaking at the same time, with one hand forward and almost standing up (from a sitting position). Comparing with the five models of conflict stages, Glasl's "debate," Hocker and Wilmot's "confrontation," and Brahm's "escalation" with the possible addition of Cornelius et al. "crisis" stage are still the relevant which again illustrate that these stages are less temporally fine grained than the stages we are suggesting.

Fig. 3.8 Climax of the conflict between Kienzle and Lafontaine (Debate 1)



Kienzle interrupts Lafontaine again, now shouting and again pointing at Lafontaine with his arm and hand. Both interlocutors are now shouting, sitting with their upper torsos forward, using one arm/hand with the index finger stretched pointing at the opponent, in a fight to gain the floor and the sympathy of the audience (Fig. 3.8).

3.7.6 Stage 6: Superiority—Having Won and Silence/Hesitation, Having Lost

A conflict sequence in a political debate can be interrupted by the program host or by other speakers. If it continues until one party wins, however, the winning party often exhibits a stance of superiority, looking determined and triumphant, often with raised chin (Figs. 3.9b and 3.10b) and gazing intently at the opponent (Fig. 3.9a) but also at the program host and/or the audience and sometimes also showing a triumphant smile (Fig. 3.10b). Returning to the five models of conflict stages, Glasl's "loss of face," Hocker and Wilmot's "fight or flight," and Brahm's "post-conflict" are possibly relevant. The comparison again points to the differences in perspective built into the five models, where perhaps, the most important difference in perspective is that our suggestion concerns short-term conflict episodes, while the other models, with the exception of Eric Brahm's model which is more neutral from a temporal point of view, concern long-term conflicts.



Fig. 3.9 The winner triumphant (a): Lafontaine (b)

Lafontaine, having counter-accused Kienzle of being cynical, turns his face in the direction of two other participants, i.e., the TV host and another participant in the debate. Then, he checks whether his opponent wants to continue the fight, gazing directly at Kienzle for 3 s (Fig. 3.9a). Kienzle has no more arguments and drops the fight: he is speechless, he does not make any gestures, though he is watching Lafontaine, the winner.



Fig. 3.10 The winner triumphant (a): Roth and Romney (Debates 2 and 4) (b)

Fig. 3.11 The defeated silent/hesitant—Perry (Debate 4)



3.8 Comparing Conflictual Communication in Different Social Activities

3.8.1 *Political Debate, Quarrel Between Neighbors, and Conflict in a Work Group*

As we have suggested above, a relevant question is whether the phases suggested for political debates are also found in conflicts taking place in other social activities, and, if so, how similar or different the phases are in different activities. Two other types of conflict we have examined are “quarrel between neighbors” and “conflict in a work group.” If we compare these three activities, illustrated in the table below, we can see how different the conditions for conflict are in the three selected activities

As we can see in Table 3.2, the initial phase can be similar in the conflict between neighbors and conflict in a work group, but is likely to be different, in terms of whether there is a latent conflict from the beginning, as is the case in a political debate. A latent conflict may perhaps also occur, but need not do so in the other two activities. If we turn to the goal of the activity, there are major differences in what can be achieved and what the best outcome is for the participants in the

three activities. This also applies to the expected result. These differences in goals and expected results will affect the type of conflict that occurs. The presence of an audience and of a leader or mediator is most likely in the political debates and would have a fairly different role in the conflict between neighbors or in a conflict in a work group.

3.8.2 Activity Comparison in Relation to Taxonomies of Conflict

In relation to the taxonomies of conflict, presented in Sect. 3.2, a political debate can be a two-party conflict or involve more participants, but often, there are two main contenders or sometimes two main groups in conflict. The relation between number of participants and the occurrence of bystanders and some type of audience can be dynamic, so that it is sometimes hard to know who is actually involved and who is a bystander or part of the audience. A neighbor conflict also typically involves two main parties (which can be groups), and a work group conflict can be between two or more parties. In the two latter cases, however, there is often no audience, whereas an audience is essential and the main real addressee in a political debate. Thus, many of the “stances” in the political debate, such as pretending outrage, sarcasm/irony and a triumphant look, gazing, and perhaps smiling demonstratively, are meant for the audience and might, for that reason, not be as prominent in the other activities.

The political debate is typically a two-way conflict, while in both the other types of activity, the conflict can be one-way or two-way. Political debates are also clear cases of overt conflicts, where exposing a conflict is actually one of the goals of the activity. The fact that the political debates are televised and in front of an audience gives them a more public and “demonstrative” function than the other two types, which are typically conducted in a small group or just between two persons.

Another related difference is that while political debates typically have a win-lose goal, the other two activities would often both benefit from some kind of solution, compromise, or reconciliation. Even though the other types of conflict can escalate and have a winner, this is less often the optimal solution in these activities, whereas it standardly is in the political debate. Strategies and stances aiming to promote joint solutions, compromises, mediation, etc. are, therefore, not very prominent in the political conflict (even though the moderator might sometimes attempt calming the argument down), but are more important in the other types.

The distribution of power can be symmetrical or asymmetrical in all the activities, depending on other circumstances. In political debates, it is usually known which of the participants has more voters than the other and which participant might be in power, e.g., part of the government, there may also be differences in political experience, thus, power differences are often present.

Political debates represent manifest, actual conflicts, rather than latent or potential conflicts, whereas this need not be as clear in the other two types. The conflict in

Table 3.2 Conditions for communicative conflict in three types of social activities

	Political debate	Quarrel between neighbors	Conflict in a work group
Initial phase	No initial unbiased dialog Latent conflict	Initial friendly dialogue or latent conflict	Initial dialog more common
Goal	Goal to win audience, voters Not to agree	Goal to win argument over some practical problem, e.g., a fence	Goal to carry out a common task, which all will benefit from
Expected result	One party wins	One party wins or compromise or breakdown	One party or the majority takes over or break up in subgroups, compromise or breakdown Task needs to be completed
Audience	Studio and TV audience/voters	No audience Or other neighbors	No audience
Leader/mediator	Talk show host Intervention	No mediator/chairman No intervention	No mediator/chairman, except possible self-selected group member No intervention

a political debate is in a sense necessary. To use the terminology of the taxonomy in Sect. 3.2, it is both permitted and obligatory, as well as certain. These properties are not the same in neighbor conflicts or work group conflicts, which very well can be merely latent and potential/possible, actual and nonpermitted as well as conceivable without being certain.

3.8.3 The Relation Between Activity Differences and Stages/Steps/Phases in Conflict

In summary, the conflicts in political debates in most respects represent very different conditions than conflict in the other two social activities they have been compared with above. Especially the beginning and the end of a conflict episode can be very different—the other two activity types often do not start with claims, instead they can start with behavior from one party which irritates the other party, possibly at first with only covert reactions. In contrast, in the political debates, there are initially usually a number of potentially confrontative claims. The three activities also vary in terms of what responses may be expected. If claims are made, acceptance of the other's claim, avoidance, and prevention of conflict are suitable in the neighbor and work group conflict cases, but not really in the political debate, because of the different purposes of the activity types. Further, even though escalation phases contain similarities in behavior, they also contain differences, depending on the different conditions, i.e., especially on the presence of an audience (in the political debate both a studio and a TV audience), which is the main addressee, and also on the more or less ritualized overt expression of conflict in political debates.

Even if manifested in somewhat different ways, the occurrence of phases of challenge/attack, response, and escalation seems to be common to most overt conflicts in all the three cases, but necessary and “obligatory” only in the political debate. The early phase can be very different between the activities, and the climax and win-lose phases are probably more common in the political debate and have alternatives like compromise and reconciliation in the two other cases.

Thus, the occurrence of stages in conflict as well as their labeling and description has to be related to the social activity in which it is pursued, in order to be detailed enough to capture stages in different types of conflict. We have also seen that the differences between types of conflict have resulted in differences between the different models that have been suggested to describe stages in conflict development and that for this reason, it would be desirable for future models to more explicitly state what type of conflict the model of stages is supposed to describe. Finally, we have suggested a six-stage model to capture conflict escalation in televised political debates.

Acknowledgments The research that has led to this work has been supported by the European Community's Seventh Framework Programme (FP7/2007–2013), under grant agreement no. 231287(SSPNet).

References

- Allwood J (1992) The academic seminar as an arena of conflict and conflict resolution. Gothenburg papers in theoretical linguistics 67. University of Gothenburg, Department of Linguistics
- Allwood J, Chindamo M, Ahlsén E (2012) On identifying conflict related stances. Proceedings of IEEE SocialCom, Amsterdam 2012
- Boersma P, Weenink D (2013) Praat: doing phonetics by computer [Computer program]. Version 5.3.51. <http://www.praat.org/>. Retrieved 2 June 2013
- Brahm E (2003) Conflict stages. In: Burgess G, Burgess H (eds) Beyond intractability. Conflict information consortium. University of Colorado, Boulder, Posted: September 2003
- Cornelius H, Faire S (2007) Everyone can win: responding to conflict constructively. Simon & Schuster, Sydney
- Glasl F (1997) Konfliktmanagement. Ein Handbuch fuer Fuehrungskraefte, Beraterinnen und Berater, 5, erweiterte Auflage. Verlag Paul Haupt, Bern
- Hocker JL, Wilmot WW (1991) Interpersonal conflict. William C. Brown, Dubuque, IA
- Kipp M (2001) [Anvil-agenericannotationtoolformultimodaldialogue](#). Proceedings of the 7th European conference on speech communication and technology (Eurospeech), pp 1367–1370
- Nivre J (2000) Gothenburg Transcription Standard (GTS). University of Gothenburg, Department of Linguistics
- Nivre J (2004) Modified Standard Orthography (MSO). University of Gothenburg, Department of Linguistics
- Noll D (2000) Conflict dynamics. Los Angeles Daily Journal Verdicts and Settlements. 20 Oct 2000

Chapter 4

Revenge and Conflict: Social and Cognitive Aspects

Francesca Giardini and Rosaria Conte

4.1 Introduction

In December 2012, 60,000 l of valuable Brunello di Montalcino produced by Gianfranco Solera were destroyed overnight by an angry former employee who wanted to get back at his boss for the recent firing. The former employee considered the firing as an unfair reduction of his power, and he wanted to make his boss suffer in a way that he considered comparable, that is, by imposing on him a substantial monetary loss (in addition to the fact that six vintages of Solera were definitely lost and nobody will ever be able to taste them). Are worth a lost job 60,000 l of Brunello?

The answer is “yes,” because when an individual wants to take vengeance on someone there is only one account rule: “make him suffer as much as you (i.e., the victim) suffered.” This makes revenge unpredictable and dangerous because it is grounded on a subjective perception of what is right and what is wrong. This feature, we posit, might provide an explanation for the fact that human societies condemn revenge and alternative forms of reaction, such as punishment (Gardner and West 2004; Henrich and Boyd 2001; Henrich et al. 2006) or sanction (Giardini et al. 2010), evolved to solve coordination problems (Andrighetto et al. 2012a). Nevertheless, revenge is still part of our everyday life, and everyone has experienced, surely more than once in a lifetime, a desire or an urge to make someone suffer because they made us suffer.

Hence the paradox of revenge: how could a risky, costly, and disruptive behavior have survived in human societies? How to reconcile the individual desire to get even

F. Giardini, Ph.D. (✉) • R. Conte

Laboratory of Agent-based Social Simulation (LABSS), Institute of Cognitive Science and Technologies, National Research Council of Italy, via Palestro 32 - 00185, Rome, Italy
e-mail: francesca.giardini@istc.cnr.it

to others and the social prescriptions against revenge? If revenge is so negative, what makes it so hard to restrain revenge even in modern societies, in which it is socially disapproved, culturally neglected, and legally prohibited?

There are two general sorts of reasons why we should be interested in the question of what motivates people to take revenge. The first reason is a psychological one. What are the cognitive mechanisms that support a socially discouraged and condemned behavior and thus ensure its persistence? Understanding revenge implies to explain why it is so difficult to resist the temptation to take revenge, and it represents an attempt to explain the desire for revenge that everybody feels and that triggers small vengeful actions in everyday life (Frank 1988).

The second reason concerns social policy and punishment institutions. We are interested not only in explaining the persistence of “cultures of honor” in which revenge is culturally accepted and even prescribed, but we would like to sketch a more general hypothesis about the evolutionary path that made possible the persistence of revenge notwithstanding the risks of feuds, conflicts, and social disruption.

In this chapter, we start from the assumption that revenge is a social behavior that creates a conflict of interest between individuals’ interests and societal needs, and we develop a cognitive theory of revenge, with the aim of offering a solution to this tension between individuals and societies. Our starting point is the following question: why a behavior that is risky (you can be counterattacked), costly (in terms of current and future costs), disruptive, and aimed not at deterrence (this can be a functional effect but it is not necessarily meant by the avenger) but at reducing the aggressor’s power or to make him or her suffer was not selected out by evolution?

Our answer is that revenge, intended as both a desire for revenge and the action itself, is grounded on and exploits the existence of several psychological mechanisms that support revenge and make it difficult to get rid of it. On the other hand, societies tried to modify the behavior, but not the mind-set of the avenger, thus allowing revenge to survive and prosper. This tension is not resolved and it is demonstrated by the fact that revenge, as a desire, a behavior, and a set of mental representations, is still present in contemporary societies.

In this chapter, we start by defining revenge, and we outline its main features also in comparison with two alternative mechanisms for conflict management: retaliation and forgiveness. In the third paragraph, we apply cognitive analysis in order to highlight the relevant sets of beliefs, desires, and intentions behind revenge, and we discuss the relationship between revenge, punishment, and sanction. In Sect. 4.1, we report ethnographic evidence from societies in which revenge is used as a way of dealing with conflicts, with the aim of describing the institutional characteristics of revenge and of identifying the main reasons for the aforementioned tension between individual desires and social prescriptions. Finally, we draw some conclusions about the role played by revenge in human societies and the supporting cognitive mechanisms, and we point to open issues that would deserve further investigation.

4.2 Defining Revenge

Revenge is a social behavior individuals choose with the purpose of evening the score after having suffered an aggression by another individual. Revenge is a counteraggression aimed to reestablish the balance of power between the actors, disrupted by an initial aggression which has to be framed as a *social damage*, intended as an intentional (or deemed intentional) disruption of someone's power. A social damage implies a reduction of power at the expenses of someone who can take revenge on the perpetrator with the aim of restoring the initial balance and of evening the score in a way that the avenger finds adequate. On the same line, Elster (1990) defines revenge as "the attempt at some cost or risk to oneself, to impose suffering upon those who made one suffer, because they have made one suffer" (p. 862).

Revenge is characterized by a special kind of "bookkeeping," which is emotionally loaded and ego-centered because what counts is the amount of suffering experienced by the victim of the initial damage. This makes revenge unpredictable and dangerous, because suffering cannot be evaluated in an objective and general manner. Since the appraisal of the situation and the context in which the offense takes place determine the intensity and kind of reaction, an aggression that the perpetrator would consider as apparently insignificant can represent a serious harm for the victim, who can become a ruthless avenger:

Vengeance actually works in equalizing the suffering. It makes part of the suffering disappear. Not all of it—the loss and the recollection of harm remain—but some of it, the poignancy of it, the loneliness of it, goes, the being-less-than-he or she, the thought of his or her gain. You get even in suffering. (Frijda 1994, p. 274)

This balance of suffering is essential for the avenger, and it can be one of its causes. For the avenger, getting even is a way of restoring equity (Stillwell et al. 2008; Tripp and Bies 1997), and it also works as a form of stress relief and self-affirmation. Making the perpetrator suffer is a way to mitigate the offense and the related emotional suffering, and it can be related with the rewarding aspects of taking revenge, which have well-defined neurological bases. De Quervain and colleagues (de Quervain et al. 2004) addressed the question of whether revenge is rewarding in a neuroimaging study using an economic game in which real money was invested by participants. In this experiment, player B received money from another player and had the opportunity to send something back to the donor. In case player B did not give back anything, player A would have the possibility to punish B by delivering penalty points and reducing him or her payoff. The results demonstrated that participants did punish the other players even at a cost to themselves, thus confirming previous behavioral results (Fehr and Gächter 2000, 2002), but also that punishment was associated with the activation of reward-related areas in our brain. More precisely, the activation took place in the dorsal striatum, which is part of the brain's reward system and is involved in many other reward-related behaviors, such as sex, food, and addictions. This result supports the view

of revenge as a primary motivation of human conduct, but it also highlights that the quest for compensation and the related willingness to make the target suffer have a distinctive neural signature.

4.2.1 *Retaliation and Revenge: Similarities and Differences*

Revenge is not the only kind of reaction to a wrong humans can resort to, and among the several available options, retaliation seems to be the safest one. Revenge and retaliation are closely related and apparently similar, and in everyday language, they are frequently used as synonyms. The Oxford dictionary definition for retaliation is “make an attack or assault in return for a similar attack,” and retaliation can be conceptualized as a form of retribution without the excesses of revenge. Retaliation is immediate and proportionate and its costs are strictly related to the damage suffered; therefore, it seems to be more adequate than revenge. If this is true, then a question arises: why do we not use retaliation, instead of revenge? Does revenge bring about something more than retaliation, thus supporting our hypothesis that humans cannot get rid of vengeful desires and behaviors?

Table 4.1 summarizes the three main dimensions along which retaliation and revenge differ, and shows whether a given element is present or not in these two behaviors. Suffering is the first and foremost difference between revenge and retaliation. The goal of having back something or of repaying a damage received with another damage is not enough to trigger revenge, because the latter implies more than a simple wrong. Kelsen (1943) highlights the importance of retribution in both behaviors, but he specifies that pairing the evil is not enough to talk of revenge and the infliction of an evil in turn is required:

The behavior interpreted as retribution is not clearly distinguished from a mere defensive reaction which arises from a desire for self-preservation or, at least among higher beings, as a countertendency to the causation of pain. It is proper to speak of ‘vengeance’ only if the reaction in question is made with the intent not only to parry the evil but also to inflict an evil in turn, either on the ‘author’ of the evil or on someone associated with him who is thus regarded as collectively responsible. (p. 49)

While in retaliation the action is a simple payback, so the wrong suffered corresponds in quality and/or quantity with the wrong imposed, the avenger wants to repay the suffering and not only the material loss. A kid hitting another kid at the playground in response for a kick is just retaliating, without any vengeful purpose. Barash and Lipton (2011) suggest that revenge, in contrast with retaliation,

Table 4.1 Defining features in retaliation and revenge

	Retaliation	Revenge
Suffering (in addition to retribution)	No	Yes
Proportionality	Yes	No
Focus on the actor	No	Yes

is delayed and disproportionate; thus, it is not a direct and immediate payback, but it is rather a more sophisticated and complex behavior that requires some form of planning.

The retribution principle behind revenge and retaliation can be conceptualized as a social function that is not explicitly represented in retaliation, but it might become the trigger for revenge. When retaliating, the individual simply strikes back, and the imposition of suffering is not necessary. It can be present as a function, which is not explicitly represented in the mind of the actor and is definitely not the reason why he undertakes that action. Few years ago, several newspapers published the story of a betrayed English woman who discovered her husband's infidelities and reacted by selling his sporty and expensive car on eBay for 1 British pound. Was she retaliating against him or was she taking revenge? According to our analysis, selling the car is not aimed to get a quantitatively similar compensation, because the betrayal and the car have completely different values (and the value of the betrayal cannot be even determined) and they cannot be compared at all. The offender was the focus of the action and making him suffer was the goal of the wife who had not repaid her husband with the same currency (betraying him in return and making him know it), but she chose to do something that was, presumably, much worse from the husband's viewpoint. The wife's goal was to rebalance the sufferance and not to reciprocate her husband with the same action, so she chose an action that was supposed to make him suffer, regardless of any apparent similarity or proportionality between the tort and the reaction.

The second important difference between these two phenomena is proportionality. In retaliation, retribution is quantitative and the reaction has to be exactly proportionate to the initial wrong, like in the *lex talionis* or "an eye for an eye." This is the classic example of retaliation in which there is a perfect symmetry between the wrong suffered and that imposed (Vidmar 2001). On the other hand, revenge is qualitative; thus, the avenger is interested in making the other suffer without any regard for proportion between the initial aggression and the reaction. The focus on quality can be accounted for by an important symbolic dimension that we find in revenge, but is completely absent in retaliation: the aggressor should not be paid back with the same currency but the effect should be comparable. The avenger may choose an action that is different from the wrong initially suffered but that triggers the same negative emotions that were experienced by the victim or that the victim expects the aggressor to suffer. A tort that calls for revenge can be either avenged or forgiven, but it cannot be repaid in any other way, because these are the only ways of getting a symbolic compensation. Retaliation can give you back the material object that was damaged by the aggressor, but it cannot repay the symbolic damage.

The third difference consists in the focus of reaction: in retaliation the focus is the action itself, whereas in revenge it is the actor. A retributive reaction implies that what individuals want is to repay the offender with an equal offense, and the suffering is not an explicit goal. Of course their counteraggression can make the other suffer, but this is not the reason why people engage in this kind of behavior. On the other hand, making the other suffer is the goal of the avenger who aims at restoring the initial balance compromised by the aggression. The offender is the

target of revenge because he or she is guilty of the aggression, as witnessed by the fact that it is possible to take revenge against people that did not take part in the aggression but who are related to the aggressor, like the aggressor's family or closer relatives, because they share some common traits. Posner (1980) views this issue the other way around: family obligation to retaliate is needed to make the threat of revenge work as a deterrent.

4.2.2 Forgiveness and Institutionalized Revenge

An alternative hypothesis about the way in which human societies overcame the costs of revenge posits that forgiveness might represent an evolution of revenge (McCullough 2008; McCullough et al. 2013).

The potentially endless increase of the costs of revenge led to the coevolution of a mechanism for inhibiting it and for developing a forgiveness system (McCullough et al. 2013). This implies a set of motivational changes that orient an individual's behavior toward signaling the harm one has incurred without taking revenge. In this view, both revenge and forgiveness serve the same function of increasing others' regard toward the self, but with quite different methods and results. *Revenge and forgiveness, we argue, have complementary biological functions: We posit that mechanisms for revenge are designed to deter harms, and that forgiveness mechanisms are designed to solve problems related to the preservation of valuable relationships despite the prior imposition of harm* (McCullough et al. 2013, p. 2).

The tight relationship between revenge and suffering has led some scholars to believe that revenge and forgiveness are complementary mechanisms evolved to solve recurrent and highly relevant social problems (McCullough et al. 2013).

Although different, revenge and forgiveness share at least the same origin: the perception of a *wrong* suffered, as opposed to a harm, which leads to the desire of compensation for the loss. Miceli and Castelfranchi (2011) propose a cognitive analysis of forgiveness in which they focus on the benefits for the individual, stressing the importance of forgiving as a means for overcoming the negative emotions associated with revenge, like resentment, and for achieving intrapsychic well-being.

Empirical evidence also shows that forgiveness may promote well-being (Mauger et al. 1992; Salman 2002; Subkoviak et al. 1995). Empirical research on forgiveness and mental health has shown the existence of a link between failure to forgive and indicators of poor mental health such as depression and anxiety (Karremans et al. 2003; Maltby et al. 2001). There is also evidence that this association is more pronounced in close relationships, in which the quality of the relationship before the transgression is important (Karremans et al. 2003).

Revenge is seen as a behavior aimed at changing others' beliefs about the avenger by deterring further attacks through the imposition of a cost on them:

Revenge is an effort to compel an aggressor to increase his or her regard for the victim's welfare—essentially, to teach the aggressor that imposing costs of the same size upon the victim in the future (should they be detected) will be met with retaliatory costs. (McCullough et al. 2013, p. 4)

According to McCullough (2008), revenge allows an individual to change other individuals' behaviors toward oneself, in terms of incentivizing benefits and deterring the imposition of costs. Such a mechanism carries costs that can offset its deterrence benefits, thus opening the way to the evolution of a new and more adaptive mechanism: forgiveness. Although different, these accounts share the view of forgiveness as a solution to revenge, highlighting the evolutionary and proximate mechanisms of it, respectively.

In this view, deterrence becomes one of the main goals of the avenger, who undertakes the action in order to change the cost/benefit ratio of the aggression, thus discouraging it in the future. Once individuals discover that revenge can be effective as a deterrent, but it can also lead to disruption of social bonds, they turn to forgiveness as a different way of obtaining deterrence without losing bonds.

Although relevant, we do not consider deterrence as one of the motivations of the avenger. If deterrence were the avenger's goal, a proportionate response would be more effective in obtaining deterrence. However, this is rarely the case and the limited importance of deterrence could partially explain why revenge is more effective in triggering counteraggressions and feuds than in avoiding them.

There are two ways to restrain revenge: an individual may forgive the perpetrator or a society may regulate revenge, by designing institutions aimed at addressing aggressions. Revenge is forbidden in many countries and serious aggressions have to be reported to designated institutions that use juridical tools to sanction the aggression. When this is not possible, because of historical and economic circumstances, like private enforcement of law and high probabilities of punishing offenses, a pure vengeance system may emerge, in which conflicts are managed through private revenge (Posner 1980).

This happens, for instance, in the so-called culture of honor in the southern United States (Cohen and Nisbett 1997; Nisbett 1993; Nisbett and Cohen 1996) or in the inner areas of Albania, where a collection of customary laws, *Kanun*, is used (see Sect. 4.3 for a detailed description). In both countries, revenge is formally banned and conflicts should be resolved by the criminal law enacted by the State, but revenge is still considered as the best way to regain honor and reputation, even if this can lead to imprisonment. In northern Albania and Kosovo, the *Kanun*, a customary set of laws, disciplines people's reactions to murder (blood revenge or *gjakmarrje*) and other offenses (*hakmarrje*), according to the roles and degree of kinship of all the people involved. Blood feud is a self-governing practice that exists parallel to the state authority and it is regulated by the *Kanun*. Book eight of the *Kanun*, entitled *Honor*, addresses the topics of honor, blood, and kinship, stating that an "offence to honor can only be paid by the 'spilling of blood'" (Article 598). Although revenge is legally prohibited and murders and serious crimes should be reported to police, the *Kanun*'s rules are followed and in some areas blood feuds are still present.

4.3 A Cognitive Theory of Revenge

Punishment is generally found in animal societies (Clutton-Brock and Parker 1995; Lorenz 1966), whereas revenge is restricted to human societies (Zaibert 2006) and to some nonhuman primates (Jensen 2010; Jensen et al. 2007). Both in human and nonhuman primates, revenge stems from an immediate reaction to a social damage, i.e., to an arbitrary reduction of power, but there are some features of human revenge that require a sophisticated cognitive machinery. If we want to identify the defining features of revenge, we should start from the desire of making the other suffer, because he made us suffer, as stated by Elster (1990). Psychological accounts of revenge usually focus on the “desire for revenge,” stressing the importance of inflicting an evil, thus equalizing the suffering and regaining self-esteem and public image damaged by the aggression (Frijda 1994).

Furthermore, the desire for revenge is usually considered as an effective means to repair negative mood and relieve the victim’s distress, although some experimental results show that people tend to overestimate the positive consequences of exacting revenge (Carlsmith et al. 2008). When asked to forecast how they will feel after taking revenge, individuals often fail to predict their emotional states, exaggerating the positive emotions due to fairness restoration and underrating the negative feelings they actually experience. This does not prevent people from exacting revenge and from trying to even the score in terms of pain, with little or no regard for how difficult, risky, or costly it can be. The costs can be really high, especially because the counteraggression is aimed at canceling the pain suffered, but the perception of it is absolutely subjective.

Baumeister (1997) uses the term *magnitude gap* to refer to the difference between the avenger and the victim in the perception of scope, importance, and consequences of an aggression. The victim of the initial aggression wants to restore equity, but in doing so, he or she creates another inequity and so on and so forth, in an endless chain of revenge.

Identifying the distinctive configuration of beliefs and goals motivating revenge is essential to set it apart from other forms of reaction to a wrong, like retaliation, punishment, and sanction (for a detailed discussion of the three phenomena, see Giardini et al. 2010). Although there are several overlapping areas among retaliation, revenge, punishment, and sanction, Andrighetto et al. (2012b) proposed to differentiate among them on the basis of few relevant dimensions, namely:

- The wrong suffered: evaluation of an offense depends both on (attributed) intentionality and on the value of the goal(s) frustrated by the aggressor.
- The purpose of the reaction: when deciding how to react to an aggression, individuals consider the goal(s) they want to achieve and then select the appropriate reaction.
- The type of cognitive influencing, i.e., the kind of intended changes in the mind-set of the victim. For instance, the avenger aims at changing the target’s and audience’s beliefs about himself or herself, whereas the punisher aims to act both at the epistemic and motivational levels, by generating in the victim’s mind the

goal—usually under threat of punishment—of abstaining from doing the action that has triggered punishment again.

- The focus of the reaction refers to the agent himself or herself (as it is in revenge), another agent (as it is in punishment), or a norm (as it is in sanction).

We are aware of the fact that, although the underlying sets of beliefs and goals are deeply different, as well as the resulting state of the world, at least from the involved actors' perspective, a punisher and an avenger could perform the same action. However, if we are able to pinpoint the distinctive mental representations of revenge (see below), we will also be able to advance our understanding of this behavior and to try to resolve the tension between the individual desire to take revenge and the social prescription against it.

4.3.1 *The Defining Features of Revenge*

The initial aggression must trigger some kind of suffering, physical, psychological, or both, and this “harm” has to be perceived as intentional. Suffering triggers a *need for compensation*, and this focus on the wrong suffered leads the avenger to concentrate on the goal of compensating her pain, with little or no regard for the risks of future aggressions. Equity restoration is a primary motive, but it can become extremely dangerous, because the avenger has “nothing to lose,” as clearly expressed in the words of an old tribesman from Montenegro:

Revenge means a kind of spiritual fulfilment. You have killed my son, so I killed yours; I have taken revenge for that, so I now sit peacefully in my chair. (Boehm 1984; p. 56)

An avenger can keep destroying material and immaterial goods until he or she finds herself happy with the outcome, no matter how disproportionate this can be in the eyes of the target, and this may lead to feuds. It is hard to assess how commensurate two damages are, since the same wrong can be perceived in two different ways, implying that compensation can be interpreted in very different manners (Kim and Smith 1993). This drive toward achieving compensation can be interpreted as coming from a broader need to believe that the world is a just place in which individuals get what they deserve, the so-called belief in a just world (Hafer and Bègue 2005; Lerner and Simmons 1966; Lerner 1977). Equity restoration (Walster et al. 1978) is linked to fairness and justice as behaviors that are not uniquely human and that may have played a role in the evolution of cooperation. Brosnan (2013) explores behaviors related with justice and fairness in human and nonhuman primates, showing that negative responses to inequity can be found in many species. Having a sense of fairness allows individuals to correctly discriminate between cooperative and uncooperative partners, thus providing useful insights for partner choice. According to this view, justice and fairness evolved to help individuals to correctly evaluate the value of their potential partners in cooperative interactions (Brosnan 2011; Fehr and Schmidt 1999).

There are also cases in which it is not possible to repay the wrong suffered in a corresponding way, because there is nothing comparable with the wrong suffered and no reaction can be *fair enough*. Compensation strictu sensu is not possible when the symbolic dimension is at stake. Honor, respect, and esteem cannot be taken back, but destroying the aggressor's honor is a way to put him or her on the same ground of his or her victim, rebalancing the situation at least in terms of damages. Going back to the Brunello example, destroying wine was not a way to get back the job, or to restore one's own good name, but it was just aimed at inflicting a comparable, in the eyes of the avenger, damage:

To fail to retaliate homicidally in many contexts used to result in severe damage to one's honor, in that the disapproval of the tribal moral community was so intense that it became almost intolerable. [...] Osveta is something born into a man. It has to do with wounds to the soul and the heart. [...] His (of the Montenegrin) is not highly enough developed so that he fully anticipates the consequences of osveta and bloody deeds, but rather he wishes to satisfy his own haughty pride [...]. If a Montenegrin does not take vengeance. . . ., he has no place and no honor among the rest of the Montenegrins. (Boehm 1984, p. 58–59)

We posit that, in order to wreak revenge on someone, an individual should believe that:

- The offense he received was intentional (*belief about the kind of harm*).
- The offender was the main or the unique responsible for it and then liable for punishment (*belief about the perpetrator*).
- There is a material and/or symbolic dimension to be restored in front of the offender but also in front of the audience (*belief about the audience*).

The above set of beliefs represent the epistemic state which is preliminary to motivations and goal setting. According to our cognitive analysis, there are three distinct goals the avenger wants to achieve:

- The goal of damaging the target and reducing his or her power, either at the material or at the symbolic level. Achieving this goal means that the avenger is able to even the score and to go back to the power relationship *before* the aggression, when the individuals were equal under some respect.
- The goal of making the target suffer, thus changing his or her beliefs about the avenger. In this way, the target becomes aware that the avenger does not passively accept the aggression and he or she is able and willing to strike back at the aggressor (influencing the target). Paying back only the material harm would be useless if the avenger would not be able to even the score in terms of suffering, thus reconstituting the power relationship.
- The goal of changing audience's representations about the avenger. In revenge the audience plays a crucial role because the damage suffered is not only material, but it usually has a strong symbolic component. Honor, for instance, is an intangible asset that can be threatened by the aggressor and that can be restored only if there is an audience in front of which the retaliatory action is performed and that recognizes that action as an attempt of restoring the initial situation.

In fact, revenge is not motivated only by the desire of evening the score per se, but achieving this goal is pivotal to the objective of changing the target's and audience's beliefs about the avenger. This is in line with the predictions of the "understanding" hypothesis (French 2001; Gollwitzer and Denzler 2009; Miller 2001; Vidmar 2001), which states that the goal of revenge is achieved if the target understood why revenge was taken on him.

Restoring one's own image in front of the audience is crucial, because the avenger is a "backward looker" (Giardini et al. 2010), who gets stuck in the past until he or she manages to even the score. Vengeance is not pursued to affect the likelihood that the wrongdoer will repeat the aggression in the future, inducing her to cooperate next time or deterring her from further aggressions. Revenge can work as a deterrent, showing that the avenger intends to and is able to counterattack, thus discouraging future attacks, but this is not the primary motive individuals have for payback. Being the victim of a social harm means that one's social power has been arbitrarily reduced and that the resulting lack of symmetry can be fixed only if the target suffers a comparable damage, no matter how much risky or dangerous this retaliation is. The avenger looks back in anger, so that what happened in the past has to be rebalanced, without any concerns for the future.

This is why the goal of changing the target's beliefs, making him or her aware that the avenger does not passively accept the aggression and is able and willing to strike back at the aggressor (influencing the target), is so important. In revenge, the audience plays a crucial role because the damage suffered is not only material, but it usually has a strong symbolic component. This is especially true for "cultures of honor" that have been documented throughout the world (see Daly and Wilson 1988; Nisbett and Cohen 1996). According to Nisbett and Cohen, who greatly contributed to unveil the main psychological and social features of cultures of honor in southern United States (e.g., Cohen 1996; Cohen and Nisbett 1994; Cohen et al. 1996; Nisbett 1993; Nisbett and Cohen 1996), a key element there is that the participant in such a culture is prepared to fight for his or her reputation for strength and toughness. For instance, honor is an intangible asset that can be threatened by the aggressor and that can be restored only if there is an audience in front of which the retaliatory action is performed and that recognizes that action as an attempt to go back to the initial situation.

4.4 Revenge as a Form of "Wild Justice"

Revenge is an individual behavior which may entail considerable risks for groups and societies. On one hand, payback can be seen as "a kind of wild justice, which the more man's nature runs to, the more ought law to weed it out," as stated by Francis Bacon in his *Essays*. On the other hand, revenge can be an effective means to prevent enlargement of conflicts, because of its deterrence effect:

Far from expressing the fundamentally collective character of life in honor societies, vengeance was a tool for preventing conflicts from becoming needlessly and dangerously collective. Knowing that they were subject to punishment if they extended help, kin of would-be aggressors had good reason to discourage violent acts, and kin of offenders had good reason to distance themselves from acts already committed. (Gould 2003, p. 134)

Is revenge functional to the maintenance of societies or should societies get rid of revenge? It is unlikely to find an unequivocal answer, and Posner (1980) is probably right in suggesting that revenge may be partially determined by historical and economic circumstances. When the juridical system is weak and individuals need to protect their families and possessions, a pure vengeance system may appear, and its maintenance depends on the structure of society but also on the presence of additional means for conflict resolution.

4.4.1 *The Yanomamo*

Payback was a fitness-enhancing behavior in traditional, strongly egalitarian, acephalous, kinship-based, or tribal societies, scattered on relatively large regions, like the Yanomamo described by Chagnon (1988, 2013). In his seminal work on warfare and blood revenge among the Yanomamo of Amazonas, Chagnon developed a theory of violence and conflict in which blood revenge serves two main functions. At the group level, groups with a reputation for “swift retaliation” are less threatened and also attacked less frequently. Revenge is equally important at the individual level, and in the Yanomamo society, being successful in exacting revenge could be even translated into higher marital and reproductive success (Chagnon 1988).

Chagnon describes warfare and blood revenge as inevitably linked to conflicts of interest that arise because as humans we need to seize resources from the environment, including our peers, to survive. In such a context, not reacting to appropriation has high material and symbolic costs. In addition to the obvious costs of losing resources, like food, in these societies, not reacting had high symbolic costs, like being downgraded in reputation and thus becoming victims of social avoidance and ostracism. Losing status was equivalent to being condemned to exploitation, starvation, falling within predators’ reach, and death, and a similar fate was presumably shared by one’s own kinship group.

In such an environment, revenge was presumably a rather effective, if not efficient, way of avoiding such lethal circumstances, more spontaneous than other institutions of social control, like third-party enforcement, and more precocious than sophisticated institutions like legal systems. In those contexts, being able to payback a damage or an offense and to signal that “nobody walks over me” were essential traits to avoid exploitation, marginalization, and death. In those contexts, not taking revenge is equivalent to being dishonored, with all the consequences in terms of losing status, power, and the possibility of being chosen as a partner. In other words, losing honor could have been equivalent to being sentenced to social exclusion and then, in extreme cases, to death. On the other hand, if we look at the

social costs of revenge, we will see how disruptive it can be in any society in which, without regulation, revenge inevitably leads to an escalation into feuds, mayhem, and destruction. The Yanomamo tribe offers an interesting example, but there are several kinds of societies and hierarchic structures in which revenge can be less positive at the group level. This is especially true when conflicts of interest arise also outside the group and there is intergroup competition. A group weakened by feuds will be more prone to appropriation by a stronger group with more resources in terms of men and power.

4.4.2 *The Culture of Honor*

Another interesting example of an environment in which revenge is quite common is offered by southern United States, where the so-called culture of honor is still present (Nisbett and Cohen 1996). Although far and characterized by really different historical, economic, and social circumstances, revenge seems to play the same role in the southern United States and in the Amazonas region where the Yanomamo live. In the southern United States, herding was the prevailing activity, introduced by herdsmen from Scotland and Ireland. Herding, more than farming, places an individual at risk for losing his entire resource base to theft. Moreover, the southern United States was a frontier region where the state was almost absent and inhabitants had to create and enforce their own system of order. According to Nisbett and colleagues (e.g., Nisbett and Cohen 1996), such an environment led to the development of an enforcement system in which minor transgressions, intended as personal aggressions, were severely punished. Shackelford (2005) provides an evolutionary explanation for this behavior, claiming that the inputs provided by the environment (herding and lack of state rules) were processed by psychological mechanisms that *may have evolved, however, as solutions to a related adaptive problem that likely was recurrently confronted by ancestral men: theft of a reproductively valuable wife. Theft of a wife might have amounted not to physical theft, per se, but to theft of her reproductive capacity, as in the form of courting her for an extra-pair copulation or raping her* (p. 390).

A set of behaviors defined as warfare among the Yanomamo or as indicators of a culture of honor in southern United States might be the output of the same psychological mechanisms that evolved in response to the adaptive problem of mate retention (see, e.g., Buss 1988; Buss and Shackelford 1997; Flinn 1988). The problem of partner retention and infidelity is universal; thus, the same psychological mechanisms and different environment could have produced the same behavior, i.e., revenge.

4.4.3 *The Kanun*

When social and environmental circumstances change, groups start to enlarge, and a different economy develops, revenge may become too costly to sustain and a need for restraining revenge and regulating it could have emerged. This poses a tension between the individual desire to take revenge and the need of societies to restrain revenge because it has become too dangerous for them. Resolving this tension requires to eradicate a behavior that is supported by evolved psychological mechanisms that can be hardly modified, even when original circumstances that produced them have changed or disappeared. Since it is unlikely that revenge can be eradicated, an alternative and most effective solution would be to restrain it somehow. There are two ways to make revenge less appealing: either limiting the contexts in which revenge may take place, thus transforming revenge into a regulated institution which is not anymore arbitrary (e.g., *Kanun* in Albania), or promoting reactions that do not trigger feuds and may enforce social order in a less costly way, like punishment and sanction (for a discussion, see Andrighetto et al. 2012a).

The *Kanuni I Lekë Dukagjinit* (Gjeov 1989) is a customary set of laws passed down mostly by oral tradition and it prescribes practices of daily life, including rules governing blood feuds. The *Kanun* is applied mostly in northern Albania and Kosovo, and it became more and more popular after the fall of the Communist party. Among other things, the *Kanun* disciplines people's reactions to murder (blood revenge or *gjakmarrje*) and other offenses (*hakmarrje*), according to the roles and degree of kinship of all the people involved. After a killing, the perpetrator and his or her family take refuge in their homes, which are considered inviolate under *Kanun*, for at least 40 days and seek forgiveness. If forgiveness is not accorded to perpetrator, isolation of all the men of the perpetrator's family can continue indefinitely. All blood feuds under *Kanun* involve honor: shirking revenge or taking it without respecting the rules mean that honor cannot be restored, and the whole family or clan is to blame. The *Kanun* puts honor at the core of social life and in doing so it offers a way to measure and weight offenses that allows individuals to even the score in a predictable and regulated manner. Turning individual revenge into a social institution can be an effective strategy to reduce the danger of a behavior based on personal suffering and aimed at inducing a harm proportional to the experienced suffering. This quest for proportionality makes revenge potentially disruptive at the social level, leading to an exacerbation of reactions and to the reciprocal destruction of the actors involved. Controlling revenge was necessary because if it is true, as we assume, that revenge is rooted in our evolutionary history, then incentives are not enough to discourage it and the development of legal and social sanctions became necessary in order to counterbalance the desire for revenge.

4.5 Conclusions

Revenge is a form of counteraggression targeted at someone who previously attacked us. Revenge is a cultural universal, which has been found in every kind of human society, all over the world, and the quest for revenge characterizes epics since the beginning of human history. However, revenge is costly; it is explicitly discouraged by modern societies in which private justice is legally prohibited. Why did the Count of Monte Cristo not give up on his revenge? Why is it so important to take revenge and how did societies manage to discourage this behavior?

In this chapter, we have highlighted the psychological mechanisms of revenge, showing how rooted such a behavior is in individuals' minds. We have posited that revenge results from cognitive mechanisms that were fitness enhancing in those contexts in which avoiding theft, exploitation, and ostracism was essential to survival and that this function is still in place, as demonstrated by the fact that we still use revenge. We have explained why revenge is different from retaliation, in an attempt to bring to light the special character of revenge, its reliance on psychological mechanisms that rule human behavior, and the resulting difficulty to turn it off. We have also sketched the epistemic and motivational representations behind revenge, making explicit the mind-set of the avenger and his or her beliefs and goals. Given these features of revenge, we have discussed the solutions different societies found for inhibiting the escalation of violence, thus allowing societies to control revenge (i.e., only groups that control the escalation of violence could survive outcompeting the others).

It is difficult to pinpoint the consequences that this tension between a natural tendency to take revenge at the individual level and the opposite tendency to deter revenge at the group level produced in the course of human history, but we would like to highlight two main general effects of it. First, revenge had to be restrained by societies in order to prevent escalation of social conflicts. This was done in two main ways: limiting the contexts in which revenge may take place, thus transforming it into a regulated institution (e.g., *Kanun* in Albania, *Codice della vendetta barbaricina* in Sardinia), or promoting alternative ways of reacting and developing new kinds of enforcing institutions, like punishment and sanction. We draw a parallel between the evolution of societies and the development of enforcing institutions: from being plain systems of revenge and retribution imposed by the individual, family, or tribe, they turned into more complex societies in which aggressions are regulated thanks to institutions characterized by a higher concern for *deterrence* and rehabilitation (Andrighetto et al. 2012a). As a deterrent, punishment serves to dissuade people from aggressions and norm violations, thus reducing the frequency and likelihood of future offenses. Deterrence theory suggests that punishment works by modifying the relative costs and benefits of the situation, so that wrongdoing becomes a less attractive option (Bentham 1962; Becker 1968). Unlike revenge, punishment is a reaction intentionally aimed to minimize the likelihood that the aggressor will repeat the wrong again in the future. This becomes

even more relevant in sanctions, where the costs imposed on the wrongdoer are paired with a message, aimed at signaling that that conduct is not approved of.

The tension between the individual desire for revenge and social prescription against it is perfectly explained by an American journalist, Susan Jacoby, who wrote a book on revenge as a form of wild justice, in which she explains the role of institutions for regulating revenge by saying:

The fact that civilized men and women adhere to a social contract requiring them to settle disputes in courtrooms instead of the corral before sundown does not mean that the impulse toward revenge has been eradicated—any more than the institution of marriage implies the disappearance of sexual impulses directed toward anyone other than one's lawful spouse. (Jacoby 1983, p. 12)

4.5.1 *Open Issues*

Before concluding, we would like to point out some open issues in the study of revenge that deserve further investigation.

Revenge as a social stigma. An alternative solution for societies that want to restrain revenge is to put a stigma on it. Taking revenge is a shameful action, it is considered socially unacceptable so that being considered as a vindictive individual is undesirable. How effective is this solution? Under what circumstances is such a solution effective?

The paradox of revenge. Revenge can be paradoxical because it implies a conflict between equity (getting things even) and justice. The wrong suffered is against equity, but taking revenge goes against the principle of not damaging others. However, if the initial aggression is perceived as inadmissible, the desire for revenge becomes legitimate. Is there a paradox between feeling entitled to avenge oneself and the normative belief against it?

Emotions. Studies of emotions in revenge refer to anger, resentment, and hate felt by the avenger, but less is known about other possible emotions. For example, does the avenger feel guilty or ashamed for his or her behavior? On the other side, what are the emotions of the target, i.e., the initial aggressor, and how does the avenger take them into account when deciding whether to act or not?

Revenge as a form of exaptation. An evolutionary explanation of the success of revenge could consider that revenge evolved as a tool for “doves,” i.e., cooperators, who were exploited by “hawks,” considered as free riders which exploits cooperators at no cost to themselves. In an evolutionary perspective, doves are supposedly not able to attack others so revenge can be seen as an exaptation. A hawk producing a social damage to a dove does not expect a counteraggression, and we also hypothesize that aggressiveness is a trait that doves do not possess, but under the selective pressure of revenge, they were able to evolve an urge for revenge and then to repay the initial aggression with the same coin. Supported by a feeling of entitlement and a need for equity, the dove turned into a hawk, and the positive outcome, gaining respect and compensation, reinforced the feeling of being right

and being entitled. This hypothesis, if supported, could lead to new questions, like: Can revenge act as a selective mechanism in enlarging societies? Does it promote stability under certain conditions or does it lead to turmoil and disruption of social order?

References

- Andrighetto G, Conte R, Giardini F (2012a) Institutions evolving through the mind. A cognitive and computational approach. In: Paglieri F, Tummolini L, Falcone R, Miceli M (eds) *The goals of cognition. Essays in honor of Cristiano Castelfranchi*. College Publications, London, pp 345–378
- Andrighetto G, Giardini F, Conte R (2012b). An evolutionary account of reactions to a wrong. COGSCI 2013 - 35th annual meeting of the Cognitive Science Society, Berlin, July 31–August 3 2013. Proceedings, Cognitive Science Society, 2013, pp 1756–1761
- Barash DP, Lipton JE (2011) *Payback: why we retaliate, redirect aggression and seek revenge*. Oxford University Press, Oxford
- Baumeister RF (1997) *Evil: inside human cruelty and violence*. Freeman, New York
- Becker GS (1968) Crime and punishment: an economic approach. *J Polit Econ* 76:169–217
- Bentham J (1962) Principles of penal law. In: Bowring J (ed) *The works of Jeremy Bentham*. Russell and Russell, New York, pp 365–580
- Boehm C (1984) Blood revenge: the enactment and management of conflict in Montenegro and other tribal societies. University of Pennsylvania Press, Philadelphia, PA
- Brosnan SF (2011) A hypothesis of the co-evolution of inequity and cooperation. *Front Neurosci* 5:43
- Brosnan SF (2013) Justice- and fairness-related behaviors in nonhuman primates. *Proc Natl Acad Sci U S A* 110(Suppl 2):10416–10423
- Brosnan SF, Freeman C, De Waal FBM (2006) Partner's behavior, not reward distribution, determines success in an unequal cooperative task in capuchin monkeys. *Am J Primatol* 68(7):713–724
- Buss DM (1988) From vigilance to violence: tactics of mate retention in american undergraduates. *Ethol Sociobiol* 9:291–317
- Buss DM, Shackelford TK (1997) From vigilance to violence: mate retention tactics in married couples. *J Pers Soc Psychol* 72:346–361
- Carlsmith KM, Wilson TD, Gilbert DT (2008) The paradoxical consequences of revenge. *J Pers Soc Psychol* 95:1316–1324
- Chagnon N (1988) Life histories, blood revenge, and warfare in a tribal population. *Science* 239:985–992
- Chagnon N (2013) *Noble savages: my life among two dangerous tribes – the Yanomamo and the anthropologists*. Simon and Schuster, New York
- Clutton-Brock TH, Parker GA (1995) Punishment in animal societies. *Nature* 373:209–216
- Cohen D (1996) Law, social policy, and violence: the impact of regional cultures. *J Pers Soc Psychol* 70:961–978
- Cohen D, Nisbett RE (1994) Self-protection and the culture of honor: explaining southern homicide. *Pers Soc Psychol Bull* 20:551–567
- Cohen D, Nisbett RE (1997) Field experiments examining the culture of honor: the role of institutions in perpetuating norms about violence. *Pers Soc Psychol Bull* 23:1188–1199
- Cohen D, Nisbett RE, Bowdle BF, Schwartz N (1996) Insult, aggression, and the southern culture of honor: an “experimental ethnography”. *J Pers Soc Psychol* 70:945–960
- Daly M, Wilson M (1988) *Homicide*. Aldine de Gruyter, Hawthorne, NY

- De Quervain DJF, Fischbacher U, Treyer V, Schellhammer M, Schnyder U, Buck A, Fehr E (2004) The neural basis of altruistic punishment. *Science* 305:1254–1258
- Elster J (1990) Norms of revenge. *Ethics* 100(4):862–885
- Fehr E, Gächter S (2000) Cooperation and punishment in Public Goods Experiments. *Am Econ Rev* 90:980–994
- Fehr E, Gächter S (2002) Altruistic punishment in humans. *Nature* 415:137–140
- Fehr E, Schmidt KM (1999) A theory of fairness, competition, and cooperation. *Quart J Econ* 114:817–868
- Flinn MV (1988) Mate guarding in a Caribbean village. *Ethol Sociobiol* 9:1–28
- Frank RH (1988) *Passions within reason: the strategic role of the emotions*. Norton, New York, NY
- French PA (2001) *The virtues of vengeance*. The University Press of Kansas, Kansas
- Frijda NH (1994) The Lex Talionis: on vengeance. In: van Goozen SHM, van der Poll NE, Sergeant JA (eds) *Emotions: essays on emotion theory*. Erlbaum, Hillsdale, NJ, pp 263–289
- Gardner A, West SA (2004) Cooperation and punishment, especially in humans. *Am Nat* 164:753–764
- Giardini F, Andrighetto G, Conte R (2010) A cognitive model of punishment, Proceedings of the 32nd annual conference of the Cognitive Science Society. Cognitive Science Society, Austin, TX, pp 1282–1288
- Gjeov S (1989) *Kanuni i leke dukagjinit*. Albanian text collected and arranged by Shtjefn Gjeov. Gjonlekaj Publishing Company, New York
- Gollwitzer M, Denzler M (2009) What makes revenge sweet: seeing the offender suffer or delivering a message? *J Exp Soc Psychol* 45:840–844
- Gould RV (2003) *Collision of Wills: how ambiguity about social rank breeds conflict*. University of Chicago Press, Chicago
- Hafer C, Bègue L (2005) Experimental research on just-world theory: problems, developments, and future challenges. *Psychol Bull* 131(1):128–167
- Henrich J, Boyd R (2001) Why people punish defectors. Weak conformist transmission can stabilize costly enforcement of norms in cooperative dilemmas. *J Theor Biol* 208:79–89
- Henrich J, McElreath R, Barr A, Ensminger J, Barrett C, Bolyanatz A et al (2006) Costly punishment across human societies. *Science* 312(5781):1767–1770
- Jacoby S (1983) *Wild justice: the evolution of revenge*. Harper & Row, New York
- Jensen K (2010) Punishment and spite, the dark side of cooperation. *Philos Trans R Soc Lond B Biol Sci* 365:2635–2650
- Jensen K, Call J, Tomasello M (2007) Chimpanzees are vengeful but not spiteful. *Proc Natl Acad Sci* 104:13046–13050
- Karremans JC, Van Lange PAM, Ouwerkerk JW, Kluwer ES (2003) When forgiving enhances psychological well being: the role of interpersonal commitment. *J Pers Soc Psychol* 84:1011–1026
- Kelsen H (1943) *Society and nature. A sociological inquiry*. University of Chicago Press, Chicago
- Kim SH, Smith RH (1993) Revenge and conflict escalation. *Negotiation J* 9:37–43
- Lerner MJ (1977) The justice motive: some hypotheses as to its origins and forms. *J Pers* 45(1):1–52
- Lerner M, Simmons CH (1966) Observer's reaction to the 'Innocent Victim': compassion or rejection? *J Pers Soc Psychol* 4(2):203–210
- Lorenz K (1966) *On aggression*. Methuen, London
- Maltby J, Macaskill A, Day L (2001) Failure to forgive self and others: a replication and extension of the relationship between forgiveness, personality, social desirability, and general health. *Pers Individ Dif* 30:881–885
- Mauger PA, Perry JE, Freeman T, Grove DC, McBride AG, McKinney KE (1992) The measurement of forgiveness: preliminary research. *J Psychol Christ* 11:170–180
- McCullough ME (2008) *Beyond revenge: the evolution of the forgiveness instinct*. Jossey-Bass, San Francisco, CA

- McCullough ME, Kurzban R, Tabak BA (2013) A cognitive system for revenge and forgiveness. *Behav Brain Sci* 36:1–58
- Miceli M, Castelfranchi C (2011) Forgiveness: a cognitive-motivational anatomy. *J Theory Soc Behav* 41(3):260–290
- Miller DT (2001) Disrespect and the experience of injustice. *Annu Rev Psychol* 52:527–533
- Nisbett RE (1993) Violence and U. S. regional culture. *Am Psychol* 48:441–449
- Nisbett RE, Cohen D (1996) *Culture of honor*. Westview Press, Boulder, CO
- Posner R (1980) Retribution and related concepts of punishment. *Journal of Legal Studies* 9:71–92
- Salman A (2002) Forgiveness: origins dynamics, psychopathology and technical relevance. *Psychoanal Q* 71:175–212
- Shackelford TK (2005) An evolutionary psychological analysis of cultures of honor. *Evol Psychol* 3:381–391
- Shackelford TK, Goetz AT, Buss DM, Euler HA, Hoier S (2005) When we hurt the ones we love: predicting violence against women from men's mate retention tactics. *Pers Relat* 12:447–463
- Stillwell AM, Baumeister RF, Priore D (2008) We're all victims here: toward a psychology of revenge. *Basic Appl Soc Psychol* 30(3):253–263
- Subkoviak MJ, Enright RD, Wu C, Gassin EA, Freedman S, Olson LM, Sarinopoulos I (1995) Measuring interpersonal forgiveness in late adolescence and middle adulthood. *J Adolesc* 18:641–655
- Tripp TM, Bies RJ (1997) What's good about revenge? The avengers' perspective. In: Lewicki RJ, Bies RJ, Sheppard BH (eds) *Research on negotiation in organizations*, vol 6. JAI, Greenwich, CT, pp 145–160
- Vidmar N (2001) Retribution and revenge. In: Sanders J, Hamilton VL (eds) *Handbook of justice research in law*. Kluwer, New York, pp 31–63
- Walster E, Walster G, Berscheid E (1978) *Equity: theory and research*. Allyn and Bacon, Boston
- Zaibert L (2006) Punishment and revenge. *Law Philos* 25:81–118

Chapter 5

Competition and Cooperation in Language Evolution: A Comparison Between Communication of Apes and Humans

Ines Adornetti

5.1 Introduction

In this paper, we analyze the topic of conflict in reference to the evolution of language. Specifically, we examine two key elements involved in conflicting interactions, competition and cooperation, and show how they are involved in the evolution of linguistic skills. We discuss a model of language origins recently proposed by Tomasello (2008) according to which human language is an evolutionary product of the cognitive systems underlying cooperation among individuals in the social group. The core assumption of this model is that the aforementioned fact makes human language qualitatively different from ape communication, which is mainly individualistic because of the competitive nature of nonhuman primates. Our aim in this paper is to call such a model into question by pointing to an “altruism of knowledge” in apes by discussing some recent experimental data on chimpanzee vocal communication. This data allows us to shed light on the evolution of the cognitive mechanisms that underlie the origin of human communication and to develop a more gradualistic and continuistic model of language evolution. We conclude with some general consideration of the necessity to integrate the cooperative model of communication with a wider and more complex conception of human language and cognition.

I. Adornetti (✉)

Department of Philosophy, Communication and Visual Arts, Roma Tre University,
Via Ostiense 234/236, 00146 Rome, Italy

Department of Human Sciences, University of L’Aquila, Viale Nizza 14,
67100 L’Aquila, Italy

e-mail: ines.adornetti@gmail.com; ines.adornetti@uniroma3.it

5.2 Why Do We Communicate? Some Answers from Evolutionary Biology

Theories on the origin of language are necessarily speculative. However, in recent years, the range of acceptable speculation has been narrowed by the recognition that any account of language origins must be consistent with the principles of Darwinian evolution by natural selection (e.g., Corballis 2011, 2013a; Fitch 2010; Hurford 2007; Pinker and Bloom 1990). One of the main problems with an evolutionary account of human language is the apparent level of altruism involved (Desalles 2007; Noble 2000). According to the orthodox position of evolutionary biology, organisms are indeed products of their selfish genes: they do not do things for the good of the group or the species but rather in order to propagate copies of their own genetic material (Dawkins 1976). In such a perspective, language (and cooperative behavior in general) can be difficult to account for. Specifically, the problem is the following: why do speakers freely exchange valuable information when the theory of natural selection predicts selfishness and competition among individuals? In addition, speaking or signaling always has a cost in terms of time and energy and may involve more indirect costs such as exposing the signaler to greater predation risk. Therefore, reaping the benefits of the informative signals of others without paying the costs of signaling themselves could have more advantages (for a discussion, see Noble 2000).

As is well known regarding cooperative behavior in general, evolutionary theory has answered these problems in terms of kin selection (Hamilton 1964) and reciprocal altruism (Trivers 1971). According to the theory of kin selection, an organism supports another (even at a cost to the organism's own survival and reproduction) because it is helping a relative: through cooperative behaviors, the helper contributes to the survival of part of its own genetic heritage, depending on the degree of genetic relatedness with the relative. According to reciprocal altruism theory, an organism offers support to others by behaving in a manner that temporarily reduces its fitness while increasing that of another organism with the expectation that the other organism will act in a similar manner at a later time.

The evolution of cooperation in relation to communication has also received several explanations (e.g., Ackley and Littman 1994; Brinck and Gärdenfors 2003; Knight 1998; Gärdenfors 2003; Noble 2000). Knight (1998), for example, maintained that the main problem in this regard is to explain the evolution of honest signals. Following Krebs and Dawkins (1984), the author started from the assumptions that animals have conflicting interests and that they seek to exploit and deceive rather than share reliable information. Communication can evolve only if there is some mechanism that makes it trustworthy for the other members of the group. In such a case, in which the advantages of defection overtake the costs of cooperation, the only reliable signals are those that are costly to fake because they cannot be imitated by free riders (Zahavi and Zahavi 1997). According to Knight, signals of this kind are rituals: group members demonstrate their allegiance to a common cause by performing costly rituals, allowing the group to believe their signals in the future.

The problem of explaining the evolution of honest information is also recognized by Dessalles (1998, 2000, 2007). He proposed a political account for the origins of language. In his opinion, in order to explain the evolution of linguistic communication, it is necessary to start from the fact that ancestral humans were capable of forming large coalitions (Dunbar 1996). Among humans, coalitions are essential to the survival of individuals (they offer some security to their members) and have an important political dimension. The power of single individuals, in fact, will depend on the number of allies they can acquire. Leadership of a group cannot be exercised without support from at least some of its members. As a consequence, when coalitions are established, individual competition for leadership is replaced by competition among the several coalitions within the group. In this context, according to Desalles, what is important is not physical strength but the ability to enter a successful coalition. The idea is that speech emerged in this context as a way for individuals to select one another when forming alliances. Relevant information may have replaced physical strength as a determining factor in the decision to join a coalition and remain in it. By living in a social group, indeed, individuals gain status from pointing out salient and correct information (about neighbors, about imminent danger, about food) in the environment. Therefore, the original motivation of human language was to trade relevant information for status (for a discussion, see Hurford 2007; Machery et al. 2010).

5.3 In Search of a Cognitive Explanation: Tomasello's Cooperation Model

The models presented briefly so far explain the evolution of communication by pointing mainly at the possible selective pressures that could have driven the evolution of language. It is not our intention here to discuss the evolutionary plausibility of these hypotheses. What we want to highlight is that any naturalistic model of human language origins has to be formulated also in reference to cognitive systems involved in the genesis and evolution of linguistic faculties (e.g., Corballis 2011; Gärdenfors 2003; Ferretti 2013a; Ferretti and Adornetti 2014; Origgi and Sperber 2000; Sperber et al. 2010). One important model for the cognitive foundations of human communication has been recently elaborated by Tomasello (2008). It is centered on the key role of cognitive systems underlying cooperation as the elements that explain the transition from ape communication to human language.

According to Tomasello, human beings are able to communicate because they have unique cognitive ways of engaging with one another socially in general. In particular, human beings cooperate with one another in species-unique ways on the basis of processes of shared intentionality. Shared intentionality can be conceived as behavioral phenomena that are both intentional and intrinsically, irreducibly social because the agent of the intentions and actions is the subject "we" (Gilbert 1989; Searle 1995; Tomasello and Carpenter 2007). The basic psychological underpinning

to participating with others in acts of shared intentionality is the ability to understand others as cooperative agents. As Searle (1990: 415) maintained, this ability “is a necessary condition of all collective behavior.” According to Tomasello, the ability to understand others as cooperative agents can be broken down into two elements: the cognitive skills for creating joint intentions and attention (a common conceptual ground with others) and the social motivations for helping and sharing with others.

Common ground represents the context of communication, that is to say, what is relevant to social interaction (Clark 1996; Levinson 1995). It includes shared knowledge among participants in social interactions, facts about the world, what people generally find salient and interesting, and so on. Common ground is necessary for the receiver to determine both what the communicator is focusing her attention on (referential intention) and why she is doing it (social intention). The critical point is that to construct common ground, people have to put aside their own egocentric perspective on things. Indeed, in the construction of common ground, people have to pursue a common goal together in order to know that together they are focusing on certain things relevant to the common goal.

The other element of shared intentionality is represented by humans’ cooperative social motivations. Tomasello’s idea (2008, 2009) is that humans have cooperative motivations because they have cooperative motivations for communication in the first place.¹ There are three such fundamental motivations that emerge earliest in ontogeny and that are products of phylogenetic processes. These motivations are:

1. *Requesting*: I want you to do something to help me (requesting help or information).
2. *Informing*: I want you to know something because I think it will help or interest you (offering help, including information).
3. *Sharing*: I want you to feel something so that we can share attitudes/feelings together (sharing emotions or attitudes).

The first motivation is a characteristic of the intentional communication of all apes. Informing and sharing, on the contrary, according to Tomasello, seem to be uniquely human. Particularly relevant to the aim of this paper is the second motivation: informing. This motivation (together with the capacity of creating common ground), indeed, is crucial for the evolution of language and represents the element that, according to the author, makes human communication qualitatively different from ape communication.

¹This idea recalls Paul Grice’s principle of cooperation (Grice 1975), which has been a theoretical milestone elaborated in linguistic pragmatics (e.g., Sperber and Wilson 1986). The principle of cooperation can be formulated in the following way: make your conversational contribution what is required at the stage at which it occurs by the accepted purpose or direction of the talk exchange in which you are engaged. On this point, see also Castelfranchi and Poggi (1998) in which Grice’s cooperation principle is conceived as an instantiation, in language, of Trivers’ reciprocal altruism: namely, they posit the existence of an altruism of knowledge. Incidentally, the importance of beliefs for human agents, as their primary route to planning, decision, and action, accounts for why deception is viewed as an aggressive act, a violation of the fundamental principle of *altruism of knowledge*, and of the *natural right* of humans to come to know beliefs relevant for their goals.

Informing comes about because individuals often want to offer help to others without even being requested to; they inform others of things even when they themselves have no personal interest in the information. Informing is a way of offering help because typically I inform you of things that I think you (not I) will find helpful or interesting given my knowledge of your goals and interests (Tomasello 2008: 85). This ability emerges early in ontogeny. An experiment by Liszkowski and colleagues (2006) showed that human infants prelinguistically informed others from as early as 12 months of age by pointing. Specifically, infants used the gesture of pointing to inform another person of the location of an object that the person was searching for. This result suggests that from very early on, humans are capable of conceiving others as intentional agents with informational states and that they have the motivation to provide such information communicatively (see also Tomasello 2009). On the contrary, according to Tomasello (2009: 15–16):

While infants consistently demonstrate understanding of informative pointing, the same is not true of apes. Apes do not point for one another, and when they do point for humans, they do so mainly to get humans to fetch food for them. Indeed, in all observed cases of apes pointing for humans, the motive is directive (imperative). Also, apes who have learned some kind of human-centered communication use it to communicate only with humans, not with one another, and they do so almost exclusively for directive purposes.

The reason for this lies in the fact that apes, that are our closest relatives—chimpanzees in particular—are extremely competitive, and their competitive nature makes it very difficult for them to share a common goal and to participate in collaborative activities (such as communication). Specifically, Tomasello's (2008) idea is that chimpanzees understand their own action from a first-person perspective and that of the partner from a third-person perspective, but they don't have a *bird's-eye view* of the interaction with the joint goal and complementary roles all in a single representational format. So while humans are capable of shared intentionality, chimps do not have the basic psychological underpinning to participate with others in acts of shared intentionality: they are not able to understand others as cooperative agents. As a consequence, chimpanzees are capable of only individual intentionality.

However, apes are capable to engage themselves in group activities. For example, in the wild, chimpanzees sometimes hunt in small groups to capture the red colobus monkey (Boesch and Boesch 1989; Boesch 2005). According to Boesch and Boesch (1989), chimps have a common goal in their hunting and play complementary roles: one individual has to chase the prey in a particular direction, others have to climb the trees to prevent the prey from changing direction, and so on. In Tomasello's (2008; Tomasello et al. 2005) opinion, this explanation is misleading: the group activities of chimps, such as group hunting, are only apparently collaborative activities. He maintained that in this process, each participant is attempting to maximize its own chances of catching the prey without any prior joint goal or establishment of roles. In addition, he affirmed that when chimpanzees engage in group hunting, they do not communicate intentionality about the ongoing activity, either to set a goal or to coordinate roles. Since they are competing in this activity, they do not engage in any intentional communication. Tomasello (2008: 181–184) wrote:

If my most immediate goal is that I capture the monkey unbeknownst to you, then I will not be doing much communicating [. . .] and so, chimpanzee group hunting would not seem to be a highly facilitative context for the emergence of cooperative communication because it is not a truly collaborative enterprise in the narrow way we have defined collaboration here, as joint goals with coordinated plans/roles.

So, since chimpanzees—and apes in general—have forms of only individual intentionality, their communication is mainly individualistic: they communicate only in an imperative way in order to request things relevant to their own scopes and not to freely exchange valuable information. They do not communicate to inform others (chimpanzees are not capable of helping by informing), and they do not comprehend pointing when it is used in an informative manner: they do not seem to grasp an informative communicative intent. For example, when apes were searching for hidden food and a human pointed to a cup to inform them of its location, the apes did not understand (Tomasello 2006). This occurs because “Chimpanzees do not operate with anything like a Gricean principle of cooperation—fittingly, in their natural worlds—and thus they have no basis for making the appropriate relevance inference” (Tomasello 2009: 18). The same is true for ape alarm calls and food calls. They are not generated by an informative intent because when apes detect a predator, for example, they give their alarm calls even if all of the other members of the group are right there looking at the predator and screaming themselves. Similarly, they give food calls when they discover a rich source of food, even if the whole group is with them already. According to this interpretation, apes do not use calls to help others since they give alarm independently of what others know.

5.4 Toward a More Continuistic View: Altruism of Knowledge in Chimpanzees

Although cooperation among humans clearly differs from cooperation among animals, recent more naturalistic studies suggest that the contrasts are not as severe as initially proposed (Brent et al. 2013; De Wall 2009; Seyfarth and Cheney 2012, 2013). For example, chimpanzees in the wild engage in several cooperative actions with long-term social partners that sometimes are risky activities (Mitani 2006). Concerning cooperation in communication, recent studies by Crockford and colleagues (2012) and Schel et al. (2013b) on chimpanzees’ vocal communication highlighted that some of the assumptions of Tomasello’s model do not apply in general.

Crockford and colleagues (2012) used an alarm-call-based field experiment, observing the response of members of a group of wild chimpanzees to a snake model, a viper, positioned on their path of travel. Although snakes are not predators of chimpanzees, they are nevertheless highly dangerous to them. Therefore, providing information about the presence and specific location of a viper will be valuable to others. At the same time, vocal production is costly and may be inhibited if it attracts the attention of predators or hostile individuals.

The results showed that chimpanzees were more likely to give alarm calls in response to a snake in the presence of unaware group members than in the presence of aware group members. According to the authors “chimpanzees keep track of information available to receivers and intentionally inform those who lack certain knowledge [. . .]. [They] communicate missing information that is *relevant and beneficial to receivers*” (Crockford et al. 2012: 145, our emphasis). In other words, chimpanzees are able to monitor the information available to others: they recognize knowledge and ignorance in others and control vocal production to selectively inform them. They inform ignorant group members of danger with such reasoning as “I know something that you don’t know, and I know that this information is useful to you.” At the same time, the receivers are able to understand the informative intent of the signalers, even if not directed to themselves. After an individual produced alert calls to inform another individual that was behind his shoulder and some 10 m away and that had not seen the snake, the ignorant chimp stopped traveling, revealing to have grasped the communicative intent of the signaler.

Similar results were obtained by Schel et al. (2013b). The authors presented wild chimpanzees with a python model and found that most alarm calls met key criteria for intentionality. Specifically, the results showed that the alarm calls were produced in the presence of socially important individuals: production was significantly mediated by the friendship between the caller and the arriving individual, with the arrival of friends more likely to be associated with an increase in calling (see also Schel et al. 2013a). Furthermore, the production was often preceded by visual monitoring of the audience with gaze alternations, and individuals were likely to persist in emitting calls until all group members were safe from the predator. As in the experiment by Crockford and colleagues (2012), chimps in this case seemed to be capable of informing others with such reasoning as “I know something that you don’t know, and I know that this information is useful to you.” Chimpanzees’ vocal behavior seems to be, indeed, influenced by prosocial motivations that are intentionally informing others of a danger.

5.5 Concluding Remarks and Future Directions

Considerations made so far seem to contradict some assumptions of Tomasello’s cooperative model, specifically the idea that chimpanzees have forms of only individual intentionality and are not able to help others by informing. The results of the studies discussed so far allow one to argue for continuity from apes to humans by pointing at a kind of altruism of knowledge in apes. It is possible to maintain, indeed, that the ability to communicate in a cooperative way is not uniquely human but has its roots in the communicative abilities of chimpanzees to help by offering information to others. These results are also particularly relevant to debates about the evolution of a theory of mind and the relationship between mental state attribution and language origins. It is not our intention to discuss the long-standing dispute on the possession by apes of a theory of mind (e.g., Call and

Tomasello 2008; Hare et al. 2000; Heyes 1998; Povinelli and Vonk 2003, 2004; Premack and Woodruff 1978; Premack 1988; Seyfarth and Cheney 2013; Whiten 2013). We simply underline the fact that the studies discussed so far on chimps in the wild have explored apes' mind-reading capacities in altruistic contexts rather than in competitive situations. Most previous research, in contrast, was conducted on the basis of the idea that it is the competitive element (generally created in the laboratory) that helped their comprehension. Such an idea was clearly inspired by Humphrey's (1976) "the social function of intellect" (or the Machiavellian intelligence hypothesis). The core idea of Humphrey's proposal is that primate intelligence is primarily an adaptation to the special complexity of social life, such as forming optimal coalitions, and that the evolution of primate cognition should be interpreted in the context of social competition. Taking a similar position, Krebs and Dawkins (1984) argued that animals indeed evolved to best guess the minds of others to manipulate them better. The results of Crockford and colleagues (2012) and Schel et al. (2013b) cast new light on this hypothesis or at least suggest that future research on apes' cognitive abilities take greater account of experimental contexts other than competition.

The ability of chimps to inform others is also relevant to the relationship between mental state attribution and language evolution. The fact that a theory of mind has a key role in language origins is not controversial (e.g., Corballis 2011; Dunbar 1996; Gärdenfors 2003; Seyfarth et al. 2005; Sperber and Origgi 2010). Indeed, several scientists have argued that a crucial stage in the evolution of language occurred when individuals began producing vocalizations with the goal of informing and thereby reducing ignorance in others (e.g., Pinker 1994; Seyfarth and Cheney 2010). As we have seen, chimpanzees are capable of it. However, chimpanzees don't speak. The question that arises from what has been discussed so far is the following: why did apes not develop language despite the fact that they are able to communicate to inform others? Trying to respond is, of course, extremely complex, and it would deserve another paper. To delineate just a schematic answer, we can posit that although a theory of mind is a cognitive device necessary for the origins of language, it is not a sufficient condition to explain the transition from animal communication to human language. In our opinion, to explain this transition, it is necessary to focus on a specific element that distinguishes human language from animal communication, namely, its flexible and creative use (language is more than informing others). Recently, it has been recognized that to explain this aspect, it is necessary to refer to an integrated network of cognitive systems including not only mindreading system, but also systems of mental time travel (the ability to project oneself into the past and future) (Corballis 2011, 2013b) and mental space travel (the ability to navigate space) (Ferretti 2013b). From this perspective, the transition from animal communication to human language would rely on the adaptive reorganization of this particular cerebral network (Ferretti and Adornetti 2014) that today has a crucial role in language processing (Ferretti et al. 2013; Ferretti and Adornetti 2011; Ferretti and Cosentino 2013). Therefore, in future research on language evolution, the cooperative model should be integrated with a more complex and articulate view of human language and cognition.

References

- Ackley DH, Littman ML (1994) Altruism in the evolution of communication. In: Brooks RA, Maes P (eds) *Artificial life IV*. MIT Press, Cambridge, pp 40–48
- Boesch C (2005) Joint cooperative hunting among wild chimpanzees: taking natural observations seriously. *Behav Brain Sci* 28(5):692–693
- Boesch C, Boesch H (1989) Hunting behavior of wild chimpanzees in the Tai National Park. *Am J Phys Anthropol* 78(4):547–573
- Brent LJ, Chang SW, Gariépy JF, Platt ML (2013) The neuroethology of friendship. *Ann N Y Acad Sci*. doi:[10.1111/nyas.12315](https://doi.org/10.1111/nyas.12315)
- Brinck I, Gärdenfors P (2003) Co-operation and communication in apes and humans. *Mind Lang* 18(5):484–501
- Call J, Tomasello M (2008) Does the chimpanzee have a theory of mind? 30 years later. *Trends Cogn Sci* 12(5):187–192
- Castelfranchi C, Poggi I (1998) *Bugie finzioni sotterfugi. Per una scienza dell'inganno*. Roma, Carocci
- Clark HH (1996) *Using language*. Cambridge University Press, Cambridge
- Corballis MC (2011) *The recursive mind: the origins of human language, thought, and civilization*. Princeton University Press, Princeton
- Corballis MC (2013a) Toward a Darwinian perspective on language evolution. In: Coello Y, Bartolo A (eds) *Language and action in cognitive neuroscience*. Psychology Press, New York, pp 33–58
- Corballis MC (2013b) Wandering tales: evolutionary origins of mental time travel and language. *Front Psychol* 4:485
- Crockford C, Wittig RM, Mundry R, Zuberbühler K (2012) Wild chimpanzees inform ignorant group members of danger. *Curr Biol* 22(2):142–146
- Dawkins R (1976) *The selfish gene*. Oxford University Press, Oxford
- De Waal F (2009) *The age of empathy: nature's lessons for a kinder society*. Random House Digital Inc, New York
- Dessalles JL (2007) *Why we talk*. Oxford University Press, Oxford
- Dessalles JL (1998) Altruism, status, and the origin of relevance. In: Hurford JR, Studdert-Kennedy M, Knight C (eds) *Approaches to the evolution of language: social and cognitive bases*. Cambridge University Press, Cambridge, pp 130–147
- Dessalles JL (2000) Language and hominid politics. In: Knight C, Studdert-Kennedy M, Hurford J (eds) *The emergence of language*. Cambridge University Press, Cambridge, pp 62–79
- Dunbar RIM (1996) Grooming, gossip, and the evolution of language. Faber and Faber, London
- Ferretti F (2013a) Was there language outside *Homo sapiens*? A cognitive perspective. *J Anthropol Sci* 91:249–251
- Ferretti F (2013b) Navigation, discourse and the origin of language. In: Brambilla P, Marini A (eds) *Brain evolution, language, and psychopathology in schizophrenia*. Routledge, New York, pp 22–32
- Ferretti F, Adornetti I (2011) Discourse processing and spatial navigation. In: Kokinov B, Karmiloff-Smith A, Nersessian NJ (eds) *European perspectives on cognitive science*. New Bulgarian University Press, Sofia
- Ferretti F, Adornetti I (2014) Against linguistic Cartesianism: toward a naturalistic model of human language origins and functioning. *Lang Commun* 37:29–39
- Ferretti F, Cosentino E (2013) Time, language and flexibility of mind. The role of mental time travel in linguistic comprehension and production. *Philos Psychol* 26(1):24–46
- Ferretti F, Adornetti I, Cosentino E, Marini A (2013) Keeping the route and speaking coherently: the hidden link between spatial navigation and discourse processing. *J Neurolinguist* 26(2):327–334
- Fitch WT (2010) *The evolution of language*. Cambridge University Press, Cambridge

- Gärdenfors P (2003) *How homo became sapiens: on the evolution of thinking*. Oxford University Press, Oxford
- Gilbert M (1989) *On social facts*. Routledge, New York
- Grice HP (1975) Logic and conversation. In: Cole P, Morgan J (eds) *Syntax and semantics volume 3: speech acts*. Academic, New York, pp 41–58
- Hamilton WD (1964) The genetical evolution of social behaviour. *J Theor Biol* 7:1–52
- Hare B, Call J, Agnetta B, Tomasello M (2000) Chimpanzees know what conspecifics do and do not see. *Anim Behav* 59:771e786
- Heyes CM (1998) Theory of mind in nonhuman primates. *Behav Brain Sci* 21:101e134
- Humphrey NK (1976) The social function of intellect. In: Bateson PPG, Hinde RA (eds) *Growing points in ethology*. Cambridge University Press, Cambridge, pp 303–317
- Hurford JR (2007) *The origins of meaning*. Oxford University Press, Oxford
- Knight C (1998) Ritual/speech coevolution: a solution to the problem of deception. In: Hurford JR, Studdert-Kennedy M, Knight C (eds) *Approaches to the evolution of language: social and cognitive bases*. Cambridge University Press, Cambridge, pp 68–91
- Krebs JR, Dawkins R (1984) Animal signals: mind-reading and manipulation. In: Krebs JR, Davies NB (eds) *Behavioural Ecology: an evolutionary approach*, vol 2, Blackwell Scientific, Oxford UK, pp 380–402
- Levinson SC (1995) Interactional biases in human thinking. In: Goody E (ed) *Social intelligence and interaction*. Cambridge University Press, Cambridge, pp 221–260
- Liszkowski U, Carpenter M, Striano T, Tomasello M (2006) 12- and 18-month-olds point to provide information for others. *J Cogn Dev* 7(2):173–187
- Machery E, Dessalles JL, Cowie F, Alexander J (2010) Symposium on J.-L. Dessalles's *Why we Talk* (OUP, 2007): *Precis* by J.-L. Dessalles, commentaries by E. Machery, F. Cowie, and J. Alexander, Replies by J.-L. Dessalles. *Biol Philos* 25(5):851–890
- Mitani J (2006) Reciprocal exchange in chimpanzees and other primates. In: Kappeler PM, van Schaik CP (eds) *Cooperation in primates and humans: mechanisms and evolution*. Springer, Berlin, pp 101–113
- Noble J (2000) Cooperation, competition and the evolution of prelinguistic communication. In: Knight C, Studdert-Kennedy M, Hurford J (eds) *The emergence of language*. Cambridge University Press, Cambridge, pp 40–61
- Origg G, Sperber D (2000) Evolution, communication and the proper function of language. In: Carruthers P, Chamberlain A (eds) *Evolution and the human mind: modularity, language, and meta-cognition*. Cambridge University Press, Cambridge, pp 140–169
- Pinker S (1994) *The language instinct*. William Morrow and Sons, New York
- Pinker S, Bloom P (1990) Natural language and natural selection. *Behav Brain Sci* 13(4):707–784
- Povinelli DJ, Vonk J (2003) Chimpanzee minds: suspiciously human? *Trends Cogn Sci* 7(4):157–160
- Povinelli DJ, Vonk J (2004) We don't need a microscope to explore the chimpanzee's mind. *Mind Lang* 19(1):1–28
- Premack D (1988) 'Does the chimpanzee have a theory of mind?' revisited. In: Byrne RW, Whiten A (eds) *Machiavellian intelligence: social expertise and the evolution of intellect in monkeys, apes and humans*. Oxford University Press, Oxford, pp 160–179
- Premack D, Woodruff G (1978) Does the chimpanzee have a theory of mind? *Behav Brain Sci* 1(04):515–526
- Schel AM, Machanda Z, Townsend SW, Zuberbuehler K, Slocombe K (2013a) Chimpanzee food calls are directed at specific individuals. *Anim Behav* 86(5):955–965
- Schel AM, Townsend SW, Machanda Z, Zuberbühler K, Slocombe KE (2013b) Chimpanzee alarm call production meets key criteria for intentionality. *PLoS One* 8(10):e76674
- Searle J (1990) Collective intentions and actions. In: Cohen P, Morgan J, Pollack M (eds) *Intentions in communication*. MIT Press, Cambridge, pp 401–415
- Searle J (1995) *The construction of social reality*. Free Press, New York
- Seyfarth RM, Cheney DL (2010) Production, usage, and comprehension in animal vocalizations. *Brain Lang* 115:92–100

- Seyfarth RM, Cheney DL (2012) The evolutionary origins of friendship. *Annu Rev Psychol* 63:153–177
- Seyfarth, R.M. & Cheney D.L. (2013). Affiliation, empathy, and the origins of theory of mind. *PNAS*, 110(Supplement 2), 10349-10356.
- Seyfarth RM, Cheney DL, Bergman TJ (2005) Primate social cognition and the origins of language. *Trends Cogn Sci* 9(6):264–266
- Sperber D, Origgi G (2010) A pragmatic account of the origin of language. In: Larson RK, Déprez V, Yamakido H (eds) *The evolution of human language: biolinguistic perspectives*. Cambridge University Press, Cambridge, pp 124–132
- Sperber D, Wilson D (1986) *Relevance: communication and cognition*. Harvard University Press, Cambridge, MA
- Sperber D, Clément F, Heintz C, Mascaro O, Mercier H, Origgi G, Wilson D (2010) Epistemic vigilance. *Mind Lang* 25(4):359–393
- Tomasello M (2006) Why don't apes point? In: Enfield N, Levinson SC (eds) *Roots of human sociality: culture, cognition and interaction*. Berg, New York, pp 506–524
- Tomasello M (2008) *Origins of human communication*. MIT Press, Cambridge
- Tomasello M (2009) *Why we cooperate*. MIT Press, Cambridge
- Tomasello M, Carpenter M (2007) Shared intentionality. *Dev Sci* 10(1):121–125
- Tomasello M, Carpenter M, Call J, Behne T, Moll H (2005) Understanding and sharing intentions: the origins of cultural cognition. *Behav Brain Sci* 28(5):675–690
- Trivers RL (1971) The evolution of reciprocal altruism. *Quart Rev Biol* 46:35–57
- Whiten A (2013) Humans are not alone in computing how others see the world. *Anim Behav* 86:213–221
- Zahavi A, Zahavi A (1997) *The handicap principle*. Oxford University Press, New York

Chapter 6

The Price of Being Social: The Role of Emotions in Feeding and Minimizing Conflicts

Alessandra Chiera

6.1 Rationality Over Feeling?

Since the time of Aristotle, humans have been considered to be social beings in nature. More interestingly, the social label seems to apply in a special way to them to the extent that it makes the human species completely different from any other one. As the Greek philosopher points out in *Politics*,

Man is by nature a social animal; an individual who is unsocial naturally and not accidentally is either beneath our notice or more than human. Society is something that precedes the individual. Anyone who either cannot lead the common life or is so self-sufficient as not to need to, and therefore does not partake of society, is either a beast or a god.

In this sense, in spite of the fact that animals are endowed with certain social abilities, those held by humans are of a different nature. Though sharing the idea that sociality represents a peculiar trait of human nature, in this article, we aim to probe the biology of human social relations by defining them in continuity with the capacities of other animals. For this purpose, we will make reference to the debate concerning the nature and the mechanisms underlying social experience in a phylogenetic perspective.

From an evolutionary point of view, a quite common consideration is that sociality has been selected because it promoted the organism's fitness by weakening conflicts among individuals (e.g., Tomasello 1999, 2009). According to many theories, eusociality—namely, the highest form of social organization characterized by hierarchical social groups—functions as a tool that minimizes conflicts (Sapolsky

A. Chiera (✉)
Roma Tre University, Rome, Italy

Roma Tre University, via Ostiense 234, 00145 Rome, Italy
e-mail: ale.chiera@gmail.com

1999): in this view, hierarchical constraints contain the natural competitive instincts. This fact has a very important consequence, that is, long-term cooperation requires individuals in survival terms. Specifically, individuals living in a group gain benefits from cooperating in various activities, for instance, in locating food, rearing offspring, or identifying predators (Aureli and de Waal 2000). These benefits explain the sacrifice of individual interests in favor of a collective interest.

Nevertheless, group life also entails costs. Living in close proximity to other members of the same species implies the simultaneous exploitation of resources, the coordination of activities, the negotiation of the status of group members, and several conditions that make competition and clashes of interests likely (Aureli and de Waal 2000). In this way, social stratification may be a source of new conflicts. To this extent, in order to preserve the benefits of group living, individuals need to maintain a balance between cooperative components and competitive components inherent to the organism—actually, in evolutionary terms, the organism's survival, namely, the selfish interests, matters above all.

A key question concerning this topic is the following: which cognitive mechanisms lead the capacity of keeping this balance? Traditionally, and intuitively, a prevalent theoretical frame of reference has claimed that rational argumentation plays the leading role in negotiation. More generally, the overriding function of rational argumentation has been thought to deal with all high mental capacities, especially including negotiation among others. For instance, Aristotle in *Ethics* regards emotion with suspicion comparing it to the foolish slave and conceiving reason as the wise master. This heavy dichotomy between rationality and feeling to the detriment of the latter has been in part overcome; a solid tradition established by Darwin (1872) underlines the crucial role of emotions in interactional cohesiveness (Gratch et al. 2006; Shamay-Tsoory et al. 2002) and conflict management (Seehausen et al. 2014) as forces that bind and drive social groups. In particular, the embodied perspective (e.g., Iacoboni 2008; Rizzolatti et al. 2004) has emphasized the constitutive link between emotions and the construction of a “we-centric space” (Gallese 2009). Specifically, a mechanism of emotional resonance allows people to capture each other's mental world in a direct way, representing a biological immediate social glue.

A heavy evidence consistent with the hypothesis that emotions are preconditions of social abilities comes by some studies that investigate emotions in psychopathological conditions. The works of Jonathan Cole (1998) on subjects with Mobius syndrome have shown the interrelation between the ability of expressing and comprehending emotional states and the development of social interactions. This congenital form of facial paralysis makes people unable to express their emotions through their face; in consequence of this impairment, the subjects do not live a normal social life. Some of them report the feeling of being spectators rather than protagonists of their social experiences. Moreover, Shamay-Tsoory and colleagues (2002) have suggested that subjects with ASD exhibit difficulties in the emotional field that in turn might be a cause of their social impairment.

This intertwined relation between sociality and emotions appears to be widespread in the animal kingdom. In fact, glancing at the phylogenesis, an

important correlation comes to light between emotional competence and social competence: as complex emotional abilities develop, more complex societies and social institutions emerge (Gratch et al. 2006; Tomasello 2009; de Waal 1996a). For instance, Tomasello states that a specific emotion, namely, empathy—the aptitude to think “he is like me” that Nagel (1986) calls “the view from nowhere”—constitutes the strongest social glue. In his opinion, empathy represents a typically human emotion that marks the dichotomy between the human species and all the rest of animals. From our perspective, this mechanism might have had a key role in driving the transition from basic to advanced social relationships. In spite of this claim, our idea is that empathy can be defined in terms of continuity (see Sect. 6.2). We will analyze this specific aspect but at present what is important to underline is that emotions regulate social interactions according to a hierarchy of levels. In other words, specific emotions are involved in increasing or deterring social relationships. For instance, Tiedens and Linton (2001) argue that people are influenced by the degree of certainty associated with a particular emotion. Thus, they claim that emotions such as happiness and anger are associated with certainty, whereas emotions such as hope and fear are associated with uncertainty. Happiness creates optimistic expectations that cooperation is likely, and fear creates the expectation of a competitive negotiation. In this way, emotions lead the trust or mistrust in developing cooperative or competitive strategies.

In light of these indications, the ability to modulate and regulate emotions amounts to an essential condition in order to make a group stable. In other words, emotions define sociality in a constitutive way. In keeping with these observations, we will side the role of emotions in enhancing the social negotiation process. On the other hand, a further main purpose of this article is to show that emotions play an additional and competitive important role, that is, feeding new forms of conflict. This hypothesis is justified by the dual nature of emotion (e.g., Ekman et al. 1983; Haidt 2003): on the one hand, emotional systems facilitate collaboration probably because they are out of conscious control and so difficult to “fake” (Ekman et al. 1983) and, therefore, provide reliable communication signals among conspecifics; on the other hand, they are embodied mechanisms which lead and favor individual behavior, preparing the organism for an appropriate response to salient stimuli (Levenson 1994). Actually, from an individual point of view, the emotional states have a basic function of directing attention to what happens in the environment preparing the organism to react appropriately. In this sense, Darwinian natural selection has fostered the evolution of this immediate answer as a means of quick evaluation and measure of the perceived situation (González et al. 2009). In other words, emotion signals interests and such interests may be both selfish and collective—that is, they may condition the subjective qualitative experience designed to defend the individual or they may promote the connection with others. Stressing on this constitutive duplicity, we will use emotional devices as a case study to investigate the relationship between conflictual and cooperative tendencies inherent to social nature.

6.2 Rituals and Emotions

The emotional phenomenon is particularly linked with a specific mechanism that presides over the maintenance of social order, namely, ritual. Our early assumption is that, by taking a phylogenetic perspective and looking at the animal kingdom, one of the most powerful mechanisms by which groups may be formed and coordinated seems to be represented by ritual activity. For instance, many primates use the ritual of social grooming to cultivate and reinforce social bonds with others upon whom they rely for coalitional support (Cords 1997; Dunbar 1991). Further, as we will discuss shortly, many social animals make use of emotionally charged rituals to respond to social conflicts (e.g., Aureli and de Waal 2000; Castles and Whiten 1998; de Waal and Lanting 1997). As it is generally well known, rituals are affiliative behaviors which also characterize human societies. They include various forms of worship and religious organizations, ceremonies and inaugurations, marriages, funerals, parties, symposia, as well as common acts like handshaking and saying hello. Considering its pervading diffusion in various social structures, from our point of view, a ritual can be considered as a hand lens in order to explore how the opposite forces embodied by emotions bind together and divide social groups.

Firstly, why does the brain create rituals? Many scholars have focused on complex human rituals such as religion describing the nature of ritual in general as a psychological invention constructed by humans in order to live with the fear of death and the uncertainty of things. In the wake of this belief, some theories (e.g., Lienard and Boyer 2006) have claimed that ritual behavior has an ancillary function and thus in evolutionary terms can simply be defined as an “accident” caused by the malfunctioning of cognitive devices developed for other purposes rather than a sheer natural phenomenon. Though such claims are so widespread, many scholars point out that ritual behaviors are rooted in the biology of brain also because of their social adaptive value. To this extent, social animals at different degrees are biologically bound to perform rituals. In particular, according to a class of models pertaining to philosophical anthropology, namely, “social solidarity theories” (Alcorta and Sosis 2005), in the evolutionary history, ritual has emerged in order to support the construction of social structure binding groups together. For instance, Bering (2006) argues that ritual may have had a role in shaping a propensity for altruism starting from the feelings of empathy and attachment developed in collective ritual practices. More generally, as ritual represents a pervasive phenomenon in social species, the idea that it is a secondary “gadget” seems to be a little likely (Ferretti and Adornetti 2014). How could cognitive mechanisms that make organisms to perform time-consuming rituals be selected for, unless these activities played some evolutionary functions (Bering 2006)?

The idea that the ritual phenomenon might represent a binding tool promoting social cognition meets reliable clues. Actually, the transposition of ordinary behaviors in ritualized displays primarily concerns the communication of social information (Rowe 1999) as in the courtship context. Although these ritualized forms involve time and resource costs, they offer important evidence about conspecifics

and contribute to limit conflicts. What does this benefit consist in? In this view, the answer is that some kinds of stereotyped behaviors performed within group dynamics are able to minimize the effects of intraspecific aggressiveness and feed social confidence. To this extent, these elements are at the base of stable communities, or, in other words, societies build upon this wide repertoire of stereotyped components. Such idea is consistent with Lorenz's model that conceived ritualization as the main regulator of aggression and hence as the foundation of the normative system. In light of this framework, ritual evolved in response to the selective pressures of the social environment and represents a bedrock of stable social structures.

An evidence in support of this interpretative model comes from rituals observed in chimpanzees and baboons, namely, *reconciliation*—the first inter-opponent contact—and *consolation*, the contact between a recipient of aggression and a third individual. With respect to reconciliation, de Waal and Lanting (1997) have underlined that after a fight, individuals come in contact with each other in the immediate aftermath of conflict and preferentially with the antagonist. Reconciliation consists of specific behavior patterns like kissing, holding out a hand, submissive vocalization, and embracing. Aureli and de Waal (2000, p. 16) describe the phenomenon with these words:

In the course of a charging display, the highest-ranking male fiercely attacked a female. This caused great commotion as other apes came to her defense. After the group had calmed down, an unusual silence followed, as if the apes were waiting for something to happen. This took a couple of minutes. Suddenly the entire colony burst out hooting, and one male produced rhythmic noise on metal drums stacked up in the corner of the hall. In the midst of this pandemonium, two chimpanzees kissed and embraced.

The result is the restoration of tolerance and the conservation of group cohesion (Castles and Whiten 1998). Cords (1992), for instance, has shown that monkeys engaged in reconciliation were more tolerant of each other's proximity close to an attractive resource than individuals that had not reconciled. To this extent, the evolution of releaser mechanisms made of fixed ritualized action sequences is critical for the preservation of group living. Eibl-Eibesfeldt (1970) had already underlined the role of appeasement gestures in animals, particularly in many species of birds: the defeated animal displays a submissive behavior that contributes to inhibit aggression, providing a foundation for bond formation.

For our purpose, the interesting fact that we intend to highlight is that reconciliation can be explained by variation in the underlying emotions (Aureli and de Waal 2000). Actually, after a conflict, an increase of anxiety that decreases after reconciliation is observed. Other examples confirm that the disturbance and the following restoration of a relationship are reflected in emotional responses. The affiliative contacts occurring in conflict resolution are likely to be associated with positive sensations: gentle touching causes relaxation and a reduction in heart rate in humans and other primates (Feh and de Mazières 1993), and allogrooming in monkeys decreases heart rate and reduces tension (Aureli et al. 1999). The same results arise from playback experiments where baboons were played the threatening screams of a dominant member that had recently charged them (Cheney et al. 1995).

The recording elicited less strongly subjects' responses when the higher-ranking opponent had engaged in friendly interactions of reconciliation after a fight than when the opponent had not shown a friendly behavior. Hence, reconciliation plays a crucial role in mollifying low-ranking members and restoring the relationships to normal tolerance levels thanks to the realignment of emotional states.

Concerning consolation, de Waal and Aureli (1996) have discovered that ritual contacts between a recipient of aggression and individuals that were present at the fight happen more often in the first moments after the conflict. Moreover, these first contacts consist of a greater number of embraces and touches than later contacts. For instance, macaques exhibit an unsolicited consolation hugging other members who had been molested by a third individual (de Waal 1996a); the behavior has the effect to help the recipient of assault to mitigate distress. To spontaneously provide consolation is thought to require some level of emotional perspective-taking, which allows the bystander to both recognize the emotional state of the victim and to provide the appropriate response to reduce distress. In this sense, consolation is a complex ritual capacity deeply tied to emotions.

Hence, the widespread rituals of reconciliation and consolation have the function of reestablishing relations and fostering social cohesion by limiting the effects of aggressiveness and raising forbearance. In this view, collective rituals seem to bind groups together specifically because of the emotionally charged symbols of which they are made of (Alcorta and Sosis 2005). In particular, emotions provide a tool for the synchronization of motivational states among participants, which in turn facilitates the creation of cooperative alliances.

Human rituals, although more complex, seem to undertake similar functions (Sosis and Bressler 2003): just reflecting on our everyday life, we see that our need to overcome competition leads us to employ various rituals, which represent critical mechanisms of conflict management. What ritual participation does is increasing confidence and cooperation among members by feeding a group identity. Recently, Levenson (2003a) has argued that in humans, ritual is deeply tied to emotional and empathetic conditions. In particular, ritual practices elicit in participants autonomic body states—involved in the control of visceral functions that are highly unconscious—that provide the context for the creation of *communally* sacred things. It is the shared emotional and empathetic response elicited by such symbols that provides a mechanism for trust and cooperation, motivating and coordinating individual behaviors across time and space. Consistent with these claims, some studies (d'Aquili and Newberg 1999; Newberg et al. 2001, 2003) have investigated the brain function of people engaged in the ritual of prayer and meditation. The interesting discovery is that these collective activities trigger a unique pattern of cerebral activation, namely, the sensation to belong to a oneness corresponds to a decrease in the activity of the left posterior parietal lobe and a wide general increase in the right hemisphere associated with unconscious emotional states. In other words, to participate in a collective ritual as prayer is proved to involve the deactivation of spatial awareness and the outbreak of ecstatic states associated with strong emotions.

Accordingly, a series of findings proves that rituals have a relevant impact on participants' health and psychological well-being (Anastasi and Newberg 2008; Hummer et al. 1999; Levenson 2003a, b; for a critical review see Lee and Newberg 2005). Actually, religious subjects benefit better health than those with low religiosity (Levin and Schiller 1987). The interplay of ritual and health is specifically associated with the claim that rituals such as religion are important sources of either emotional distress or support (Turner et al. 1995). For instance, ritual contributes to have an effect of reduction on the level of anxiety (Anastasi and Newberg 2008).

On the one hand, animal and human rituals have in common several features as the fact to be tied to emotional components that favor social affiliation as well as individual fitness; on the other hand, the emotional aspect seems to be more complex in human ritual behaviors. It is plausible to suppose that emotions such as empathy might have had a main role in the transition from basic to advanced ritual practices—thus, from simple social structures to more complex social institutions. In fact, empathy arises in different grades: a basic *affective* aspect of empathy is explained by the simulation mechanisms typical of emotional contagion; however, there is a more complex *cognitive* component of empathy that is not limited to sharing but includes a mental understanding of others (Davis 1994). The affective empathy shared with animals might have triggered complex forms of social behaviors as altruism (e.g., see Bering 2006), leading to forms of cognitive empathy underlying more advanced rituals.

This passage leads us to the last part of our argument where we will highlight the implications of a sociality driven by more complex emotionally charged rituals.

6.3 Emotions and “Sacred Values”

As yet, we have highlighted the role of emotionally charged rituals in restoring and maintaining relationships with individuals considered to be particularly important in group dynamics, and we have given support to the hypothesis that these mechanisms might have served the main purpose of fostering more evolved social structures having a bond-strengthening effect. To this extent, we have drawn attention to the positive role of emotions in order to show how they minimize conflicts. However, in our opinion, this point of view renders an incomplete truth about human nature: sociality, especially in its complex forms, has a high price. In our perspective, the same mechanisms that lead the social order—emotionally charged rituals—are also the foundation of new social conflicts.

Actually, another essential way to account for the role of emotions in social life is to look at their character of enhancing conflicts. Probably, this complementary function arises when the two natures of emotions reported previously (see Sect. 6.1) conflict with each other. This process appears to be particularly clear in human societies. Consider the case of religion. Sheikh et al. (2012) present evidence that participation in religious ritual led people to be more likely to consider preferences as *sacred values*. The more people participate in religious

ritual, the more likely they are to treat preferences as protected values (Sheikh et al. 2012). Interestingly, perception of threat to the in-group accentuates the positive relationship between members but feeds intergroup conflicts. Namely, the creation of sacred values by connecting in-group members accentuates intergroup conflict and disputes. Some of the greatest atrocities have been caused by groups defending or advancing their sacred values.

What is about ritual that contributes to in-group cohesion and out-group hostility? The hypothesis carried on in this article is that emotional arousal tied to ritual is exactly the mechanism underlying the forces that bind groups together in opposition to outsiders (Whitehouse 2012). On the one hand, emotions minimize conflict feeding a group identity, and on the other hand, they transform the common values in sacred values pitting the group against out-group individuals (Ginges et al. 2007). As Tomasello states (2009, p. 88), the better way to drive people thinking as a group is to identify some enemies. Several experiments show that humans from an early age quickly divide into groups and side for one's own degrading the other ones. The in-group modality undertakes a protective role, especially in humans that have developed what Cavalli-Sforza and Padoan (2013) call *we-ness*. Identifying common enemies strengthens this preference for one's own group that becomes a strong evolutionary vehicle: it produces a partition between friends and enemies by defining *circles of we*, increasingly more normative, that are driven by different configurations of empathy. Thus, *we-ness* includes both the positive and negative sides of the supremacy of *we-thinking*: a form of empathy inherent to positive *we-ness* inclines us toward the other, perceived as alike, whereas a destructive form leads to exclude the other that is external to the circle. As underlined by Cavalli-Sforza and Padoan (2013), such forms of *we-ness* contribute to breed ideologies like racism. Consistent with these observations, Harris and Fiske (2006) have shown that viewing certain social groups perceived as stranger elicits dehumanized emotions. The subjects involved in the experiment showed to feel disgust in the face of photographs of extreme out-groups similarly of viewing objects. The neural response supports the idea that out-groups may be considered as less than human, in virtue of a natural tendency to prefer one's own group.

In this sense, cooperation seems to have evolved especially to interact with the local group, namely, has been selected in intragroup contexts. Paradoxically, this tendency is one of the greater causes of intergroup conflicts; in other words, the sources of cooperation are also the sources of tribalism.

A recent study by Clay and de Waal (2013) offers interesting evidence in this direction: the authors show that social closeness facilitates empathy in both humans and other animals. Partners sharing stronger affiliative bonds are more likely to make repairing behaviors and to be sensitive to each other's distress. This fact demonstrates that emotions like empathy do not work in an undifferentiated way but are linked to the degree of social relationship: empathy does not produce a conflict mitigation process in general, does not work universally, and does not make us indiscriminately kind and cooperative individuals. Consistent with this observation, for instance, Rozin and colleagues (1999) have shown that an emotion like disgust does not work similarly toward members of the same group or toward

out-group individuals. Rossi (2013) underlines that these dynamics have several relevant implications on important human aspects, for instance, they are deeply involved in the foundation of morality.

The overall view comes to light that emotional systems define behaviors in terms of cooperative/conflictual attitudes. This dialectics is generally considered to be an inherent aspect for social cognition. For instance, in his relational model, de Waal (1996b) has stressed the need for integrating the competitive nature of primates with the constraints on it in order to solve conflicts and protect cohesion. In this view, the role of aggression is not denied, but rather the sake of cooperation has favored the evolution of relationship-repair behaviors.

Leaning on the evolutionary arguments offered by Edward Wilson (2012), in this last section, we will make reference to a possible explanation of the complex dynamics underlying the dialectics between conflicts and integration involved in sociality. According to his model, complex social intelligence is the product of a multilevel selection which consists of traditional individual selection and group selection (Richerson and Boyd 2008). While the former shapes selfish instincts and therefore tends to produce competitiveness, the latter shapes instincts that tend to make organisms altruistic toward each other, that is, to adopt unselfish behaviors that favor group cohesion. The paradox inherent in the relation between these two forces intrinsic to multilevel selection produces the in-group/out-group attitude giving rise to the ambivalent human mind. Although the plausibility of group selection theory is rather controversial (e.g., Price 2012; for a review see Okasha 2006), in our opinion, the role of a mechanism working in shaping behaviors that benefit the group is likely. In this perspective, the selective forces are also implicated at the level of competing groups of individuals and not just at the level of competing individuals. That is, groups become units of selection at the level of the species' genome; in this way, a conflict between individual selection and group selection does not occur.

Following this argument, it is plausible to claim that emotions—with their double nature: individualistic and interpersonal—underlie the forces of multilevel selection that get selfish or cooperative pressures to prevail; in other words, emotions govern the balance between hostility and cooperation characterizing the human nature in a complex way in which conflict and integration are two intertwined aspects of sociality. Pinker's following quote (2002; pp. 237–238) really well catches the tangle of human nature underlined by emotions:

The other-condemning emotions - contempt, anger, and disgust - prompt one to punish cheaters. The other-praising emotions - gratitude and an emotion that may be called elevation, moral awe, or being moved - prompt one to reward altruists. The other-suffering emotions - sympathy, compassion, and empathy - prompt one to help a needy beneficiary. And the self-conscious emotions - guilt, shame, and embarrassment - prompt one to avoid cheating or to repair its effects.

The frame emerging from our argument has important implications on open topics such as morality. In this perspective, emotions have a main function in defining the roots of what is right and what is wrong; for instance, Singer (1982) highlights their role in expanding and reducing the circle of moral concern. On the same line, Tomasello (2008) hypothesizes that the sharing of emotions and attitudes

is a function of group identity that has been critical for the creation of social norms. The need to keep a good reputation and the mutual expectations about helpfulness in cooperative groups fostered the pressure for social conformity—that is a first step in the path toward the creation of norms. In the context of multilevel selection, the conformity with the group identity evolved in order to optimize intragroup bonds and intergroup differentiation has been achieved in an extraordinary way by humans.

To this extent, the ambivalent forces that characterize the animal and human mind appear to be at the base of the complex phenomena implicated in social cognition and are worthy of further consideration in the future research.

References

- Alcorta C, Sosis R (2005) Ritual, emotion, and sacred symbols: the evolution of religion as an adaptive complex. *Hum Nat* 16:323–359
- Anastasi MW, Newberg AB (2008) A preliminary study of the acute effects of religious ritual on anxiety. *J Altern Complement Med* 14(2):163–165
- Aureli F, de Waal FBM (2000) *Natural conflict resolution*. University of California Press, Berkeley, CA
- Aureli F, Preston SD, de Waal FBM (1999) Heart rate responses to social interactions in free-moving rhesus macaques (*Macaca mulatta*): a pilot study. *J Comp Psychology* 113:59–65
- Bering JM (2006) The folk psychology of souls. *Behav Brain Sci* 29(05):453–462
- Castles DL, Whiten A (1998) Post-conflict behaviour of wild olive baboons. I. Reconciliation, redirection and consolation. *Ethology* 104:126–147
- Cavalli-Sforza L, Padoan D (2013) *Razzismo e noismo. Le declinazioni del Noi e l'esclusione dell'altro*, Passaggi Einaudi
- Cheney DL, Seyfarth RM, Silk JB (1995) The role of grunts in reconciling opponents and facilitating interactions among adult female baboons. *Anim Behav* 50:249–257
- Clay Z, de Waal FBM (2013) Bonobos respond to distress in others: consolation across the age spectrum. *PLoS ONE* 8(1)
- Cole J (1998) *About face*. MIT Press, Cambridge, MA
- Cords M (1992) Post-conflict reunions and reconciliation in long-tailed macaques. *Anim Behav* 44:57–61
- Cords M (1997) Friendships, alliances, reciprocity and repair. In: Whiten A, Byrne RW (eds) *Machiavellian intelligence II: extensions and evaluations*. Cambridge University Press, Cambridge, MA
- D'Aquili E, Newberg A (1999) *The mystical mind: probing the biology of religious experience*. Fortress Press, Minneapolis, MN
- Darwin CR (1872) *The expression of the emotions in man and animals*. John Murray, London
- Davis MH (1994) *Empathy: a social psychological approach*. Brown and Benchmark, Madison WI
- De Waal FBM (1996a) *Good natured: the origins of right and wrong in humans and other animals*. Harvard University Press, Cambridge, MA
- De Waal FBM (1996b) Conflict as negotiation. In: McGrew WC, Marchant LF, Nishida T (eds) *Great ape societies*. Cambridge University Press, Cambridge, MA, pp 159–172
- De Waal FBM, Aureli F (1996) Consolation, reconciliation, and a possible cognitive difference between macaques and chimpanzees. In: Russon AE, Bard KA, Parker ST (eds) *Reaching into thought: the minds of the great apes*. Cambridge University Press, Cambridge, MA
- De Waal F, Lanting F (1997) *Bonobo: the forgotten ape*. University of California Press, Berkeley
- Dunbar RIM (1991) Functional significance of social grooming in primates. *Folia Primatol* 57:121–131

- Eibl-Eibesfeldt I (1970) *Ethology. The biology of behavior*. Holt, Rinehart and Winston, Inc., New York, NY
- Ekman P, Levenson RW, Friesen WV (1983) Autonomic nervous system activity distinguishes between emotions. *Science* 221:1208–1210
- Feh C, de Mazières J (1993) Grooming at a preferred site reduces heart rate in horses. *Anim Behav* 46(6):1191–1194
- Ferretti F, Adornetti I (2014) Biology, culture, and coevolution: religion and language as case studies. *J Cogn Culture* 14(3–4):305–330
- Gallese V (2009) Mirror neurons, embodied simulation, and the neural basis of social identification. *Psychoanal Dialogues* 19(5):519–536
- Ginges J, Atran S, Medin D, Shikaki K (2007) Sacred bounds on rational resolution of violent political conflict. *Proc Natl Acad Sci* 104:7357–7360
- González MAE, J. A RL, Piqueras Rodríguez J, Oblitas Guadalupe LA (2009) Emoción y Cognición: Implicaciones para el tratamiento. *Terapia Psicológica* 27(2):227–237
- Gratch J, Mao W, Marsella S (2006) Modeling social emotions and social attributions. In: Sun R (ed) *Cognition and multi-agent interaction: extending cognitive modeling to social simulation*. Cambridge University Press, Cambridge, MA
- Haidt J (2003) The moral emotions. In: Davidson RJ, Scherer KR, Goldsmith H (eds) *Handbook of affective sciences*. Oxford University Press, New York, NY, pp 852–870
- Harris LT, Fiske ST (2006) Dehumanizing the lowest of the low neuroimaging responses to extreme out-groups. *Psychol Sci* 17(10):847–853
- Hummer RA, Rogers RG, Nam CB, Ellison CG (1999) Religious involvement and US adult mortality. *Demography* 36(2):273–285
- Iacoboni M (2008) *Mirroring people: the new science of how we connect with others*. Farrar, Straus, Giroux, New York, NY
- Lee BY, Newberg AB (2005) Religion and health: a review and critical analysis. *Zygon* 40(2): 443–468
- Levenson RW (1994) Human emotion: a functional view. In: Ekman P, Davidson RJ (eds) *The nature of emotion: fundamental questions*. Oxford University Press, New York, NY
- Levenson RW (2003a) Autonomic specificity and emotion. In: Davidson RJ, Scherer KR, Goldsmith HH (eds) *Handbook of affective sciences*. Oxford University Press, New York, NY
- Levenson RW (2003b) Blood, sweat and fears: the autonomic architecture of emotion. In: Ekman P, Campos JJ, Davidson RJ, DeWaal FBM (eds) *Emotions inside out*, vol 1000, *Annals of the New York Academy of Sciences*. New York Academy of Sciences, New York, NY, pp 348–366
- Levin JS, Schiller PL (1987) Is there a religious factor in health? *J Relig Health* 26:9–36
- Lienard P, Boyer P (2006) Whence collective rituals? A cultural selection model of ritualized behavior. *Am Anthropol* 108(4):814–827
- Nagel T (1986) *The view from nowhere*. Oxford University Press, New York, NY
- Newberg A, Alavi A, Baime M, Pourdehnad M, Santanna J, d'Aquili E (2001) The measurement of regional cerebral blood flow during the complex cognitive task of meditation: a preliminary SPECT study. *Psychiatry Res Neuroimaging* 106(2):113–122
- Newberg A, Pourdehnad M, Alavi A, d'Aquili EG (2003) Cerebral blood flow during meditative prayer: preliminary findings and methodological issues. *Percept Mot Skills* 97(2):625–630
- Okasha S (2006) Evolution and the levels of selection, vol 16. Clarendon Press, Oxford
- Pinker S (2002) *The blank slate*. Viking, New York, NY
- Price ME (2012) Group selection theories are now more Sophisticated, but are they more predictive? *Evol Psychol* 10(1):45–49
- Richerson PJ, Boyd R (2008) *Not by genes alone: how culture transformed human evolution*. University of Chicago Press, Chicago, IL
- Rizzolatti G, Gallese V, Keysers C (2004) A unifying view of the basis of social cognition. *Trends Cogn Sci* 8(9):396–403
- Rossi MG (2013) *Il giudizio del sentimento. Emozioni, giudizi morali, natura umana*. Editori Riuniti University Press, Roma

- Rowe C (1999) Receiver psychology and the evolution of multi-component signals. *Anim Behav* 58:921–931
- Rozin P, Lowery L, Imada S, Haidt J (1999) The CAD triad hypothesis: a mapping between three moral emotions (contempt, anger, disgust) and three moral codes (community, autonomy, divinity). *J Pers Soc Psychol* 76(4):574
- Sapolsky R (1999) Hormonal correlates of personality and social contexts: from non-human to human primates. In: Panter-Brick C, Worthman C (eds) *Hormones, health and behavior*. Cambridge University Press, Cambridge, MA, pp 18–46
- Seehausen M, Kazzer P, Bajbouj M, Heekeren HR, Jacobs AM, Klann-Delius G, Menninghaus W, Prehn K (2014) Talking about social conflict in the MRI scanner: neural correlates of being empathized with. *Neuroimage* 84:951–61
- Shamay-Tsoory SG, Tomer R, Yaniv S, Aharon-Peretz J (2002) Empathy deficits in Asperger syndrome: a cognitive profile. *Neurocase* 8:245–252
- Sheikh H, Ginges J, Coman A, Atran S (2012) Religion, group threat and sacred values. *Judgm Decis Mak* 7:2
- Singer P (1982) *The expanding circle: ethics and sociobiology*. New American Library, New York, NY
- Sosis R, Bressler ER (2003) Cooperation and commune longevity: a test of the costly signaling theory of religion. *Cross-Cult Res* 37:211–239
- Tiedens LZ, Linton S (2001) Judgment under emotional certainty and uncertainty: the effects of specific emotions on information processing. *J Pers Soc Psychol* 81(6):973
- Tomasello M (1999) *The cultural origins of human cognition*. Harvard University Press, Cambridge, MA
- Tomasello M (2008) *Origins of human communication*. MIT Press, Cambridge, MA
- Tomasello M (2009) *Why we cooperate*. MIT Press, Cambridge, MA
- Turner RP, Lukoff D, Barnhouse RT, Lu FG (1995) Religious or spiritual problem. A culturally sensitive diagnostic category in the DSM-IV. *J Nervous Mental Dis* 183:435–44
- Whitehouse H (2012) Ritual, cognition, and evolution. In: Sun R (ed) *Grounding the social sciences in the cognitive sciences*. MIT Press, Cambridge, MA
- Wilson EO (2012) *The social conquest of earth*. Liveright Publishing Corporation, New York, NY

Part II
Argumentation and Conflict

Chapter 7

Arguments, Conflicts, and Decisions

Fabio Paglieri

7.1 Introduction

In English, “to argue” has two very different meanings: on the one hand, it refers to the practice of giving reasons for or against a certain position, thus submitting them to intersubjective scrutiny; on the other hand, it indicates the act of verbally fighting against an opponent, often viciously and with little or no exchange of reasons. Let us label these two senses as, respectively, the rational and the polemical view of argument. While different languages do not exactly share the same ambiguity, other forms of polysemy are found in relation to arguments: for instance, the Italian noun *argomento* either means “argument” in its rational sense (a concatenation of reasons in favor of a conclusion), or it indicates instead a topic of discussion; in the latter sense, it is often used to label such topic as inappropriate or dangerous, e.g., *Non è un bell’argomento* (It is not a nice thing to discuss), *Non tocchiamo l’argomento, per favore!* (Let us skip the issue, please!), *È un argomento delicato* (It is a sensitive matter).

This tension between a rational and a polemical view of arguments is not just a quirk of natural languages. On the contrary, we find it mirrored in different theoretical perspectives on argumentation. In pragma-dialectics (van Eemeren and Grootendorst 2004), a critical discussion is defined as an ideal model of argumentative discourse, in which argumentation is directed at resolving a difference of opinion through reasonable means: thus arguing is seen as a rational exercise to remedy an intersubjective conflict. In sharp contrast, other argumentation scholars (e.g., Goodwin 2007) emphasize that promoting conflict is the main functional effect of arguing, whether or not the arguers intend it, and that sometimes conflict escalates

F. Paglieri (✉)

Goal-Oriented Agents Lab (GOAL), Istituto di Scienze e Tecnologie della Cognizione,
Consiglio Nazionale delle Ricerche, via san Martino della Battaglia 44, 00185 Rome, Italy
e-mail: fabio.paglieri@istc.cnr.it

as a consequence of argumentative engagement, rather than in spite of it (Cohen 2005; Paglieri 2009). This more conflictual view of argumentation is supported by some empirical findings on how arguers perceive their dialogical interaction: Hample and Benoit (1999) found that people do not consider a dialogue as being argumentative in nature, unless it is adversarial and somehow unresolved (see also Hample et al. 1999).

This leaves us with a puzzle on the relationship between argument and conflict. Does argumentation promote conflict resolution, or does it tend to exacerbate conflicts? Crucially, this is not just a matter of theoretical curiosity, but it has also important practical implications. Consider, for instance, its ramifications in critical thinking education: if we see argumentation as a rational salve to heal social conflicts, then of course its fostering should be regarded as a key priority for any educational system worth its salt; but if, on the contrary, training people to argue risks turning them into bickering, obdurate, insufferable know-it-alls, then great caution should be exerted.

In light of such potential repercussions, it is important to establish whether the dichotomy between a rational and a polemical view of argumentation is real or apparent, before we even begin considering how to solve it. Section 7.2 of this paper is devoted to articulate a decision-theoretic approach to argumentation, which is suggested to put this tension in better focus and allows to make sense of it. Section 7.3 reviews extant evidence in favor of that approach, highlighting several roles that conflicts play in our argumentative decisions. Finally, Sect. 7.4 outlines future developments and open issues in this line of research.

7.2 The Role of Conflict in Argumentative Decisions

When we argue, we make several decisions, sometimes without even realizing it: we choose whether to enter the argument or not, what arguments to use and how to present them, how to respond to the arguments of the counterpart, how to address challenges and objections, how to solve potential ambiguities, when and how to end the argument, and so forth. In fact, argumentation can be seen as the result of a complicated decision-making process or, more exactly, as the interaction of multiple decision-making processes performed by autonomous agents.

In spite of its obvious relevance in everyday argumentation, decision making has been taken for granted rather than explored in argumentation theories, with few exceptions (see Hample 2005 for a review of some of them, as well as my own efforts in the same direction, summarized in Paglieri 2013a). This partial neglect originates from an insistence on what is the right move in a given argumentative situation and not on how the subject may decide to opt (or not) for that move. It is not that argumentation theorists are unaware of argumentative decisions, of course—as arguers, if nothing else, they are bound to be familiar with those. They just do not see it as their business to produce a theory of such decisions (for a diagnosis of why this is the case, see Paglieri 2013a).

In contrast with this traditional lack of scholarly interest for argumentative decisions, recent studies have proposed to analyze argumentation as a decision-making process, both theoretically (Paglieri 2009, 2013b; Paglieri and Castelfranchi 2010) and experimentally (Cionea et al. 2011; Hample et al. 2011). In what follows, I will summarize the main results of this line of research, but first let us clarify its relevance to the problem at hand—to wit, developing a balanced view of the relationship between arguments and conflicts. In a nutshell, the guiding idea is that our expectations on how arguing will impact on conflict (of beliefs, goals, values, etc.) are crucial in determining *whether* to engage in argumentation at all and in deciding *how* to do so. Thus, while most of the times¹ we indeed argue with the aim of solving or defusing a conflict, our awareness of the inherent risks of conflict escalation may justify avoiding or fleeing an argument, for fear of making things worse. Seen in a decision-theoretic light, there is no longer a puzzle here: it is precisely because we want our arguments to defuse conflicts that we factor the dangers of polemical escalation in our argumentative decisions. If we want our dialogical engagements to be rational and productive, we must take great care in avoiding those moves that would make them violent and destructive. Thus conflict looms large in argumentative decisions, both as a problem to solve and as a pitfall to avoid.

Let us now look in greater detail to what a theory of argumentative decisions looks like. In presenting a typology of the choices we make while arguing, it is useful to follow the typical chronological order in which the arguer has to face them. This criterion is by no means the only possible one, and the resulting taxonomy is not necessarily intended to be exhaustive. But it does provide a convenient starting point, both by giving some order to the discussion and by helping to better illustrate argumentative decisions “in real life.” In light of these considerations, what follows is meant as a first tentative process-based taxonomy of argumentative decisions (discussed in greater detail elsewhere; see Paglieri 2009, 2013a, b):

- (a) *Argument engagement*: the decision to enter an argument or not, either by proposing one or by accepting to be drawn into one by the counterpart. Considering engagement as a decision implies acknowledging that arguing is not always the best option, and sometimes it is actually the worst (Martin and Scheerhorn 1985; Hample and Benoit 1999; Cohen 2005; Goodwin 2005, 2007; Paglieri 2009). More generally, the strategic considerations that are relevant in choosing whether to argue or not are best understood in terms of costs and benefits, as exemplified by various contributions in Artificial Intelligence

¹But not always, e.g., in instances of what Walton (1998) labels eristic confrontation, where the dialogical goal is to vent one’s feelings at the opponent, not to solve any underlying conflict of opinion—on the contrary, often with the reasonable expectation of exacerbating it.

- (Amgoud and Maudet 2002; Karunatilake and Jennings 2005; Riveret et al. 2008; Rahwan and Larson 2009; Paglieri and Castelfranchi 2010).²
- (b) *Argument editing*: all decisions concerning what arguments to use (*selection*) and how to present them to the audience (*presentation*), in order to maximize their intended effects (Hample and Dallinger 1990, 1992; Hample 2005; Hample et al. 2009). These argumentative decisions also have obvious relevance in rhetoric, and in fact, most of the five canons of Western classical rhetoric (*inventio*, *dispositio*, *elocutio*, *memoria*, *pronuntiatio*) can be seen as different sub-decisions concerning effective argument presentation.
- (c) *Argument timing*: the decision on when it is time to speak and when it is time to listen. An appropriate timing of one's argumentative contribution and an awareness about the optimal length of one's speech are essential elements for making an effective argument in almost every dialogical context, including one-way presentations in front of an audience. In contrast, most argumentation theories lack a systematic study of this aspect, and the only form of timing which is considered consists in the kind of highly stylized, improbably well-mannered turn-taking assumed by some dialogical approach—for an illustration, look at the otherwise informative example of argumentative dialogue provided by Walton and Krabbe (1995, pp. 86–91).
- (d) *Argument interpretation*: if there are ambiguities in what the counterpart is saying, the decision on whether to criticize them, ask for more clarity or additional information, or solve them autonomously—and if so, favoring what interpretation, on what grounds, and to what ends. A well-studied case of argument interpretation concerns enthymemes, but most theories of enthymemes focus on what is the normatively correct/legitimate reconstruction of the argument. To highlight the arguer's underlying decisions, pragmatic approaches such as relevance theory (Sperber and Wilson 1986; Wilson and Sperber 2002) are more useful (for additional details on this point, see Paglieri 2007; Paglieri and Woods 2011a, b).
- (e) *Argument reaction*: decisions concerning whether to accept or challenge an argument, an objection, or a counterargument raised by the counterpart. This area has been so far dominated by normative concerns: the widely received wisdom in argumentation theories is that arguments should be challenged and critical questions should be asked, whenever appropriate, but there has been little consideration on what reasons (other than being right) might guide this choice (see Gilbert 1997, for some in-depth discussion of these issues, as well as a critique of the enduring lack of attention they suffered in argumentation theories).
- (f) *Argument termination*: the decision on when and how to end an ongoing argument. Clearly, the arguer cannot unilaterally “decide” to win the argument

²Costs and benefits are crucial also in relevance theory (Sperber and Wilson 1986; Wilson and Sperber 2002), but with several important differences with respect to the decision-theoretic approach outlined here (for discussion of this point, see Paglieri 2013b).

(or to reach whatever goal she/he pursued by arguing), since, for this to happen, the agreement of the counterpart and/or the satisfaction of some objective criteria is required. But each arguer can and does decide whether to let the other win by conceding the point, or shelve/postpone the argument, or move on to other matters, or some other way of terminating the argumentative exchange for the time being (for some preliminary findings on this point, see Benoit and Benoit 1987, 1990; Vuchinich 1990; Hicks 1991; Hample et al. 1999).

Even though this taxonomy of argumentative decisions is based on a procedural view of argumentation, starting from its inception (engagement) and ending with its conclusion (termination), the temporal succession of many decisions is not fixed, and some decisions are bound to co-occur: for instance, it is typically the case that arguers have to decide simultaneously on when to speak and what to say (timing and editing), as well as on how to interpret the counterpart's utterance and how to respond to it (interpretation and reaction). Moreover, many of these categories refer to families of decisions, rather than to a single act of choice.

In spite of these limitations, this preliminary taxonomy suffices to convey the sheer ubiquity of decision making in argumentation, and it provides a principled starting point to look at arguments in this relatively new light. In the next section, I will review some empirical evidence pertaining to three of the categories mentioned above (engagement, editing, termination), to highlight the key role conflict plays in all these decisions.

7.3 Conflict Matters: Empirical Evidence on Argumentative Decisions

So far, the relevance of conflict considerations in orienting our argumentative decisions has been defended mostly on intuitive grounds. Based on our everyday experience, there is *prima facie* evidence that we do take into account how arguing (or not) will affect intersubjective conflicts, either defusing or escalating them, and then use these expectations in deciding how to proceed. However, is it possible to muster more reliable evidence on the role of conflict in argumentative decisions? Is it truly so pivotal as I have been arguing, or is it often trumped by other more important considerations?

In this section, I summarize empirical evidence (Sects. 7.3.1 and 7.3.2) and testable predictions (Sect. 7.3.3) that speak in favor of considering conflict as a cornerstone of argumentative decision making. My overall assessment is that while the evidence is not yet conclusive, it is more than sufficient to motivate further research along these promising lines.

7.3.1 *Argument Engagement: Likelihood of Success and Contextual Appropriateness*

Previous work (Paglieri 2009) has suggested that arguers do not carelessly waltz into any potential dialectical confrontation, but on the contrary are very careful in “picking their fights,” more often than not declining the opportunity to engage in argument, and with good reasons. More generally, this emphasizes that arguing is not always the best option, not only because arguments may occasionally backfire and lower the credal status of their conclusion (Cohen 2005) but also because the original disagreement (when arguments originate from a divergence of opinions, as they often do) may in fact escalate due to argumentative intercourse, rather than resolving itself (a point also discussed in Martin and Scheerhorn 1985; Hample and Benoit 1999; Goodwin 2005, 2007). Crucially, arguers are fully aware of these dangers and factor them in deciding whether to engage in argument or not.³

Hample (2009) noticed that this view, originally proposed as a theoretical speculation based on commonsense observations, could be easily operationalized in terms of costs and benefits (see also Paglieri and Castelfranchi 2010), thus producing some testable predictions on what conditions would make an arguer more or less willing to engage in argument. The following is the list of predictive factors proposed by Hample and colleagues (2011), to empirically test the hypothesis that argument engagement is affected by such considerations:

- (a) Costs, including effort, emotional exposure, and unwelcome consequences
- (b) Benefits, that is, whatever an arguer gets out of the interaction if it goes well
- (c) Likelihood of winning/success
- (d) Reasonableness of the counterpart/audience
- (e) Civility, that is, how pleasant and productive the interaction is expected to be, as opposed to angry and destructive
- (f) Resolvability of the argument
- (g) Contextual appropriateness

This list was meant as a simplified version of a richer but less precise model (Paglieri 2009) and intended for a first exploratory attempt to individuate factors affecting argument engagement. As such, it leaves much to be desired, both in terms of completeness (are there other relevant factors not listed here?) and correctness (are these factors all necessary or are some of them irrelevant?), as well as ignoring some potential interdependence between different factors (for

³The studies presented in this section refer to persuasion dialogues (Walton 1998), in which each party tries to prove a point that the other party is resistant to accept. Such focus on persuasion will result in frequent reference to the notion of “winning the argument” as a criterion for success. This does *not* imply that winning is invariably the purpose of arguing in general: as it is often emphasized in the literature (Walton and Krabbe 1995; Walton 1998; Paglieri and Castelfranchi 2010), arguing can be motivated also by other goals, personal or dialogical. Nonetheless, arguers are still concerned with *success*, defined in terms of whatever goal they are striving to achieve.

instance, it seems likely that items from D to G determine C and thus impact on A and B). Nevertheless, this tentative list of criteria provided useful guidance in the first two studies devoted to investigate experimentally how arguers make the decision to engage in argument, as outlined in what follows.⁴

The first experimental test of the model was conducted on a sample of 509 undergraduates (of which 473 completed the survey) at a large public Mid-Atlantic university in the United States. The seven factors listed above were studied in interaction with two personal variables, verbal aggressiveness (Infante and Wigley 1986) and argumentativeness (Infante and Rancer 1982), and a situational variable, consisting in three different topics of discussion: public, private, and workplace topic (Johnson 2000). After compiling the personality scales and providing basic personal information, participants were presented with a short dialogical scenario, regarding one of the three topics under study, in which a disagreement with a close friend was described, suggesting (but not necessarily forcing to accept) the possibility of engaging in argumentation to resolve it. After reading the scenario, respondents compiled several multi-item scales, to probe their assessment of the factors listed above with respect to that situation, as well as their willingness to engage in argumentation under such circumstances (the dependent variable). Data analysis was performed using a structural equation model (SEM), connecting personal, situational, and argumentative variables to intention to engage, to verify the relative weight and role of each factor in determining the willingness to engage in argument, and to explore the impact of different situational contexts on the decision to engage (i.e., what factors are more relevant in different situations?).

Overall, this first study provided partial support to the model and in particular revealed that:

- Personal variables do *not* affect consideration of argumentative factors: in assessing costs and benefits of arguing, people seem to be fairly objective—at least as far as their degree of argumentativeness and verbal aggressiveness is concerned.
- The intention to engage in argument is *strongly predictable* (the final equations explained most of the variance in the data), but the relevant factors *differ* across situations.
- *Likelihood of success* and *contextual appropriateness* are always present as relevant predictors of the decision to engage, and success is by far the most important consideration.
- More generally, manipulating situational variables proved to be instructive, not only to highlight different minor predictors but also to make some unexpected and/or counterintuitive findings emerge, potentially spurring more focused research (for additional details on this point, see Hample et al. 2011).

⁴Both studies are published (Study 1: Hample et al. 2011; Study 2: Cionea et al. 2011), and readers are referred to those articles for further methodological elucidation.

The study was subsequently replicated with a population of 201 Romanian respondents, using the same multi-item scales but employing a different manipulation of the situational variable. In this new study (Cionea et al. 2011), both the topic (private vs. public) and the relationship with the counterpart (friend vs. romantic partner) were manipulated. The basic findings were largely consistent with those reported by Hample and colleagues: perceived likelihood of success and contextual appropriateness were still the main predictors of the intention to argue, and again personal variables (i.e., argumentativeness and verbal aggressiveness) did not affect either the decision to argue or the assessment of costs and benefits relative to that decision.

Replicating this study with Romanian respondents also aimed to address the role of *cultural differences* in argument engagement. In this respect, it was more interesting to look at significant discrepancies, rather than similarities, with the data on US respondents collected by Hample and colleagues. Two interesting results emerged: first, the perceived reasonability of the counterpart is important for Romanians in deciding whether to enter the argument or not, whereas US Americans do not seem to care at all for that factor; second, Romanians in general appear to be more argumentative than US Americans, even if both populations have similar scores in verbal aggressiveness. The latter finding is especially intriguing, since it contradicts the hypothesis that people from collectivistic cultures (such as Romania is usually considered an instance of) are less argumentative than people from individualistic cultures (of which the United States is the often cited stereotype; see Prunty et al. 1991 for an example of this type of interpretation).

Overall, these preliminary experimental investigations of argument engagement provide partial support to the theoretical model outlined by Paglieri (2009) and demonstrate that a decision-making approach to argumentation is fruitful. The very fact that the decision to engage in argument is better predicted by strategic considerations, such as likelihood of winning and contextual appropriateness, rather than personality traits, such as argumentativeness and verbal aggressiveness, is an important finding. That this remains true even across cultures, and even if these cultures sharply differ in terms of argumentative habits, speaks for the potential generality of these results and invites further attempts to probe engagement decisions in argumentation, both theoretically and experimentally.

Of particular significance for the aims of this paper is the key role played by expectations of success and perceived contextual appropriateness: arguing is considered worth pursuing only if (1) the subject sees it as conducive of a solution to the original disagreement, (2) in favor of the subject's own position, and only if (3) the social setting is considered suitable for conducting an argumentative exchange. All these factors highlight how conflict considerations shape our argumentative practices: (1) and (2) demonstrate that people tend to argue only if they think this will reduce, rather than escalate, their differences of opinions, possibly to one's own advantage; on the other hand, (3) indicates a sensitivity to context which is also conflict related, since certain situations may be less conducive of agreeable solutions than others (e.g., public arguments may lead to entrenched positions due to fear of reputation losses) and also increase the likelihood of generating additional conflicts

as a side effect (e.g., arguing at a social gathering with one's in-laws may be seen as a deliberate insult, thus spurring a quarrel unrelated to the original disagreement). Thus this data confirms that arguers pick their fights carefully—that is, they enter an argument based on some understanding of how this will affect the range of intersubjective conflicts faced by both parties.

7.3.2 *Argument Editing: Editorial Standards and Styles*

Dale Hample and colleagues conceptualize argument editing as a process that occurs between an initial, private conception of a potential message or move in an argument and its final public production (or lack thereof). The basic insight is that if, upon reflection, the initial idea for an argumentative contribution comes to be regarded as inappropriate or unwise, it may be edited (i.e., properly modified or utterly censored) to better suit the arguer's goals (for in-depth discussion of this view of editing, see Hample 2005). In a series of empirical studies (e.g., Hample and Dallinger 1990, 1992; Hample 2006; Hample et al. 2009), it has been shown that people vary in the degree to which they bother to edit at all (some arguers are just “blurters,” uttering whatever comes to mind with no filter), as well as in what standards they use to reshape their utterances.

Importantly, this individual variation is not random. Hample and collaborators have identified three basic classes of *editorial standards*, roughly as follows:

- *Effectiveness*: the original contribution is edited because it is either expected to fail or it is considered to be too negative and thus likely to hinder the interaction.
- *Person-centered considerations*: the contribution is edited because its original form is considered at risk of being too harmful to either oneself, the other, or the relationship.
- *Discourse competence*: the contribution is edited because its first formulation is considered either false or irrelevant to the matter of discussion.

In turn, whether a person favors one class of standards or another depends on his/her goals. In empirical studies to date, two main *editorial styles* have emerged as dominant among arguers:

- *Effectiveness editors*: these people represent themselves as willing to say nearly anything that will work argumentatively, no matter the consequences.
- *Person-centered editors*: these arguers will not say things that have negative identity or relational repercussions, even when this leads to deliberately swallowing potentially effective messages.

Editorial preferences have also been studied in connection with various personality measures (see Hample 2005 for a summary), such as argumentativeness (Infante and Rancer 1982), verbal aggressiveness (Infante and Wigley 1986), and psychological gender (i.e., masculinity vs. femininity, Bem 1974). While none of these factors is the sole determinant of editorial style, they do have an influence.

With respect to argumentativeness, the interaction emerges at the level of its two subscales, argument approach and argument avoidance. People with high approach scores tend to bother less with editing overall, while avoidant arguers show strong concern for not harming others when editing. Not surprisingly, verbal aggressiveness reduces the amount of editing and also correlates with lack of interest for harming others. Finally, masculinity associates with a general disregard for harm to self, other, or relationship and a predominance of effectiveness considerations, whereas femininity scores do not predict editorial choices at all.

Taken together, these results offer a view of the role of conflict in argument editing that is remarkably coherent with its parallel role in argument engagement. Indeed, it is very significant that the predominant editorial styles are focused either on the potential for success in conflict resolution (effectiveness editors) or on minimizing the dangers of generating additional conflicts and/or escalating the original disagreement (person-centered editors). In contrast, the only editorial standard not directly focused on conflict (discourse competence) is typically delegated to an ancillary role, as something to be considered only after other conflict-related concerns have been satisfied. Sometimes, we do edit our arguments to avoid falsity or irrelevance, but this is not our dominant inclination: making sure that every argument counts toward conflict resolution and/or avoidance is considered to be by far more urgent and important. Once again, we see here how argumentative decisions strive to achieve a balance between two opposing forces in conflict management: resolving the original disagreement, without flaring any other unwanted controversy along the way.

7.3.3 Argument Termination: Duration Undermines Positive Resolution and Facilitates Conflict Escalation

So far, we have reviewed empirical evidence on how conflict shapes argumentative decisions, in particular regarding argument engagement and argument editing. In this section, the focus shifts on argument termination: while there are not yet empirical findings to be considered on this topic, it is possible to formulate some predictions on what factors will influence the decision to terminate an argument and what role conflict may play in that choice. This, in turn, will help stimulating empirical verification of these hypotheses.

Once an argument is joined, each participant typically remains free to withdraw from it at any time: special circumstances might limit this freedom, e.g., in legal trials, but this is rarely the case in everyday arguments. What an arguer typically cannot unilaterally decide is to bring the argument to a conclusion that satisfies his/her goals, since this often requires the agreement of the counterpart (in persuasion or negotiation) or the achievement of some objective criteria (greater understanding in inquiry, satisfactory self-expression in eristic confrontation, etc.). But it is well within their power to concede the point to the counterpart, to cut short

the discussion, to shelve an issue or postpone debating it, and more generally to find suitable ways of concluding the argumentative interaction before the matter under discussion has been settled. The factors affecting such decision are, by and large, the same that are responsible for entering the argument in the first place: whatever reason makes it worthwhile to argue with someone, it is typically also the reason that justifies prolonged discussion of that matter with the same interlocutor. However, the history of the ongoing argument also has a rational influence on the arguers' decision to either continue or interrupt it.

Two main aspects are relevant here: how the argument progressed so far (*quality*) and how long it has been going on (*duration*). Obviously, the quality of the argument is likely to predict the arguer's intention to continue it: if an argumentative exchange is perceived as productive, interesting, engaging, informative, etc., it stands to reason that both parties will be willing to keep at it; if, on the contrary, the argument is experienced as pointless, frustrating, circular, depressing, etc., the arguers will be ready to abandon it at the first chance. Once again, here we see the double-edged role played by conflict considerations in argumentative decisions. However, the quality of an argumentative exchange needs to be assessed on a case-by-case basis, since it depends on the specifics of the interaction, rather than on abstract criteria. As such, considerations of quality are unlikely to yield general predictions on argument termination, aside from the obvious—good quality discourages argument termination; bad quality encourages it.

Argument duration is a more objective metric, one that has general and verifiable effects on argument termination.⁵ To see how, let us consider the arguer as an expected utility maximizer (or any other type of rational decision maker, as discussed below): whenever continuing to argue yields an expected utility higher than terminating the argument, this is what the agent will (and should) do. Thinking in terms of expected utility implies focusing on two aspects of the decision problem: the payoffs and the perceived likelihood of securing such payoffs. In the case of argumentation, the payoffs can be roughly characterized as the benefits of arguing (whatever positive goal the arguer expects to obtain) minus the costs (whatever resources are needed to carry out the argumentative activity) minus the dangers (whatever undesired side effect is likely to occur, due to argumentation). As for the perceived likelihood of all these outcomes, benefits and dangers of arguing are typically uncertain (arguers can be more or less sure of reaching a satisfactory conclusion and more or less afraid of incurring harmful side effects), whereas the costs of argumentation tend to be fairly certain—all other things being equal, a long

⁵The reason why the argument happens to be prolonged will also affect the arguer's choice: that is, argument quality and argument duration are not entirely independent from each other. Consider for instance *counter argumentation*, that is, arguments from the other party aimed at undermining or undercutting one's own arguments (see Pollock 1992; Prakken 2000; Besnard and Hunter 2001). Given extant evidence on how preexposure to counterarguments affects resistance to persuasion (e.g., Petty and Cacioppo 1977), the amount of counter argumentation experienced so far is likely to matter for deciding whether and how to terminate the exchange.

argument requires more resources than a short one, regardless of how it is concluded or what other effects it produces.

Based on these commonsense considerations, some hypotheses can be formulated on how argument duration might affect (negatively, for the most) the expected utility associated to the choice of prosecuting the argument, in three different ways:

- (a) *Duration increases the costs but not the benefits*: the costs of arguing inevitably increase with duration, in terms of time, breath, cognitive effort, social exposure, and lost opportunities—all the other things one could have done, instead of being stuck in the argument. In contrast, benefits are typically independent from the duration of the argument: for instance, in persuasion or in negotiation, what is gained by winning the argument or by reaching an agreement remains the same, whether it took one minute or one year to achieve that result—again, assuming no other factor is at play, such as a need for indirectness to avoid being perceived as impolite or aggressive. Even in the special circumstances where prolonged discussion may increase the benefits of arguing (e.g., in inquiry dialogues, according to the taxonomy in Walton 1998), benefits typically increase at a slower rate than costs, with very few exceptions—one such exception would be a very productive brainstorming session, in which all participants are getting out of the exchange more than they put into it. However, we are all too familiar with how depressing brainstorming sessions can be, precisely because we often perceive them as a huge waste of time for relatively small payoffs, and not the other way around. This, I argue, is the typical dynamics of costs and benefits in prolonged discussion, even in inquiry dialogues. Whenever costs increase and benefits remain stable, or costs increase faster than benefits, the cost/benefit balance of arguing deteriorates as a function of duration (the exact shape of each function is irrelevant here). This means that (1) the longer the argument, the less convenient it becomes, up to the point where (2) arguing becomes an absolute liability, since the arguer loses utility even if the argument is successfully concluded.
- (b) *Duration increases the likelihood of negative consequences (dangers)*: prolonged argumentation is likely to put a strain on the arguers' relationship, and arguers factor this risk in their decision to persevere in an argument, rather than dropping it (see also Gilbert 1997). Indeed, some empirical evidence suggests that arguers perceive full-blown arguments as inherently dangerous and potentially damaging, so much so that polite, well-mannered exchanges are considered as incomplete arguments that were cut off before they could harm the relationship (Hample et al. 1999). In general, argument duration can only heighten the perception of similar dangers, not only because people quickly tire of being opposed and become more and more resentful toward their opponent (cognitive fatigue and emotional strain) but also because exploring the initial disagreement without finding any solution changes the perception of the counterpart's standpoint. As discussed more thoroughly in previous work (Paglieri 2009), interpersonal disagreement is more easily condoned before discussing it, and the very fact of engaging in argumentation raises the stakes for

social interaction. Prior to argument, the parties were in a situation of respectful disagreement (“We have different views”), but once the issue has been debated in details without reaching any consensus, the arguers find themselves in a condition of stubborn disagreement (“You will not listen to reason!”), which directly threatens their relationship. The longer the argument, the higher the danger of facing such quandary.

- (c) *Duration reduces the likelihood of a satisfactory conclusion (benefits)*: prolonged argumentation can indicate either that the subject matter is especially complex or that the parties are failing to find any common ground or both. Whatever the reasons for prolonging the debate, it would thus seem that duration speaks against reasonable expectations for a positive conclusion. This negative correlation between argument duration and success expectations is understandable, in view of the following considerations: first, by having to repeatedly defend their respective positions against criticisms and objections, the arguers become more and more entrenched in their views, hence less willing to accept alternative standpoints or compromises (radicalization)⁶; second, exploring the matter in greater details is likely to uncover further issues of disagreement between the parties (e.g., we begin discussing why we like different political candidates and end up heatedly debating abortion rights and alternative energy sources), which in turn makes it harder to reach a satisfactory conclusion on the original topic (explosion); third, in terms of rhetoric value, long-winded arguments are more complex to follow and persuasively less effective, thus reducing the chances that the counterpart will be swayed by them (deterioration); fourth, the very fact that the arguers have been discussing for a long time over the same issue is taken as evidence that further debate is futile and no satisfactory conclusion will ever be achieved, and once this conviction is in place, it acts as a self-fulfilling prophecy (disillusion). Due to the combined effects of all these factors, argument duration typically reduces the arguer’s reasonable expectations of reaping the benefits of the discussion, whatever they might be.

The upshot of this analysis is that the expected utility of prolonging an argument should be expected to steadily decline as a function of its duration, since the cost/benefit balance deteriorates, various dangers increase, and likelihood of success decreases. This yields the straightforward prediction that the arguer’s willingness to disengage from the argument will grow as the discussion progresses. The presumed

⁶In group discussion, this interacts also with *polarization*, as an effect of being exposed to novel arguments (from other like-minded arguers) in favor of one’s previous position: as it has been repeatedly shown in social psychology (for a review and discussion, see Myers and Lamm 1976; Isenberg 1986; Sunstein 2002), this leads groups of like-minded people to develop positions far more extreme than those previously held by their individual members. The interplay of radicalization and polarization explains why multiparty argumentation, e.g., debates on social media, easily produces partisanship, that is, the predominance of relatively few views, each quite extreme and fanatically opposed to any dissenting voice.

relative impact of each factor is also linked to verifiable predictions: if the decision of truncating the argument is inspired by excessive costs (factor A above), one would expect that manipulating perceived costs would have an effect on such decision; if, on the contrary, the choice is due to increased dangers of negative side effects (factor B), reducing or enhancing the likelihood of such dangers should affect argument termination; finally, if what matters is duration-induced skepticism on the chances of reaching an agreement (factor C), the arguers should be sensitive to manipulations that reduce such skepticism. Moreover, different factors may also prompt different strategies on how to terminate the argument: for instance, if argument termination is motivated by a desire to avoid antagonizing the counterpart, generously conceding the point and negotiating an amicable truce are the most likely options; on the contrary, if the argument is ended out of dissatisfaction for one's chances of overcoming opposition, an abrupt interruption is more probable, with no concessions or niceties toward the counterpart.

Before turning to consider the role of conflict in decisions on argument termination, it is important to defuse a potential objection to the hypotheses ventured so far. At this point, in fact, some readers might feel that these predictions on argument termination, and the proposed analysis of argumentative decisions in general, rely too much on expected utility theory, which is, after all, a debated model of human choice (for an early, balanced, and still authoritative review of its merits and limits, see Schoemaker 1982; for a recent discussion of its theoretical implications, see Secchi 2011). However, expected utility is used here only for the sake of simplicity, since it allows for a very straightforward illustration of what kind of predictions can be generated on argument termination, once we assume that arguers have to make a decision about it. But the tenability of a decision-theoretic perspective on argumentation is largely independent from the particular theory of rational choice that arguers are supposed to apply. Whereas specific predictions will change according to different theories of choice, this does not undermine the case for the relevance of decision making for argumentation. On the contrary, as extensively discussed elsewhere (Paglieri 2013b), one of the key advantages of marrying argumentation and decision making is in the variety of competing theories of rational choice that become available for modeling the strategic rationality of arguers.

Still, with respect to argument termination, adopting a different theory of choice might well result in different predictions. For instance, one could speculate that arguers decide whether to terminate an argument by comparing the two options (continue vs. quit) according to various criteria (e.g., quality of the relationship with the other party, social image, self-esteem, etc.), which they learn by experience to rank in terms of their importance in guiding their choice, and thus consult lexicographically: that is, they start the comparison from the most important criterion, then proceed in order of descending importance, and stop the procedure as soon as one criterion allows to discriminate between the options (Gigerenzer and Goldstein 1996). So, with an arguer who gives greater stock to preserving good relationships than to make a strong impression as a competent speaker or boost his/her self-esteem, any discussion that endangers the quality of the relationship

will be dropped, without even considering the other criteria—contrary to what an expected utility maximizer would do.⁷

In turn, this heuristic view of decision making would also affect how duration influences argument termination: whereas prolonged discussion is likely to discourage further continuation on some criteria (as mentioned, longer arguments are more prone to create frictions and hostility between the parties; see also Gilbert 1997; Paglieri 2009), it might have the opposite effect on other criteria (e.g., the more invested arguers have become in the discussion, the more damaging backing off will be for their self-esteem). Thus, the overall impact of argument duration on argument termination would depend on what different criteria the arguers consider and how they rank them in their lexicographic search for a discriminating cue. Ultimately, the validity of different theories of choice in accounting for argumentative decisions will have to be ascertained empirically, along the lines pioneered by the studies reviewed in previous sections.

If we now turn our attention to conflict considerations, we see them looming large in factors B and C above. Among the dangers increased by prolonged discussion, both exacerbation of the original disagreement and emergence of new conflicts figure prominently, albeit not exclusively. In parallel, expectations of a positive solution to the initial conflict of opinions shrink as a function of duration, thus making arguing a less compelling option. By now, readers will be familiar with this pattern, since it is the same observed in argument engagement and argument editing: to a large extent, the decision to either continue or stop an argumentative engagement depends on how this will impact on conflict management.

7.4 Conclusions and Future Directions

Empirical evidence on a garden variety of argumentative decisions confirms that a decision-theoretic approach can make better sense of the apparent tension between a polemical view of argumentation (arguments as fights) and a rational one (arguments as rational debates). To put it simply, the expectation of rationally resolving a conflict of opinions is a prime motive in prompting us to argue, but at the same time we are keenly aware of the dangers of conflict escalation inherent to argumentative exchanges, and such awareness can lead us to avoid arguing, to proceed cautiously, or to abandon an argument before it reaches its completion. As arguers, we constantly tread the thin line between “finding an agreement” and “making things worse”: conflict is a paramount concern on both sides, but in very different ways—as something to be resolved on the one hand and as something to be avoided on the other. From these considerations, we can take home two important lessons.

⁷Incidentally, the findings reviewed in Sect. 7.3.1 provide preliminary support to lexicographic decision making in argument engagement, since arguers consider only a very limited and rigidly prioritized subset of considerations in deciding whether to enter an argument or not.

Firstly, differences of opinions, even when potentially pernicious, are the dialectical equivalent of ticking bombs, so that, unless they are handled with care, argumentation may produce early conflagration instead of safe resolution. Crucially, arguers appear to be acutely aware of this fact, given how they weight both conflict resolution and escalation avoidance as key concerns in various argumentative decisions. On the contrary, doctrines that put exclusive emphasis on finding a “reasonable” solution to the original disagreement, no matter the costs for both arguers and innocent bystanders (think of a couple so absorbed in their own debate to ignore its potential impact on their children), are a recipe for disaster, as far as social conflict is concerned. The fact that some prominent argumentation theories tend to lean in that direction (see Introduction) should give pause to any argumentation scholar interested in the practical applicability of what she/he preaches.

Secondly, the balancing act between solving a disagreement and avoiding its escalation has an important corollary: sometimes, or even often, the original difference of opinions *will not be settled by arguing*, and *arguers should not even try too hard* to settle it, lest a trivial dispute degenerates in a bitter feud. Those who feel frustrated by this limitation of our discursive practices suffer from a pernicious epistemological bias: the belief that agreement is always superior to disagreement. This is, quite simply, wrong. As argued more thoroughly elsewhere (Paglieri 2013c), some forms of disagreement are highly beneficial, both for individuals and for their social group. While the prolonged inability to reach a binding consensus on what to accomplish together is a problem for society, because it blocks the possibility of reaping the benefits of cooperation (e.g., consider the substantial costs of negotiation impasses in bargaining situations; see Crawford 1982; Svejnar 1986), a certain level of differences of opinion is inevitable in any dynamic social group, especially if its members enjoy high degrees of autonomy. Moreover, such a variety of views is often beneficial to the group itself, inasmuch as it allows the exploration of several possible courses of action and avoids premature fixation on suboptimal plans. Indeed, the accuracy of the so-called wisdom of crowds has been linked to the variety of opinions represented within a group and to the independence of judgment of its members: lacking one or both of these parameters, the collective ability to converge on a correct belief or find an effective plan of action dramatically decreases (for discussion, see Surowiecki 2004). So it would seem that well-functioning social groups do not eradicate disagreements among their members, but rather develop effective methods to negotiate temporary truces when (and only when) consensus is required on a given matter. This delicate balance between ad hoc agreement formation and a permanent reservoir of disagreement conveys important lessons also for argumentation theories. Schematically, a well-adapted “agreement ecology” includes both techniques for removing disagreements and renewable sources of further disagreement. On the contrary, a unilateral focus on the former factor runs the risk of killing the goose that laid the golden eggs, that is, promoting social practices in which the volume of agreements rapidly escalates (possibly gravitating toward very poor equilibrium points), without at the same time maintaining a healthy level of baseline disagreement (and thus precluding recovery from the local minimum where groupthink often precipitates us; see Janis 1982; Esser 1998).

Finally, turning to consider future work, it is apparent that there is plenty to be done. Research on argumentative decisions is still in its infancy, both theoretically and empirically, and yet, it shows much promise: thus further efforts should be made in this direction, for instance, looking for empirical validation of the hypotheses outlined in Sect. 7.3.3 on argument termination. Even more importantly, this research should strive for a closer integration with related studies in social psychology and communication studies. Here, we are faced with a specific instance of a general problem: the relative lack of cross-fertilization between argumentation theories and psychological studies of communication and group behavior. This is not the place to discuss the historical roots of this situation, but we still can (and should) lament its enduring existence. As of today, most argumentation scholars, in spite of their otherwise rich interdisciplinary background, are utterly unaware of things such as persuasive arguments theory (Burnstein et al. 1973) and the elaboration likelihood model of persuasion (Petty and Cacioppo 1986), even though these are key theoretical constructs in social psychology, with profound implications for our understanding of argumentative practices. In turn, most social psychologists and communication scholars remain blissfully ignorant of even the most widespread theoretical models of argumentation, such as pragma-dialectics (van Eemeren and Grootendorst 2004) or Walton's argument schemes (Walton et al. 2008). The study of argumentative decisions tries to charter the *terra incognita* between these two areas of research, with an eye to facilitate communication across them. Here, a greater focus on conflict, as a topic of interest for both communities, has the potential of offering a much needed bridge between argumentation theory and psychological research.

Acknowledgments This research was supported by the project *PRISMA—PiattafoRme cloud Interoperabili per SMARt-government*, funded by the Italian Ministry of Education, University and Research (MIUR) and by the research network *SINTELNET—European Network for Social Intelligence*, funded by the European Union. I am indebted to Cristiano Castelfranchi, Jean Goodwin, Dale Hample, Marcin Lewinski, Steve Patterson, and Frank Zenker for their helpful comments on some of the ideas detailed in this paper and to the organizers and participants of the events where this work was presented and discussed: the OSSA conference on “Virtues of argumentation” (Windsor, May 23–25, 2013), the workshop on “Computational models of natural argument” (Rome, June 14, 2013), the conference on “Conflict and communication” (Rome, October 29–31, 2013), and the ArgLab workshop on “Argumentation and rational decisions” (Lisbon, December 5, 2013). While I greatly benefited from debating my ideas with other argumentation scholars, I take full responsibility for any lingering defect of this work.

References

- Amgoud L, Maudet N (2002) Strategical considerations for argumentative agents. In: Benferhat S, Giunchiglia E (eds) Proceedings of NMR-2002. IRIT, Toulouse, pp 399–407
- Bem SL (1974) The measurement of psychological androgyny. *J Consult Clin Psychol* 42:155–162
- Benoit W, Benoit P (1987) Everyday argument practices of naïve social actors. In: Wenzel J (ed) *Argument and critical practices*. Speech Communication Association, Annandale, pp 465–474

- Benoit P, Benoit W (1990) To argue or not to argue: how real people get in and out of interpersonal arguments. In: Trapp R, Schuetz J (eds) *Perspectives on argumentation: essays in honor of Wayne Brockriede*. Waveland, Prospect Heights, pp 55–72
- Besnard P, Hunter A (2001) A logic-based theory of deductive arguments. *Artif Intell* 128:203–235
- Burnstein A, Vinokur A, Trope Y (1973) Interpersonal comparison versus persuasive argumentation: a more direct test of alternative explanations for group induced shifts in individual choice. *J Exp Soc Psychol* 9:236–245
- Cionea I, Hample D, Paglieri F (2011) A test of the argument engagement model in Romania. In: Zenker F (ed) *Argumentation: cognition & community*. Proceedings of OSSA 2011. CD-ROM. OSSA, Windsor
- Cohen D (2005) Arguments that backfire. In: Hitchcock D (ed) *the uses of argument*. Proceedings of OSSA 2005. OSSA, Hamilton, pp 58–65
- Crawford V (1982) A theory of disagreement in bargaining. *Econometrica* 50(3):607–638
- Esser J (1998) Alive and well after 25 years: a review of groupthink research. *Organ Behav Hum Decis Process* 73(2–3):116–141
- Gigerenzer G, Goldstein D (1996) Reasoning the fast and frugal way: models of bounded rationality. *Psychol Rev* 103:650–669
- Gilbert M (1997) *Coalescent argumentation*. Lawrence Erlbaum Associates, Mahwah
- Goodwin J (2005) What does arguing look like? *Informal Logic* 25:79–93
- Goodwin J (2007) Argument has no function. *Informal Logic* 27:69–90
- Hample D (2005) *Arguing: exchanging reasons face to face*. Lawrence Erlbaum Associates, Mahwah
- Hample D (2006) Argument production. In: van Eemeren FH, Hazen MD, Houtlosser P, Williams DC (eds) *Contemporary perspectives on argumentation: views from the Venice argumentation conference*. Sic Sat, Amsterdam, pp 9–22
- Hample D (2009) Commentary on Fabio Paglieri's "Ruinous arguments: escalation of disagreement and the dangers of arguing". In: Ritola J (ed) *Argument cultures: proceedings of OSSA 2009*. CD-ROM. OSSA, Windsor, pp 1–4
- Hample D, Benoit P (1999) Must arguments be explicit and violent? A study of naive social actors' understandings. In: van Eemeren F, Grootendorst R, Blair JA, Willard C (eds) *Proceedings of ISSA 1998*. SicSat, Amsterdam, pp 306–310
- Hample D, Dallinger J (1990) Arguers as editors. *Argumentation* 4:153–169
- Hample D, Dallinger J (1992) The use of multiple goals in cognitive editing of arguments. *Argumentation Advocacy* 28:109–122
- Hample D, Benoit P, Houston J, Purifoy G, Vanhyfte V, Wardwell C (1999) Naive theories of argument: avoiding interpersonal arguments or cutting them short. *Argumentation Advocacy* 35:130–139
- Hample D, Werber B, Young D (2009) Framing and editing interpersonal arguments. *Argumentation* 23:21–37
- Hample D, Paglieri F, Na L (2011) The costs and benefits of arguing: predicting the decision whether to engage or not. In: van Eemeren F, Garssen B, Godden D, Mitchell G (eds) *Proceedings of ISSA 2010*. Sic Sat, Amsterdam, pp 718–732
- Hicks D (1991) A descriptive account of interpersonal argument. In: Parson DW (ed) *Argument in controversy*. Speech Communication Association, Annandale, pp 167–181
- Infante D, Rancer A (1982) A conceptualization and measure of argumentativeness. *J Pers Assess* 46:72–80
- Infante D, Wigley C (1986) Verbal aggressiveness: an interpersonal model and measure. *Commun Monogr* 53:61–69
- Isenberg DJ (1986) Group polarization: a critical review and meta-analysis. *J Pers Soc Psychol* 50(6):1141–1151
- Janis I (1982) *Groupthink: psychological studies of policy decisions and fiascoes*. Houghton Mifflin, Boston
- Johnson R (2000) *Manifest rationality: a pragmatic theory of argument*. Lawrence Erlbaum Associates, Mahwah

- Karunatilake N, Jennings N (2005) Is it worth arguing? In: Rahwan I, Moratiš P, Reed C (eds) *Argumentation in multi-agent systems. Proceedings of ArgMAS'04*. Springer, Berlin, pp 234–250
- Martin RW, Scheerhorn DR (1985) What are conversational arguments? Toward a natural language user's perspective. In: Cox JR, Sillars MO, Walker GB (eds) *Argument and social practice*. Speech Communication Association, Annandale, pp 705–722
- Myers DG, Lamm H (1976) The group polarization phenomenon. *Psychol Bull* 83:602–627
- Paglieri F (2007) No more charity, please! Enthymematic parsimony and the pitfall of benevolence. In: Hansen H, Tindale C, Johnson R, Blair JA (eds) *Dissensus and the search for common ground*. OSSA, Windsor, pp 1–26, CD-ROM
- Paglieri F (2009) Ruinous arguments: escalation of disagreement and the dangers of arguing. In: Ritola J (ed) *Argument cultures: proceedings of OSSA 2009*. OSSA, Windsor, pp 1–15, CD-ROM
- Paglieri F (2013a) Choosing to argue: towards a theory of argumentative decisions. *J Pragmat* 59B:153–163
- Paglieri F (2013b) Argumentation, decision and rationality. In: Mohammed D, Lewinski M (eds) *Virtues of argumentation: proceedings of OSSA 2013*. OSSA, Windsor, pp 1–13, CD-ROM
- Paglieri F (2013c) Agreements as the grease (not the glue) of society: a cognitive and social science perspective. In: Ossowski S (ed) *Agreement technologies*. Springer, Berlin, pp 43–52
- Paglieri F, Castelfranchi C (2010) Why argue? Towards a cost-benefit analysis of argumentation. *Argument Comput* 1:71–91
- Paglieri F, Woods J (2011a) Enthymematic parsimony. *Synthese* 178:461–501
- Paglieri F, Woods J (2011b) Enthymemes: from reconstruction to understanding. *Argumentation* 25(2):127–139
- Petty R, Cacioppo J (1977) Forewarning, cognitive responding, and resistance to persuasion. *J Pers Soc Psychol* 35(9):645–655
- Petty R, Cacioppo J (1986) *Communication and persuasion: central and peripheral routes to attitude change*. Springer, New York
- Pollock J (1992) How to reason defeasibly. *Artif Intell* 57:1–42
- Prakken H (2000) On dialogue systems with speech acts, arguments, and counterarguments. In: Ojeda-Aciego M, de Guzmán IP, Brewka G, Moniz Pereira L (eds) *Logics in artificial intelligence: proceedings of JELIA 2000*. Springer, Berlin, pp 224–238
- Prunty A, Klopff D, Ishii S (1991) Argumentativeness: Japanese and American tendencies to approach and avoid conflict. *Commun Res Rep* 7:75–79
- Rahwan I, Larson K (2009) Argumentation and game theory. In: Rahwan I, Simari G (eds) *Argumentation in artificial intelligence*. Springer, Berlin, pp 321–339
- Riveret R, Prakken H, Rotolo A, Sartor G (2008) Heuristics in argumentation: a game-theoretical investigation. In: Besnard P, Doutre S, Hunter A (eds) *Proceedings of COMMA 2008*. IOS Press, Amsterdam, pp 324–335
- Schoemaker P (1982) The expected utility model: its variants, purposes, evidence and limitations. *J Econ Lit* 20(2):529–563
- Secchi D (2011) *Extendable rationality: understanding decision making in organizations*. Springer, New York
- Sperber D, Wilson D (1986) *Relevance: communication and cognition*. Blackwell, Oxford
- Sunstein CR (2002) The law of group polarization. *J Polit Philos* 10(2):175–195
- Surowiecki J (2004) *The wisdom of crowds: why the many are smarter than the few and how collective wisdom shapes business, economies, societies and nations*. Little/Brown, London
- Svejnár J (1986) Bargaining power, fear of disagreement, and wage settlements: theory and evidence from the U.S. Industry. *Econometrica* 54(5):1055–1078
- van Eemeren F, Grootendorst R (2004) *A systematic theory of argumentation: the pragma-dialectical approach*. Cambridge University Press, Cambridge
- Vuchinich S (1990) The sequential organization of closing in verbal family conflict. In: Grimshaw AD (ed) *Conflict talk*. Cambridge University Press, Cambridge, pp 118–138

- Walton D (1998) *The new dialectic: conversational contexts of argument*. University of Toronto Press, Toronto
- Walton D, Krabbe E (1995) *Commitment in dialogue: basic concepts of interpersonal reasoning*. SUNY Press, Albany
- Walton D, Reed C, Macagno F (2008) *Argument schemes*. Cambridge University Press, Cambridge
- Wilson D, Sperber D (2002) Truthfulness and relevance. *Mind* 111:583–632

Chapter 8

Common Ground or Conceptual Reframing? A Study of the Common Elements in Conflicting Positions in French Interactions

Alfredo M. Lescano

8.1 The Notion of “Common Ground”

According to Stalnaker, the common ground of an interaction is the set of pieces of information that the participants in an interaction accept as shared belief (Stalnaker 2002).

It is common ground that φ in a group if all members accept (for the purpose of the conversation) that φ , and all believe that all accept that φ , and all believe that all believe that all accept that φ , etc.
(Stalnaker 2002: 716)

In this formulation, φ stands for a piece of information (a proposition), say an objective description of a state of affairs towards which a speaker may have a certain attitude, such as believing in its truth or falsehood. Common ground is based on the idea that meaning consists *ultimately* in pieces of information; that is why the notion of common ground makes sense from an informational/descriptive perspective, where speakers exchange descriptions of states of affairs (i.e. pieces of information) together with their psychological attitudes towards them. This is one of the assumptions that I question. But first of all, it must be made clear that the notion of common ground can be treated independently from another notion which is usually associated with it, the notion of presupposition. In fact, the notion of common ground was created by Stalnaker to give a pragmatic account of presupposition (or we should say “the phenomena usually called *presupposition*”). The most unequivocal example of presupposition may be presented by saying that the proposition [There is a unique king of France] is presupposed by the sentence

A.M. Lescano (✉)
ENFA, EFTS (U. Toulouse); CRAL (EHESS/CNRS),
16 rue de Metz (31000), Toulouse, France
e-mail: alflescano@gmail.com

“The present king of France is bald”. Contrary to Strawson (1950), who, like Frege (1892), treated the use of sentences containing false presuppositions as *spurious* (the truth of the presupposition being a condition for the sentence to be either true or false), Stalnaker suggests that presupposition is not a problem of how the truth value of the presupposition affects that of the sentence, but rather a problem of the speakers’ attitude towards information (Stalnaker 1973, 1996, 2002). In Stalnaker’s view, presuppositions must be explained by observing what speakers take for granted when using a sentence, which implies that a sentence can have a truth value even if its presuppositions are false. Thus, a notion like “common ground” was necessary in order to theorise the pragmatic principles governing presupposition in conversation. Stalnaker suggests that to presuppose that φ is to believe that φ belongs to the common ground of the interaction. Yet, if we admit that in a given interaction, the parties involved have a particular attitude, like “to accept as shared beliefs”, towards a particular set of pieces of information, it doesn’t seem incoherent to suppose that presupposition is not the only way to denote that attitude. Arguably, a sentence like *We both believe that φ* would explicitly accomplish that task without presupposing that φ . This is the way in which we will understand “common ground” in this work.

In the next sections, I will present some descriptive problems with the idea that what is presented as “shared” has something to do with pieces of information and hence with common ground. But, of course, these are not very original claims; I am preceded by a long and strong tradition of non-truth-conditional semantics or pragmatics. However, what I believe is that it is possible to allow a different kind of description of the “shared” elements of an interaction by focusing on the way elements of content (in my terminology, “concepts”) are related. Furthermore, I will try to show that this approach, which rests upon a conception of meaning as socially dependent connections of linguistic forms rather than as individual mental states, can be useful to identify speakers’ strategies to gain the upper hand in a discursive conflict. In particular, we will try to characterise the strategy that consists in modifying the interpretation of the opponent’s point of view.

8.2 The Common Ground of a Conflicting Interaction

The first interaction we will analyse is an excerpt from a political debate. In 2007, the two main candidates in the presidential elections in France were confronted in a televised debate. Nicolas Sarkozy, from a right wing party (UMP), had been Interior Minister during the period that was just coming to an end. Segolène Royal, from the French Socialist Party, was the challenger, and was at that time in an excellent position according to the opinion polls—many people had the impression that her bad performance in the forthcoming elections was linked to her bad performance as an *orator* in precisely this debate, from which we analyse the following excerpt:

Fragment 1

S. Royal: But it is also true that [...] the ethics of politics [...] demands that politicians account for what they do. I also know that what interests the French people is the future. Nevertheless, Mr. Nicolas Sarkozy, do you think you are partly responsible for the situation in which France finds itself today?

N. Sarkozy: [...] Am I responsible for a portion of the government's record? Yes, I was Minister of the Interior for four years, I was faced with a situation that was catastrophic, which, Ms. Royal, largely accounted for the defeat of your friends, the government to which you belonged at the time¹ [...] Regarding the failures of the republic, the Left and the Right are both partly responsible²
liberation.fr

Let us reduce the conflict to its minimal expression from an informational perspective. The sentences that I list here intend to respect a declarative form so their translation into logical propositions should not be complicated, even if issues, such as the eventual presence of deontic operators, may be taken into account (here, *p* means [France finds itself in a critical situation]).

Royal's Position

(1) [Sarkozy is partly responsible for *p*]

(2) [Sarkozy must account for *p*]

Sarkozy's position

(3) (a) [The Left is partly responsible for *p*] & (b) [The Right is partly responsible for *p*]

(4) (a) [The Left must account for *p*] & (b) [The Right must account for *p*]

Proposition *p* [France finds itself in a critical situation] is presupposed both by Royal (*do you think you are partly responsible for the situation in which France finds itself today?*) and Sarkozy (*Regarding the failures of the republic, the Left and the Right are both partly responsible*), hence it is a legitimate element of the common ground of the interaction. This is unproblematic. But what I am interested in is the fact that Sarkozy's position somehow includes Royal's position. So let us look in more detail what both candidates say.

My description of Royal's position might seem debatable since she introduces proposition (1) [Sarkozy is partly responsible for *p*] using an interrogative sentence, which does not denote a propositional attitude like "believe to be true" but rather, in Fregean terms, denotes the suspension of judgment about the truth value of the proposition (Frege 1918). However, from a pragmatic perspective, one could

¹Segolène Royal was a former minister in the government of Lionel Jospin, prime minister until 2002. The Socialists were eliminated in the first round of the presidential election of 2002.

²S. Royal : Mais il est vrai aussi que [...] la morale politique [...] demande que les responsables politiques rendent des comptes par rapport à ce qu'ils ont fait. Je sais aussi que ce qui intéresse les Français, c'est le futur. Malgré tout, Monsieur Nicolas Sarkozy, estimez-vous que vous avez une part de responsabilité de la situation dans laquelle se trouve la France aujourd'hui?

N. Sarkozy: [...] Suis-je responsable d'une partie du bilan du gouvernement? Oui, j'ai été ministre de l'Intérieur pendant quatre ans, j'ai trouvé une situation qui était catastrophique, qui a compté pour beaucoup, Madame Royal, dans la défaite de vos amis, du gouvernement auquel vous apparteniez à l'époque. [...] Sur les défaillances de la République, gauche et droite, confondues, nous avons chacun notre part.

interpret it differently. Royal's sentence (*do you think you are partly responsible for the situation in which France finds itself today?*) may be described as questioning Sarkozy's beliefs: the question tends to oblige Sarkozy to publically accept or deny (1). According to Searle's framework (Searle 1969), one would say that this is not a real question about [you are partly responsible for *p*] because the speaker already "knows the answer" (there is no suspension—not even a *mock* suspension—of any judgment about this being true or false), it is actually a real question concerning the content [Sarkozy thinks that he is partly responsible . . .]. So in Stalnaker's terms, Royal does not presuppose (1), because she presents this content as a personal belief and is asking Sarkozy precisely whether he believes it too. In other words, Royal wants to know whether (1) is shared belief (i.e. common ground) or not.³

Sarkozy's position is instead quite transparent. Conjunction (3) comes from the sentence *Regarding the failures of the republic, the Left and the Right are both partly responsible*. Conjunction (4), the acceptance of the fact that the Right must account for what happened, plus the statement that the Left must do the same, can be identified in the word *record* (*Suis-je responsable d'une partie du bilan du gouvernement ? Oui . . . /Am I responsible for a portion of the government's record? Yes . . .*) which refers, in this precise context, to an explanation of how things went the way they did, and which we can consider as a synonym for the expression *rendre des comptes* (*to account for*) used by Royal.

In conclusion, from the moment that both Royal and Sarkozy accept contents (1) and (2) as true, that is to say the two pieces of information in Royal's position, these pieces of information belong to the common ground of the interaction—for the sake of argument, we will consider *Sarkozy* as synonymous with *the Right*, so that (1) is synonymous with (3)*b* (reproduced here as (5)) and (2) with (4)*b* (reproduced here as (6)):

The Common Ground of the interaction contains:

(5) [The Right is partly responsible for *p*]

(6) [The Right must account for *p*]

The first point I want to make is that by isolating pieces of information, we are separating what seems to be deeply connected in meaning. In neither of the interventions is *being responsible for p* independent from *having to account for what happened to p*. In Royal's view, (1) and (2) are not isolated descriptions of independent states of affairs, one being that Sarkozy had important responsibilities during the government still in power at the time, the other, that Sarkozy must explain why everything had gone wrong in France in the previous few years. These are not two independent entities put one next to the other; they are connected in a way

³I am not taking into account that there are non-verbal manifestations (intonation, gestures) of this kind of attitudes, because I am focusing in the verbal features of the interaction. But this must obviously be addressed in a complete study of how the "conceptual space" (cf. *infra*) is modified by the speaker's interventions.

that they make sense together. For Royal, Sarkozy must explain why everything has gone wrong in France *because* he is mainly responsible. But this causal link relating *responsibility* to *accounting for* is more profound than it may seem. It does not merely emerge as a textual causal connection between two independent propositions. This link pertains to the meaning that Royal gives to one single expression: “the ethics of politics”. Indeed, Royal presents this expression as a vehicle for a kind of *responsibility*, which is something which is followed by the obligation to *account for* what you have done during its exercise. This kind of *responsibility* is different from many others. For instance, it is different from the one generally attached to fatherhood. When we say that we are responsible for our children, we are usually saying something quite different, we are saying, for instance, that a father must “take care” of his children. But how could we describe this causal link?

In truth-conditional approaches, causality is analysed as a peripheral phenomenon as regards propositions, and by *peripheral* I mean that the causal connection is situated outside the *core* of the semantic content, the proposition itself. I will argue, along the lines of Carel’s Semantic Blocks Theory (Carel 1992, 2005, 2011) that causality (in fact, something similar to what is usually called “causality”) may appear at the deepest conceptual level. So the first way in which the common ground approach appears to be unsatisfactory is that in order to understand Royal’s position, a causal link between *responsibility* and *accounting for* must be supposed as internal to the meaning of *the ethics of politics*, while in the propositional approach, on which the common ground notion is based, this link is external to semantic content.

The second criticism of the notion of common ground specifically concerns the description of the *common* element in the conflicting positions. As we have seen, if we stick to pieces of information, we can conclude that what is shared in both Royal’s and Sarkozy’s stances is the pair formed by (5) [The Right is partly responsible for p] and (6) [The Right must account for p]. So that Sarkozy’s strategy would consist in adding these pieces of information:

- (7) [The Left is partly responsible for p]
- (8) [The Left must account for p]

Although it overlooks the causality issue, this description might allow an explanation of the fact that in Sarkozy’s reply, Royal’s point of view appears as an incomplete interpretation of the situation. Nevertheless, one of the crucial properties of Sarkozy’s strategy is that he shows his own position to be a particular manifestation of *the ethics of politics* as Royal defines it (this is developed in Sect. 8.4). This kind of strategy seems to imply non-informational contents. But even if we imagine an informational description of Royal’s *ethics of politics*, the way both positions are related does not seem something which it is possible to interpret using the common ground approach. In the next section I present a framework that avoids these problems, and come back to the analysis of this fragment in Sect. 8.4.

8.3 The Connective Concepts Framework

If semantic content is not made up of pieces of information, what is it made up of? I will say that it is made up of “concepts”. But I will give to this term a definition that is noticeably different from usual ones. In my vision, a concept is a specific kind of link between linguistic forms that can function as lexical meaning, the semantic contents of utterances and the units of social thought (i.e. what is usually treated as “stereotypes”, “common sense”, “social representations”, etc.). Let’s address the idea of “link”. In her theory about linguistic meaning, which is an outcome of Anscombe & Ducrot’s “Theory of argumentation in language” (Anscombe and Ducrot 1983), Carel supposes that some utterances link terms together by way of a *normative* link, link defined on a linguistic basis: A and B are normatively linked in an utterance U, if U can be paraphrased by using a causal or a consecutive connector, like *so, hence, thus, therefore, if... then, because*, etc., that links A and B (Carel 2011). Utterance (9) would then express a normative link between *to be a boy* and *not to cry*, since (10) could be an acceptable paraphrase of it in a particular communicative situation:

- (9) Boys don’t cry.
 (10) You’re a boy, so don’t cry.

Carel suggests that this kind of link between terms constitutes an “argumentative predicate”. Something like *Peter doesn’t cry because he is a boy* expresses the same normative link between *boy* and *cry* as (9) and (10). Carel notes this predicate by marking the link with the letters “DC”, which is an abbreviation of the French connector *donc*, but we will use instead, for clarity’s sake, the English word THEREFORE. We could say that both (9) and (10) express predicate (11):

- (11) to be a boy THEREFORE NEG to cry

One of the most important features of Carel’s view on linking is that normativity is complementary to another kind of link: *transgression*. Similarly to normativity, transgression is defined on a linguistic basis: two terms A and B are transgressively connected in an utterance U if U can be paraphrased by connecting A and B thanks to an “opposite” connector like *however, yet, although, despite, even if*, etc. For example, sentence (12) allows the expression of the argumentative predicate (13), where HOWEVER marks the fact that it involves a transgressive link (Carel notes this link with the letters PT, abbreviation of *pourtant*):

- (12) Peter cries (all the time), even if he is a boy.
 (13) to be a boy HOWEVER to cry

In our vision, a concept (hence the units of lexical meaning, the semantic content of utterances and the units of social thought) have the same semantic structure as Carel’s argumentative predicates: a concept is a normative or a transgressive connection of signifiers. It’s in this sense that we can say that a concept is a *connective* entity.

But concepts are not free atoms. For instance, as it can be noticed, concepts (11) [to be a boy THEREFORE NEG to cry] and (13) [to be a boy HOWEVER to

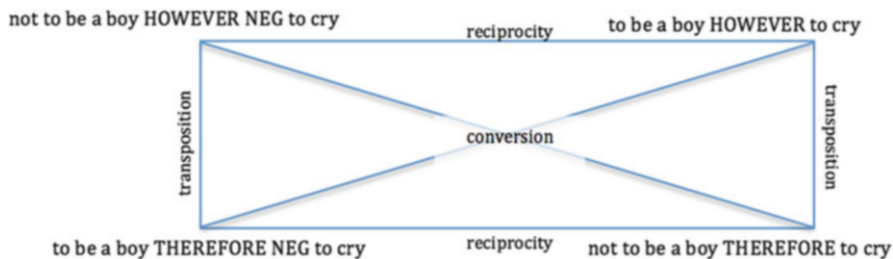


Fig. 8.1 Formal relationships between the concepts of a block

cry] share what can be intuitively called a “local world conception”, or a “micro-ideology” concerning boys and crying. In fact, concepts pertain to coherent sets of concepts called “conceptual blocks” (after Carel’s “semantic blocks”). Figure 8.1 (Carel and Ducrot 1999) outlines the formal relationships between the concepts of a block.

Let’s stress that, in our framework, contrary to Fodor (1998) (amongst others), a concept is not a mind state but a public productive structure. A concept is *public* because it acquires its concreteness when it is put into circulation. This is not to deny that the private level plays a role in the formation and the circulation of concepts, nor that concepts influence the private level of individuals. But the private level appears as *external* to the concept itself (Paveau makes a similar stance about what she calls “*pre-discours*” (Paveau 2011)).

A concept is *productive* in three ways: *discursively*, *cognitively* and *praxeologically*. A concept is *discursively productive* because it can engender infinite discourses (concept (11) can engender discourses where boyhood appears as interpenetrated with the absence of crying); it is *cognitively productive* since it can become one of the structures through which individuals and social groups perceive (the way boys and crying are perceived, for instance); and it is *praxeologically productive* because it can be applied by performing an action (a boy may apply concept (11) by avoiding to cry).

So, in a way, many conflicting interactions can be seen as struggles in which what is at stake is the productivity of concepts. In conflicting dialogues, as well as in other kinds of interaction, to refute a concept is to try to cease the productivity of this concept.

Finally, the analysis of conflict is crucially dependent on the way we conceive interaction. When people interact, individual activity is subordinated to what the participants suppose is going on in a shared space. This is maybe quite visible in a multi-party interaction, like a business meeting in which speaking turns are strictly regulated. One can imagine a very schematic interaction where the first speaker states a position, then a second participant makes this position more precise, a third one refutes the precision of the second participant but maintains the first stance, and so on. All of this happens at a level that is not the individual mind, but a collective space. Not everyone in the interaction has exactly the same idea of what is going

on, but if one wants to describe what is happening with respect to the concepts involved in the interaction, one can objectivate a common dynamic “workspace”. So that when studying how people manipulate concepts in interactions, we need a notion that allows the description of the operations performed by each individual intervention at the collective level. We must be able to say, at a given point of the interaction, whether a concept has been added, modified, refuted, etc. We call this instance the “conceptual space” of the interaction.⁴ Even if this idea raises many problematic questions that require our attention (what about plurivocity? is this the only thing happening at the collective level of an interaction?, etc.), we will treat it as an operative notion that makes the description of the conceptual level of interactions accessible. I suppose that every concept or conceptual configuration that will be postulated as describing (a part of) a particular point in the evolution of an interaction, takes place within its conceptual space.

8.4 The Common Element in Two Conflicting Positions in a Connective Framework

In this section, we shall study a strategy set up in replies within conflicting interactions. This strategy involves a conceptual operation that is defined as *reframing*. Generally speaking, reframing consists in reinterpreting the opponent’s point of view, by modifying its *frame*, i.e. the concept that gives the key to the interpretation of the particular point of view. We describe two different kinds of reframing. If we call *A* the speaker to which the reply is addressed and *B* the one who addresses the reply, we could schematise the two kinds as follows:

- B reframes A’s position:
 - By respecting the frame that A proposed (internal reframing) (Sect. 8.4.1)
 - By proposing a higher level frame (external reframing). Two different cases of external reframing are considered (Sect. 8.4.2).

8.4.1 Internal Reframing

Let us now begin the connective analysis of the excerpt from the debate between Royal and Sarkozy. First of all, we make a connective interpretation of what Royal

⁴Some of the terms that are used here to refer to technical notions (like “conceptual space”, “frame”, “reframing”...) are already in use to refer to notions that are not comparable to those defined here, mainly because they do not describe what is going on in a shared space but in individual minds. One of the most used meanings of the term “conceptual spaces” comes from the cognitive theory presented in (Gärdenfors 1995). For a good account on the cognitive conception on “frames” and “reframing”, cf. (Kaufman et al. 2003).

calls “the ethics of politics”. As we have already seen, Royal means by that a particular kind of responsibility, one that engenders the obligation on the part of the subject of this responsibility, to account for what happened to the object of the responsibility while exercising it.

The Ethics of Politics (according to Royal)

(14) [to have the responsibility for something THEREFORE to account for what happened to it]

If one accepts this, a straightforward description of Royal’s and Sarkozy’s positions is possible. According to Royal, Sarkozy must account for the way things went in France because he was (or was partly) responsible for its government, so she states the following concept:

Royal’s position

(15) [Sarkozy is partly responsible for the situation in which France finds itself today THEREFORE he must account for it]⁵

Sarkozy shows himself as respecting “the ethics of politics”, he does not contest the link between having the responsibility for something and the obligation to account for what happens when exercising this responsibility: instead he contests being himself solely responsible for the situation. Sarkozy’s position can be described using almost the same concept that describes Royal’s position, the difference being that in his position, besides the Right (or himself), the previous Socialist party government has also held a *responsibility that engenders accounting for*.

Sarkozy’s position

(16) Left and Right are both partly responsible for the situation in which France finds itself today THEREFORE Left and Right must account for it

Sarkozy not only accepts in his discourse “the ethics of politics” as defined by Royal, he even situates his own position within this concept, but does so in a way that makes Royal’s position appear as a particular (partial) interpretation of his own (larger) vision of the world. Hence, it becomes necessary to consider the way concepts can relate to one another. Here, the relevant relationship seems to be “specification” (Lescano 2013). A concept C1 *specifies* another concept C2, when C1 appears—in a given conceptual space—as a particular version of C2. One crucial feature of specification is that the more abstract concept of the two plays a role in the interpretation of the more specific one, in a way that makes it pertinent to say that the more abstract concept functions as the conceptual “frame” of the more specific one. If we take the concept [to be tired THEREFORE NEG to work], it is easy to see

⁵If we remember that a concept is a connection between signifiers, it may become suspicious to make concepts connect entire sentences. Nevertheless, the whole sentence may function as a unique signifier which relates to other sentences-signifiers with different kinds of links. Normative and transgressive links are the links that form concepts, but that does not exclude the existence of other kinds of links, since the sentence-signifier, as well as every signifier, may be linked to other concepts in a given conceptual space.

that it can be interpreted in different ways depending on its conceptual frame. It can be interpreted as the manifestation of *lazy* behaviour, if it appears as a specification of [obstacle THEREFORE NEG to do], assuming that *laziness* corresponds to the behaviour of people who do not work whenever the smallest obstacle intervenes, and that the fact of being tired can function as an obstacle to action (Figs. 8.2). On the contrary, if the conceptual frame is the concept [risk THEREFORE to avoid], *not working because of being tired* is not to be *lazy* but to be *careful* (8.3). In Figs. 8.2 and 8.3, the embedded concept specifies the “embedder”, which acts as a frame. As we shall see, the specification relationship may take, as the most specific concept: (1) a concept that functions as the frame to another concept (so that specification can be characterised as being recursive) as well as (2) two (or maybe more) different concepts at the same time.

Conceptual specification plays an important role in Sarkozy’s strategy. Let us proceed step by step. Royal sets first a general concept that she associates with the expression “the ethics of politics” and then gives a specific version that is no longer *general* but specifically concerns Sarkozy, his responsibility as the leader of France and hence his obligation to account for how things went in France (Fig. 8.4).

As we saw, Sarkozy does not oppose the general concept that defines “the ethics of politics”, even recognising himself that he is partly responsible for the current situation and accepting implicitly the possibility of accounting for it. Instead, he proposes something close to a correction to Royal’s more specific concept, a correction that *reframes* Royal’s position: Royal’s demand to Sarkozy no longer appears *only* as a specification of the ethics of politics, but *also* as specifying a concept where Left and Right together are items of a *responsibility which engenders accounting for*. Sarkozy’s strategy consists in, on the one hand, maintaining the

Fig. 8.2 Not to work because of being tired is to be lazy

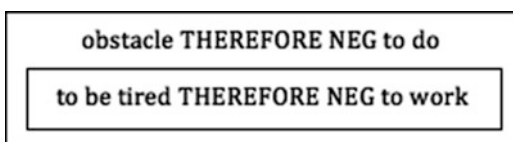


Fig. 8.3 Not to work because of being tired is to be careful

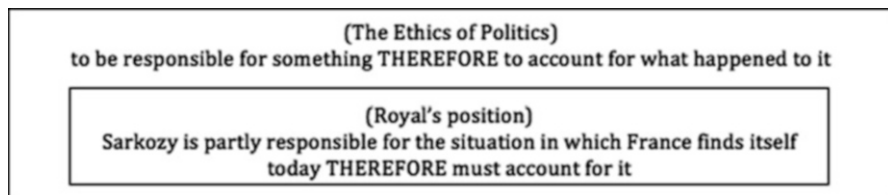
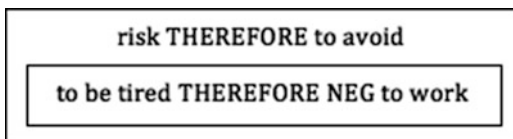


Fig. 8.4 A region of the conceptual space of the interaction as determined by Royal’s intervention

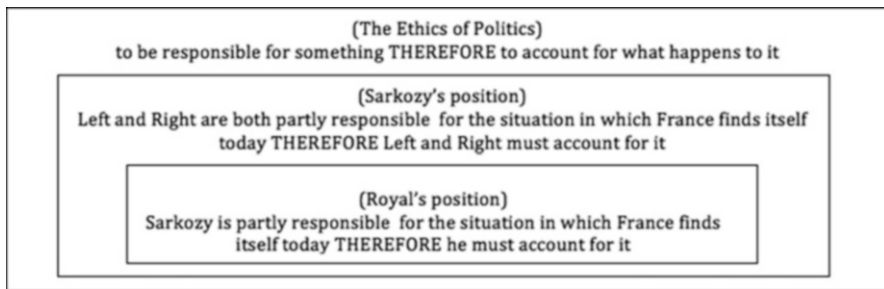


Fig. 8.5 A region of the conceptual space of the interaction as determined by Sarkozy’s intervention

ethics of politics as the frame of his own concept, and on the other hand, setting his own position as the frame for Royal’s demand. This strategy gives the impression that Royal’s demand is a *partial* view of the situation, while Sarkozy’s position appears to be more general in scope (as far as history is concerned), as well as not escaping from the laws of the ethics of politics (Fig. 8.5).

The fact that the reframing of Royal’s position performed by Sarkozy takes place within the frame that Royal herself has set, leads us to class this as a case of *internal* reframing. The new frame is internal with respect to the previous frame of the reframed concept, i.e. the frame in which Royal includes her position. In the following sections we examine two cases of external reframing.

But, before going through the next analysis, let us briefly point out how the description we have just made can shed light on our initial claim of the insufficiency of the notion of common ground to grasp the common element in conflicting interactions.

The first reason for the unsatisfactory result of the “common ground” view is that—given the informational perspective to which it pertains—it prevents us from describing some concepts which seem to be relevantly described as connective entities, like “the ethics of politics”, which we have identified as the frame of both positions. The ethics of politics cannot be defined as the conjunction of two states of affairs [X is responsible for p] and [X must account for what happens to p]. The ethics of politics, as Royal defines it, conceptually connects responsibility to accounting for.

The second reason is that by focusing on the pieces of information that the speakers accept as true rather than on the relationships between semantic entities that result from discursive activity, the common ground notion lacks descriptive adequacy regarding the strategies used in conflicting discursive interactions, even when these strategies concern entities that appear to be common to both conflicting positions. We cannot describe the strategy of reframing the opponent’s position, which involves “*keeping it but giving it a different interpretation*”, by describing what is believed to be true by both participants. The result of applying the notion of common ground would be a list of what is accepted as true by both participants, but not how the replier’s position affects the opponent’s one.

In the next sections, I present two other kinds of strategies that involve presenting a common concept as the frame to two conflicting positions, which tend to show that connective concepts, specification and reframing might be useful tools to reveal the nature of the operations performed in the conceptual space of conflicting interactions.

8.4.2 External Reframing

We have just seen a first reply strategy in a conflicting interaction which consists in making the adversary's position seem too specific, partial, local, with respect to *the real situation*. This "real situation", which is nothing but the replier's position, becomes what I have called the *frame* of the opponent's position, a more abstract concept that reinterprets the adversary's statement. This strategy is characterised by the respect of the *frame* of the opponent's position, and that is why it can be viewed as an *internal reframing*. The excerpts that we will analyse in this and the following sections, concern two different kinds of external reframing, defined as the strategy of setting the highest level frame for the reinterpreted position. The two kinds of external reframing differ in that while the first replaces the frame proposed by the opponent (substitutive reframing), the second encompasses it in the new frame (comprehensive reframing).

8.4.2.1 Substitutive Reframing

The following interaction takes place on a forum of a French web site addressed to women (*aufeminin.com*). A member of the forum, that we shall call "ONE", starts a "topic" by expressing concern about the growth of racism in France. The day after, the user we call "TWO" replies to ONE by giving a different view of the situation.

Fragment 2

ONE⁶ posted on 06/23/10:

With no intention of playing the lefty or run of the mill idealist, I think people are becoming more and more racist in France. The French affirm particularly loudly and clearly, their hatred of Blacks and Arabs [. . .] I don't know where we're going with these mentalities, but it's frightening [. . .] I admit that there are problems in the suburbs, there are the idiots that don't give a shit and provoke, but that doesn't excuse everything. Tone down your language! You cannot love everyone, that's true, but I hope there are still tolerant people in France?⁷

⁶The surnames of the participants have been changed.

⁷Je ne vais pas jouer la gauchiste et idéaliste de bas étages mais je trouve que les gens sont de plus en plus racistes en France. Les français affirment haut et fort leur haine anti-noirs et anti-arabes en particulier [. . .] Je sais pas ou on va avec ces mentalités, mais ça fait peur [. . .] Je veux bien admettre qu'il y'a des problèmes en banlieue, y'a des petits cons qui foutent la merde et qui

TWO replies to ONE on 06/24/10:

We must understand how exasperated and “fed-up” the French, who are forced to endure situations they never asked for, feel. [. . .] Yes, it’s disturbing, but unfortunately I think it’s too late to turn the clock back [. . .]⁸
aufeminin.com

According to ONE, explicit racism against Blacks and Arabs is growing among the French. In our terms, she describes the French using a THEREFORE connection between “racial” categories and hate: “the French” are characterised by a concept in which being Black or Arab engenders hate—obviously (17) is not the only concept expressed by ONE, but we focus on it to describe the way the two posts relate to each other.

(17) to be Arab/Black THEREFORE to hate him/her

Concepts like this one, which construct a kind of hate with a racial basis manifestly constitute one of the many types of concepts that can epitomise a racist ideology. Words like *racist* or *racism* must be stably associated to a concept similar to (17) but in a more abstract form, like (18).

(18) x has other ethnic origins than y THEREFORE NEG y sympathises with x

This ties in with the description that would receive the typical justification *I’m not racist! I’ve got Black/Arab friends*, which evokes a specific version of the converse concept⁹ to (18), namely (19)—which hence belongs to the same conceptual block:

(19) some of the people I know are Black HOWEVER they are my friends

ONE’s position is analogous to Royal’s position in that concept (17) [to be Arab/Black THEREFORE to hate him/her] specifies a more general concept, i.e. (18) [x has other ethnic origins than y THEREFORE NEG y sympathises with x]—lexically associated to *racism* or *racist*—, the latter functioning as a conceptual frame, hence giving an interpretation of (17) as manifesting *racism* in a particular way (Fig. 8.6).

This is a schematic view of the conceptual counterpart, in particular, of this fragment of ONE’s post: *I think people are becoming more and more racist in France. The French affirm particularly loudly and clearly, their hatred of Blacks and Arabs*. In her reply, TWO does not at all argue against this vision of the French as hating Black and Arab people precisely because they are Black and Arab. Instead,

provoquent mais ça n’excuse pas tout. Modérez vos propos ! On peut pas aimer tout le monde, c’est vrai mais j’espère qu’il existe encore des gens tolérants en France ?

⁸Il faut aussi comprendre l’exaspération et le ras-le-bol des français qui subissent des situations qu’ils n’ont jamais demandé. [. . .] oui, c’est inquiétant, et malheureusement je crois que maintenant, on ne pourra plus faire marche arrière [. . .]

⁹Conversion is a formal relationship between two concepts. There are four possible converse pairs: [A THEREFORE B] and [A HOWEVER NEG B]; [NEG A THEREFORE NEG B] and [NEG A HOWEVER B]; [A THEREFORE NEG B] and [A HOWEVER B]; [NEG A THEREFORE B] and [NEG A HOWEVER NEG B]. Cf. Carel, 2011.

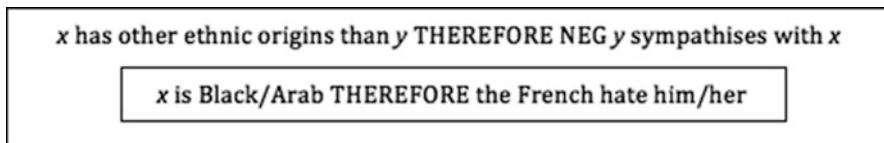


Fig. 8.6 A region of the conceptual space of the interaction as determined by ONE's intervention

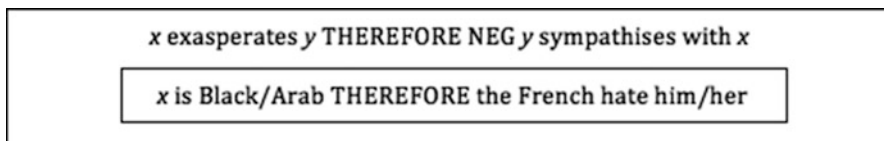


Fig. 8.7 A region of the conceptual space of the interaction as determined by TWO's intervention

she aims to reinterpret this attitude as the expression of the “exasperation” caused by these ethnic groups. For her, French hatred of Arabs and Blacks stems from the fact that Arabs and Blacks make French “endure situations they never asked for” (which can be easily interpreted as evoking a commonplace which links immigration to crime and violence, but also those concerning “irritating” habits like praying in public places, etc., all of which receive a high level of coverage in the French media). In this manner, TWO proposes a modification to the conceptual space as ONE sets it, but she does not express a completely dissimilar version of reality. In fact, she retains concept (17) (according to which being Black or Arab engenders French's hatred), but she embeds it in a different frame, and hence reinterprets it. According to TWO, being Black or Arab engenders French's hatred because these populations have “exasperated” the French, and not because of a racist ideology anchored in the French population. Figure 8.7 represents TWO's position.

To summarise, TWO replaces the frame of ONE's position by a concept which is not an accusation of the French, but of Blacks and Arabs, who engender hate by their “exasperating” behaviours. In other words, she retains the most specific component of ONE's position, yet she gives a new interpretation of it. So that, similarly to what happens in the fragment we have analysed from the Royal/Sarkozy debate, the *core* concept of the two conflicting positions of Fragment 2 is precisely their common element. That is because TWO, like Sarkozy, replies by reinterpreting (or technically, by *reframing*) the most specific concept of her opponent's position. Yet TWO's strategy is not exactly the same as Sarkozy's. In fact, whereas Sarkozy respects the frame proposed by Royal (he accepts to talk within the frame of “the ethics of politics” as Royal conceives it), TWO replaces the frame of ONE's position, substituting a concept associated to *racism* with another that presents the French as victims of exasperating behaviours of Blacks and Arabs and whose hatred is an “understandable” reaction to those behaviours. In other words, for ONE, the French are racist, while for TWO if French hate Blacks and Arabs, it's rather Black's and Arab's fault.

The main difference between Sarkozy's and TWO's reframing strategy is that while Sarkozy *corrects* Royal's position *from within her own perspective* (that is

why it is an *internal* reframing), TWO situates ONE's most specific statement *in a new perspective*, hence accomplishing an *external* reframing. In order to distinguish this kind of external reframing from the one that is analysed in the following section, we class it as a *substitutive* external reframing.

As what concerns these conceptual strategies as *argumentative* strategies, one can notice that internal reframing makes much more difficult to the speaker A (in our case, S. Royal) to counter-attack (to try to cease the productivity of her opponent's concepts), since speaker B is actually keeping the frame that speaker A had proposed: speaker A is prisoner of her own statement, which has been now turned against her position. Instead, when a speaker performs a substitutive reframing, she gives a very different account of the situation proposed by speaker A, so that this kind of reframing is overtly polemical. Hence, after a substitutive reframing performed by speaker B (in our case, TWO), speaker A (ONE) is free to develop her own point of view as well as to attack speaker B's vision, because only the most specific part of her statement is kept within the opponent's position.

How would a description of this interaction using the notion of common ground give an unsatisfactory result, as I claimed before? First of all, by applying the notion of common ground we would have found that both speakers agree that *the French hate Blacks and Arabs*, but we could not have accounted for the fact that this hatred is not just a coincidence, as if we said that the French hate Blacks and Arabs independently of their origin or colour of skin (those who are hated just *happen* to be Black and Arab). According to this interaction, there is something in the fact of being Black and Arab that *engenders* hate, "something" which is interpreted by ONE as a *racist attitude* and by TWO as some kind of *understandable behaviour*. More precisely, on this level, if we wanted to describe the common ground of this dialogue, we could say that the piece of information that both ONE and TWO accept as true is:

The Common Ground of the interaction contains:
(20) [the French hate the Arabs and the Blacks]

However, unlike what happens in the debate between Royal and Sarkozy, one could say that the impossibility of describing the causal link that connects being Black or Arab to hate, which is included in both positions, does not prevent us from describing the common element in both points of view. This link would just be left underspecified, yet the piece of information (20) could count as being what the two participants think to be true (that is not the case with respect for "the ethics of politics", which is irreducible to a description of a state of affairs). On the contrary, if we applied the common ground notion, we would not succeed in obtaining a clear representation of the way the two positions differ. As far as I know, there is no tool for describing the idea of frame, and *a fortiori* that of frame substitution, within an informational perspective. There is no room in the meta-language of an informational approach to see a semantic entity, like (20),

as a particular manifestation of a more abstract one (this is not comparable to an extensional relationship like, for instance, “inclusion” since what is at stake here is not the elements of sets but connections between linguistic forms).

8.4.2.2 Comprehensive Reframing

In the first strategy we have analysed, the one we have called *internal reframing*, the replier maintains the frame proposed by his opponent, while in the one we have just described, the replier’s position is formed by giving to the opponent’s position a new frame. Because of the fact that the resulting highest level concept is set by the replier, we qualify this strategy as an *external reframing*. We noted also that in the case of the latter, the replier *replaces* the opponent’s frame with her own frame, and that is why this kind of external reframing can be named *substitutive reframing*. If we need this detailed classification it is precisely because in this section we look at another case of external reframing, which is perhaps a more complex strategy, in which the replier reinterprets not only the most specific position but also its frame, by adding a higher level frame. For this reason, this kind of external reframing is called *comprehensive reframing*.

Let us consider Fragment 3. Some contextual elements are useful: the French Interior Minister, Manuel Valls, said in a recent interview that Romani people who are in France are destined to return to their countries of origin (Bulgaria and Romania). The Ecology Minister, Cécile Duflot, strongly opposed this statement by saying that it does not respect French republican values. After this, a highly publicised poll said that two thirds of the French agreed with Valls. This is—roughly speaking—the context of this brief interaction found on the forum of the website of a well-known French magazine, *Le Nouvel observateur*:

Fragment 3

ALPHA¹⁰ posted on 3-10-2013 at 09:57.

A CSA-BFM TV poll released Wednesday, October 3 said that two thirds of French are closer to Valls than to Duflot. No! This is untrue! Totally untrue! The survey lies, the survey does not tell the truth. The truth is that Valls is closer to 2/3 of the French than Duflot. If tomorrow two thirds of the French were opposed to the “bournats” who were intending to take over all the bars in France again, Valls would be closer to that 2/3. Valls knows the laws of populist mathematics very well. Regardless of the proposed axiom, if it is the axiom of 2/3 of the French, Valls will follow it [...]¹¹

BETA posted on 3-10-2013 at 11:00. (Reply to ALPHA’s post)

¹⁰The surnames of the participants have been changed.

¹¹Un sondage CSA-BFM TV publié mercredi 3 octobre dit que 2/3 des Français sont plus proches de Valls que de Duflot. Non c’est faux !! archi faux !!! Le sondage ment, le sondage ne dit pas la vérité. La vérité c’est que Valls est plus proche de 2/3 des Français que Duflot. Si demain les Français étaient pour les 2/3 opposés aux bournats qui se remettraient à accaparer tous les bistrotts de France, Valls serait plus proche de ces 2/3 là. Valls connaît par excellence les lois de la mathématique populiste. Peu importe l’axiome proposé, si le chiffre de 2/3 le tient pour axiome, Valls le suivra [...]

The French [. . .] are exasperated. That politicians attempt to surf on this exasperation—I mean, in their discourses—is quite normal. This is not populism, as the ill-intentioned would say. It’s just the game of democracy, the politician must RESPOND to the desires of those who voted or intend to vote for him [. . .]¹²

Nouvelobs.com

In order to get an image of the common ground in this interaction we could start by looking for pieces of information that can be considered to be believed as true by both participants. I take into account only two pieces of information: (21) and (22).

The Common Ground of the interaction contains:

(21) 2/3 of the French are for the return of the Romani to Romania and Bulgaria

(22) Valls is for the return of the Romani to Romania and Bulgaria

It is easy to see that (21) and (22) are accepted as true by ALPHA and are taken for granted by BETA. We could then state that (21) and (22) are part of the common ground of this interaction. I know I am oversimplifying but this extremely simplified version reveals the core of the problem. Exactly as in Royal’s and Sarkozy’s positions, we cannot describe ALPHA’s position by juxtaposing pieces of information. We could summarise ALPHA’s opinion by saying that she states that what is going on qualifies as a specific instance of some general idea that she calls the “populist mathematics”. But which piece of information may qualify to instantiate this general way of doing politics? If ALPHA sees a manifestation of populism it can neither be in (21) nor in (22) taken as isolated pieces of information. Take (22), the information about Valls. It just describes Valls as being for the return of the Romani to their countries of origin. That is not, in itself, what ALPHA is calling populism. If we don’t relate it to (21), it does not manifest Valls’ populism that ALPHA is talking about.

What manifests populism, for ALPHA, is the THEREFORE link between Valls’ declarations and popular opinion. We can note this entity as in (23):

(23) two thirds of the French are for the return of the Romani to R&B (Romania and Bulgaria) THEREFORE Valls is for the return of Romani to R&B

ALPHA opposes this concept to the one that she associates with the poll, namely (24):

(24) Valls is for the return of the Romani to R&B THEREFORE 2/3 of the French are for the same thing

In this concept—that ALPHA rejects—, Valls’ opinion about the Romani appears to be the cause of public opinion. According to the poll, people would follow Valls, while ALPHA’s position is that Valls follows popular opinion.

We can see, once again, that the THEREFORE connection may occur at the deepest conceptual level in the impossibility of describing what *populism* is—

¹²[. . .] Les Français [. . .] sont exaspérés. Que des politiques tentent de surfer sur cette exaspération—en paroles s’entend—est tout à fait normal. Ce n’est pas du populisme comme diraient les mal-pensants. C’est simplement le jeu de la démocratie, l’homme politique devant REpondre aux désidérata de ceux qui votent ou ont l’intention de voter pour lui [. . .]

at least for ALPHA—if we do not take this conceptual linking into account. For ALPHA, *populism* is the very fact that popular opinion conditions political decisions. We can obtain this general concept of what ALPHA calls *populist mathematics* by giving a more abstract form to (24).

Populist Mathematics (according to ALPHA)

(25) the opinion of the majority is something THEREFORE the politician defends that something

If we recall that the “common ground” isolates pieces of information, we can see why this notion fails to describe what populism is for ALPHA, hence how ALPHA depicts Valls. We cannot describe what *populism* is here by saying that on the one hand, we have popular opinion, and on the other hand, a politician who defends a point of view—even if it happens to be the same point of view. From an informational perspective—which is, I insist, the perspective within which the notion of “common ground” makes sense —, if you have established a cause-consequence relationship it is because you have performed some kind of operation on propositions, not within them. My claim is that understanding what *populism* is for ALPHA involves a THEREFORE connection that is internal to concept (25).

Let us turn now to what is common to both positions. ALPHA’s position, which I schematise in Fig. 8.8, is formed by concept (25)—which describes Valls’ attitude—specifying in a particular way concept (23)—which defines populism.

In this figure I intend to show that, in ALPHA’s text, Valls’ attitude appears as a particular manifestation of *populism*. Let us remind ourselves that, in a conceptual specification like this, the more abstract concept acts as the frame for the more specific one. Here (25), which defines *populism*, is the frame in which the link is made between popular opinion and Valls’ opinion.

Conceptual specification plays also an important role in BETA’s strategy. Let us see how it works. Concept (26) gives a connective form to BETA’s position about Valls.

(26) voters desire the return of the Romani to B & R THEREFORE Valls responds to this desire

In BETA’s discourse, when Valls responds to the desire of voters, he acts by strictly *playing the game of democracy*, that is to say that in BETA’s text, concept

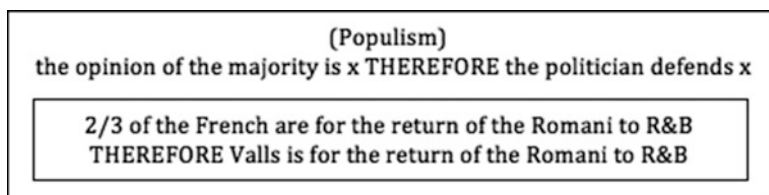


Fig. 8.8 A region of the conceptual space of the interaction as determined by ALPHA’s intervention

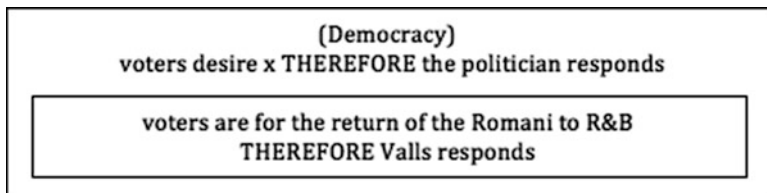


Fig. 8.9 A region of the conceptual space of the interaction as determined by BETA's intervention

(26), concerning Valls, is a particular specification of a more general concept involved in the very nature of democracy, which I note in (27):

The Game of Democracy (according to BETA)

(27) voters desire something THEREFORE the politician responds to this desire

Figure 8.9 schematises the way BETA's position can be described with a conceptual specification, which respects the same pattern that we attributed to ALPHA's position: a concept concerning Valls specifies a more abstract concept defining a way of doing politics.¹³

Now, what is the conceptual element common to both these positions? This element stems from a subtle difference between ALPHA's and BETA's views: the opposition between the verbs *to follow* (ALPHA) and *to respond* (BETA). *To respond to the desire of the people* includes the possibility of not following popular opinion. So that *to follow popular opinion* appears as a specific kind of *response*. This means that BETA's *democracy* concept (which is about *responding*) appears as a more abstract (or a less specific) concept than ALPHA's *populism* (which is about *following*). Put another way, by defining *democracy* by encompassing ALPHA's concept of *populism*, BETA treats ALPHA's description of Valls as a particular interpretation of democratic behaviour. In fact, if BETA can qualify ALPHA as ill-intentioned, it is because she presents ALPHA as specifying the concept of democracy in a way that makes *following popular opinion* the only possible kind of *response* Valls can make to satisfy the desire of voters. Consequently, BETA makes her own view of *democracy* function as the common conceptual frame for both positions.¹⁴

¹³It is true that BETA does not mention Valls explicitly, but concept (26) is triggered by the fact that BETA's post appears as a reply to ALPHA's, so the link to Valls is made by the textual association of the two posts.

¹⁴If we compare the three descriptions we have made, we can see that *specification* is the generic name for a variety of possible relationships. In Sarkozy's case, the fact that Royal's position appears as a specification of his own view, makes Royal's position appear to be partial in the sense of *incomplete*. In the example of French hatred towards Arabs and Blacks, the most specific concept appears as the "fact" to which the frame provides an interpretation. In the present interaction, to include the opponent's position in a more comprehensive frame allows for it to be qualified as a biased (ill-intentioned) position.

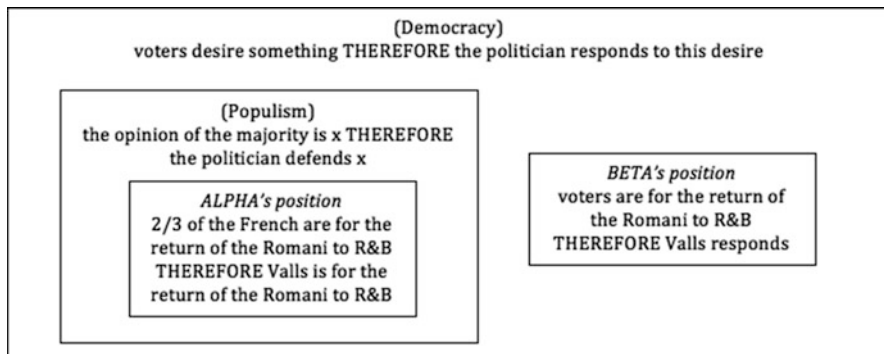


Fig. 8.10 A region of the conceptual space of the interaction as defined by BETA's reply

A final interesting element of BETA's strategy is that by situating the two positions as specifying the same frame, she gives a view of the conflict as confronting *two different ways of seeing the same thing*. This is similar to the previous analysis in which TWO's reply accepts that the French hate the Blacks and the Arabs yet not taking it as a racist attitude but as "understandable" behaviour. However, what is specific to BETA's strategy is that she presents her opponent's position as a particular manifestation of a higher level concept¹⁵ (Fig. 8.10).

Would ALPHA agree with this? Maybe or maybe not. One can imagine that ALPHA has a different conception of democracy, but this is not the point. What counts here is that BETA has made ALPHA's position enter into a conceptual frame that appears to be common to both of them. In order to get out of this frame, ALPHA would have to show in what way what she calls *populism* is not a democratic attitude. Thus, comprehensive reframing appears as a highly efficient argumentative strategy: by making the opponent's position appear as a partialised vision of reality, the speaker who reframes is associated to the "ethos" of an *orator* who has a more impartial vision of the situation.

8.5 Conclusion

In this paper, my intention has been to show that taking content to be connections of linguistic forms instead of pieces of information reveals a particular kind of replying strategy in conflicting interactions, and, at the same time, allows a more accurate

¹⁵In this chapter I have not dealt with the problem of the "status" of the concepts within the conceptual space. In BETA's reply, ALPHA's position is given a "rejected" status, whereas this is not the case in the previous fragments we analysed. Other related issues are overlooked here, like the fact that BETA treats ALPHA as *ill-intentioned* because of her use of the term *populism* and not because of the concept that is asserted.

description of the individual positions per se. In particular, we have seen that a possible strategy in replying consists in reinterpreting the other speaker's position by encompassing it in a more abstract concept, strategy which we have referred to using the term *reframing*. This idea rests upon the hypothesis that in a given conceptual space, two concepts can enter into a relation (*specification*) in which one of them appears as a specific version of the other one, which functions as its frame, and hence participates in the interpretation of the more specific concept.

We have identified two main kinds of reframing, one in which the reply keeps the original frame, i.e. the most abstract concept, of the opponent's position as the higher level frame (internal reframing), and one in which the higher level frame is set by the replier (external reframing). The latter case has been illustrated by two of its possible sub-kinds: in the first one, the frame set by the replier replaces the one that belonged to the opponent's position (substitutive reframing) and the second embeds the opponent's frame into a new one (comprehensive reframing). This is obviously not the whole catalogue of every possible reframing strategy, but the heterogeneity between the cases we have analysed suggests that the reframing strategy might be a generalised phenomenon that manifests itself in a variety of procedures.

To accept the existence of a strategy like "reframing" challenges the idea of conflicts as opposing contrary contents, as well as a conception of discursive interaction as an exchange of pieces of information accompanied by propositional attitudes. To reframe is to modify the way that the position defended by the interlocutor must be interpreted. In other words, two conflicting positions may have in common the precise element that defines their most specific item, while they differ in their most abstract component. This has nothing to do with a picture in which the two speakers are opposed in believing one that P is true and the other that P is false (or that non-P is true). The conception of the discursive interaction that the strategy of reframing depends on sees speakers' interventions as performing operations within a collectively elaborated conceptual space. This space is not comparable to Stalnaker's common ground on several counts. Firstly, it is not composed, at any level, of pieces of information. Secondly, its elements are not determined by what the speakers believe to be mutually accepted, but by whatever concept is set by the utterances of the interaction, even when it is explicitly shown to be an unshared position. Thirdly, it is an organised space, in which elements are not isolated items on a list; the specification relationship is just one of its possible organisational principles. Thus, a conflicting interaction seen from this angle appears as a struggle to set the final form of the resulting conceptual space, hence—if we recall that concepts are productive in what concerns discourse, thought and action—to enlarge or maintain the productivity of certain concepts, as well as diminishing or ceasing the productivity of others.

References

- Anscombe J-C, Ducrot O (1983) L'argumentation dans la langue. Mardaga, Bruxelles
- Carel M (1992) Vers une formalisation de la théorie de "l'argumentation dans la langue". École des hautes études en sciences sociales, Paris
- Carel M (2005) La construction du sens des énoncés. *Revue romane* 40(1):79–97
- Carel M (2011) L'entrelacement argumentatif. *Lexique, discours et blocs sémantiques*. H. Champion, Paris
- Carel M, Ducrot O (1999) Le problème du paradoxe dans une sémantique argumentative. *Langue française* 123:6–26
- Fodor J (1998) *Concepts: Where cognitive science went wrong*. Oxford UP, Oxford
- Frege G (1892) Sens et dénotation. In: Imbert C (ed) *Écrits logiques et philosophiques*. Seuil, Paris, pp 102–126, 1971
- Frege G (1918) La négation. In: Imbert C (ed) *Écrits logiques et philosophiques*. Seuil, Paris, pp 194–213
- Gärdenfors P (1995) Conceptual Spaces. *Kognitionswissenschaft* 4:185–201
- Kaufman S, Elliott M, Shmueli D (2003) Frames, framing and reframing. In: *Beyond intractability*. Burgess G, Burgess H (eds.) Conflict research consortium. University of Colorado, Boulder. <http://www.beyondintractability.org/essay/framing/>. Accessed June 10, 2009
- Lescano A (2013) Stéréotypes, représentations sociales et blocs conceptuels. *Semen Revue de sémio-linguistique des textes et discours* 35:153–170
- Paveau MA (2011) Quelles données entre l'esprit et le discours? Du préconstruit au prédiscours. In: Azouzi A (ed.) *L'analyse du discours. Notions et problèmes*. Éditions Sahar, Tunis, pp. 19–37
- Searle J (1969) *Speech acts: an essay in the philosophy of language*. Cambridge University Press, Cambridge
- Stalnaker R (1973) Presuppositions. *J Philosophical Logic* 2:447–457
- Stalnaker R (1996) Pragmatics. In: Geirsson H, Losonsky M (eds.) *Readings in Language and Mind*. Blackwell, Cambridge 77–88
- Stalnaker R (2002) Common ground. *Linguistics Philosophy* 25:701–721
- Strawson P (1950) On referring. *Mind* 59:320–344

Chapter 9

Disaffiliation and Pragmatic Strategies of Emotive Communication in a Multiparty Online Conflict Talk

Laura Bonelli

9.1 Introduction

Conflict talk is about challenges and counter-challenges, defenses, and retreats (Labov and Fanshel 1977: 59). Such moves and their possible impacts on the interactant's stances, on their communicative choices, and even on their relationship statuses are not only, but also determined by considerations on affect (Grimshaw 1990: 12).

The path I am walking in this chapter is an argued attempt of how emotive communication, or the strategic and co-constructed signaling of affective information in conversational interactions, constitutes a prerequisite of more general forms of connection (and disconnection) among people: what Malinowski (1923) referred to as a capability of aggregation of interests and attitudes or, using metaphors of authors who are closer to psychological and linguistic researches, what Watzlawick et al. (1967) called *interpersonal convergence*, what Clark (1996) defines as *joint actions*, and what Caffi (2001) more specifically connects to the ability of empathetic attunement among individuals. In order to achieve a state of *interpersonal convergence*, one has to be able to relate emotionally to her interlocutors, and, in particular, one needs to be attuned to their expressions of affect, both linguistic and paralinguistic. When speakers fail at this, conflictive exchanges are one possible consequence. Although conflicts lead to divergence and disconnection, they occur as joint actions nonetheless: our experience of speakers offers everyday confirmations of this possibility.

A phenomenon which is curious, however, and which became object of study only recently is how computer-mediated communication (henceforth CMC) can

L. Bonelli (✉)
Università "La Sapienza" di Roma, Rome, Italy
e-mail: laura.bonelli@istc.cnr.it

provide similar opportunities of emotive cohesion or lack thereof through, *mutatis mutandis*, macro- and micro-stylistic strategies similar to those occurring in conversational settings.

Exchanges on CMC are a potentially fruitful unit of analysis in the research on emotive communication, since the impressions of interpersonal distance and proximity inferable from their tokens of interaction are so heavy and clear. These impressions contrast with the contextual “coldness” which distinguishes them: as in e-mail, so in online discussion groups (namely, *fora* or the Internet message boards) or on social network platforms (excluding those with integrated instant messaging systems), the interaction often occurs among strangers; the communication is asynchronous, which means that the production of a message and its answers occur at different times; and language is “disembodied” by its producers. Nevertheless, the intensity of the exchanges produced and communicated via these media is often strong enough to make communicative activities such as affiliation and disaffiliation or involvement and detachment particularly evident. At a first glance, it even seems that such manifestations of emotion and positive or negative affectivity are more heavily communicated online than in face-to-face conversations.

If CMC has received special attention since the early 1990s of the last century,¹ recently more attention has been paid to the expressions of emotionality through CMC (to mention a few recent works: Pistolesi 2002; Fabri et al. 2005; Provine et al. 2007; Rodham et al. 2007; Kleinke 2008; Gill et al. 2008; Hancock et al. 2007; Angouri and Tseliga 2010; Thelwall et al. 2010, 2011; Marwick and Boyd 2010; Chmiel et al. 2011; Langlotz and Locher 2012; *inter alia*). It is, however, an amount of researches still quite heterogeneous with regards to investigated objects and platforms, methodologies, and theoretical paradigms of reference.

The analysis I propose is an attempt to employ the resources and analytic tools related to the concept of emotive communication on contextualized tokens of conflict talk in CMC: in particular, I will try to consider how disaffiliation could be detected and measured by means of markers such as Caffi and Janney’s (1994) emotive devices. The methodological and disciplinary framework that I adopt is that of an “integrated” or “holistic” pragmatics (Caffi 2001, 2007), a framework which takes into account strategies and concepts coming from different perspectives (i.e., linguistic pragmatics, social psychology, rhetoric, stylistics, possibly also prosody and nonverbal communication).² The CMC platform I chose for this analysis will be that of Internet message boards. The language of reference is Italian or what Giuseppe Antonelli defined as “digital Italian.”³

¹Two important references with these regards are the volumes of *Language@Internet* and the *Journal of Computer-Mediated Communication*, respectively available online from 2004 to 2012 and from June 1995 to July 2007.

²For an alternative approach to CMC tied to a framework of cognitive pragmatics, see Yus (2013).

³For an overview of features that the Italian language assumes through CMC, see Orletti (2004), Antonelli (2007), Tavosanis (2011), and Fiorentino (*in press*).

But what exactly is referred as emotive communication and what kind of tools are relevant for its investigation? What kind of pertinent concepts does the conversational analytic research offer and how do they relate with the former in the study of conflict talk? Before heading to data and charts, I will briefly try to answer these two questions.

9.2 Emotive Communication: Psychology and Linguistic Pragmatics at Their Interface

At the beginning of the twentieth century, *emotive communication* was broadly defined by Marty (1908: 364) as the strategic and intentional signaling of affective information in speech and writing aimed at influencing the interlocutor's communicative actions, perlocutions, dispositions, stances, and goals. This idea was set by the author against that of *emotional communication*, meaning the spontaneous bursting out of emotion in speech. Leaving behind the discussion on how emotional communication in this sense could also (both intentionally and unintentionally) modify the interlocutor's dispositions and perlocutions (e.g., Haakana 2012), I would like to draw attention on the type of commitments and stances speakers linguistically adopt to influence their interlocutors, either in contexts of negotiation or conflict, and quickly present how they have been treated in the literature.

9.2.1 Linguistic Markers of Psychological Attitudes

Conversations are overflowing with polyfunctional signals or markers (Hölker 1988) which indicate the quality of the self-presentation enacted by the speaker and the quality of her cooperation with her interlocutors on different levels (e.g., prosodic, morpho-syntactic, stylistic, and rhetorical levels). From psychological and sociological points of view, markers might act as cues of extralinguistic behaviors and attitudes: for example, they may index the speaker's belonging to a given social group, specific features of the speaker, or the degree of adherence to an uttered state of affairs and the affective bonds connected with it.

More or less evident and intense tokens of emotive communication are inferable from these cues. The idea of strategic *markedness* of discursive contents and modalities (Hübler 1987) has a long tradition in semiotic studies, as well as in social sciences (see, for instance, Abercrombie 1967). The signaling of speech markers is itself a communicative activity through which the speakers can negotiate needs, request and express information, and regulate personal attitudes (Caffi 2001: 26).

From the point of view of a pragmatics of emotive communication, it is important to identify a comprehensive operational category of markers able to detect and integrate the speaker's attitudes and the modality in which the conversational content is expressed. Tentatives in this direction are, among others, Goodwin et al. (2012), Couper-Kuhlen (2012), Selting (1994, 2010), Caffi (2001), Caffi and Janney

(1994), and Arndt and Janney (1987). Especially Caffi and Janney's emotive devices are a direct attempt of gathering Giles et al.'s (1979) speech markers and Gumperz's (1982) contextualization cues in a unique polyfunctional type of analytic tools.

9.2.2 *Caffi and Janney's Emotive Devices (1994)*

Caffi and Janney's (1994) research effort aims at connecting psychological and linguistic research perspectives to the theme of emotive communication. The authors identified six different *emotive devices* based on the three most recurrent psychological dimensions of affect in the history of psychology—*evaluation*, *potency*, and *activity* (Osgood et al. 1957)—and on the most widespread linguistic categories up to the early 1990s. Rather than focusing solely on the propositional content of the conversational units of analysis (thus investigating emotive communication not exclusively on its semantic and lexical levels), Caffi and Janney (1994: 354) preferred to specify the communicative phenomena that could highlight a certain global affective tonality of the conversation, and they did so by systematically organizing the different types of rhetorical, stylistic, and possibly prosodic and paralinguistic choices that the speakers use in order to strategically produce different evocative effects connected with the kind of emotive stance they display.

The devices they proposed are:

1. *Evaluation devices* (polarity: positive/negative), which include all the verbal and nonverbal choices used to assess the speaking partner or the discursive content and context (e.g., friendly or hostile tones of voice, modal adverbs, adjectives, vocatives, diminutives, lexical, or stylistic choices conveying a positive or a negative attitude). According to the authors, these choices can be interpreted as indexes of pleasure or displeasure, agreement or disagreement, and sympathy or antipathy.
2. *Proximity devices* (polarity: close/far), which include all the verbal and nonverbal choices that can modify the metaphorical distances between the speakers and their conversational topics, between the speakers and the spatial and/or temporal objects belonging to their speaking context, or among the speakers themselves. Proximity is intended as a subjective dimension emotively experienced by the speakers and aimed at the shortening (or at the widening) of their own perceived distances, including the communicative ways of approach or withdrawal toward specific objects of appraisal.
3. *Specificity devices* (polarity: clear/vague), which include all the lexical choices, conversational techniques, and those organizational patterns in the utterance that can express a variation in the level of clarity and accuracy regarding objects and states of affairs, the interlocutor, and the conversation itself. Examples are direct or indirect vocatives, definite articles and pronouns versus indefinites, generic references to the whole versus specific references to parts of a whole (e.g., "Lunch was great"/"The salad dressing was great"), and explicit subjects versus generic subjects (e.g., "I think that"/"One thinks that").

4. *Evidentiality devices* (polarity: confident/doubtful), which include all the linguistic strategies that can regulate the speaker's subscription to the correctness and credibility of what she intends to speak of. From the point of view of an emotive approach to conversation, the most interesting feature of these devices is their ability to convey the speaker's level of confidence or insecurity toward specific topics and interlocutors (1994: 357). Examples are strategic uses of modal verbs (e.g., "It's correct"/"It might be correct"), the degree of explicitness of an intention (e.g., "I'm coming tomorrow"/"I might be coming tomorrow"), other sorts of parentheticals, modal adverbs, hedges (Brown and Levinson 1987; Lakoff 1974), verbal forms of epistemic commitment (Schiffrin 1987; Lyons 1977), verbal forms of self-identification with the conversational topic (Tannen 1989), and more generally all the prosodic and nonverbal choices that can express a major or minor level of intended clearness.
5. *Volitionality devices* (polarity: assertive/nonassertive), which include all the linguistic and conversational strategies that can give the conversational agents an active or a passive role. Examples are, again, strategic uses of modal verbs in requests (e.g., "Would you mind passing the salt?"/"Can you pass the salt"/"Give me the salt") or active versus passive verbal forms in regard to expressing opinions (e.g., "I thought that"/"It was claimed that"). The research on volitionality phenomena is central in studies of Western politeness (inter alia: Brown and Levinson 1987; Blum-Kulka 1987; see Locher and Graham 2010 for a recent overview).
6. *Quantity devices* (polarity: more intense/less intense), which include all the lexical, prosodic, and sometimes kinesic choices aimed at enhancing or reducing the level of conversational intensity (Volek 1987; Labov 1984). Heterogeneous examples are unexpected prosodic stress (e.g., "Don't do that"/"DON'T do that!"), emphatic adjectives (e.g., "It was a good experience"/"It was an awesome experience"), adverbs (e.g., "It was quite/definitely fun"), and various rhetorical strategies of repetition (e.g., "I'm happy, really happy we have met").

The emotive devices of evaluation, specificity, and evidentiality often seem to foreground the speaker-content relationship and to background the speaker-interlocutor relationship, while the devices of volitionality appear to be crucial in the speaker-interlocutor relationship but less important in the speaker-content relationship. When the *focus* of the communicative act is the interlocutor, preferred choices are rhetorical and stylistic strategies aimed at expressing the willingness to maintain the interlocutor's approval, displays of respect (i.e., low levels of assertiveness, recurring positive evaluations, high levels of vagueness, and politely doubtful choices), and face-saving strategies (Brown and Levinson 1987; Goffman 1971, among others). When the *focus* of the communicative act is the speaker herself instead, preferred choices are self-disclosures and choices related to the speaker's own attitudes and desires, primarily marked by devices of evaluation and proximity and enhanced by devices of quantity. Finally, when the *focus* of the communicative act is the conversational content, devices of (2.4) and generally the order in which the elements appear in each utterance are especially central in the expression of relevance and proximity to specific objects and states of affairs.

However, it is important to notice that this kind of perspective, more theoretically than practically clear-cut, may be valid mostly on micro-level units rather than on macro-level units (i.e., conversations, texts, or discourses as a whole), and it may vary depending on different speaking contexts, situations, registers, and cultures. In the extract I analyzed, recurring patterns of devices are organized around hearer-centered utterances, and they mostly present strategies of empathetic proximity and devices of low evidentiality and volitionality when understanding and affiliation is solicited and contrasting sets of devices of low empathetic proximity, negative evaluation, and high intensity in the interlocutor's replies.

9.2.3 *Mitigation: An Umbrella Category of Emotive Communication*

The communicative actions aimed at adjusting at one's interlocutor may also be seen from a perspective of cautious accounting of the risks and responsibilities that conversations generate per se, as well as a way of careful distancing from the possible negative perlocutionary effects that conversations lead to and a manner of protecting the interlocutor or the speaker herself from unwanted interactional outcomes. This form of adaptation is addressed in pragmatics metalanguage with the term *mitigation* (Fraser 1980) and potentially include all the communicative choices aimed at reducing the possible unwanted effects of a given speech act (e.g., indirect acts, justifications, impersonal or passive constructions as a means of non-immediacy, disclaimers, parentheticals, modal adverbs used in order to decrease the emotive subscription to an uttered state of affairs, tag questions, and hedges).

The multidimensionality of mitigation is given by the different resources every speaker has at her disposal in her metapragmatic awareness, resources which can be expressed prosodically (e.g., quieter tone of voice, less emphatic intonations), morpho-syntactically (e.g., impersonal and passive constructions), lexically (e.g., parentheticals, diminutives, modal adverbs aimed at expressing a minor degree of epistemic confidence), and on the conversational level (e.g., topic shifts, digressions). Other examples of mitigating devices are also phatic expressions, vocatives, empathetic datives and honorifics (especially in Asian languages), lexical markers of common ground, fillers, and discursive markers of agreement.

Mitigation is a *nomen actionis*: it can be referred to as the act of mitigating something or as a result of the mitigating process. On the one hand, the former can be seen as part of the speaker's metapragmatic competence where emotive, social, and linguistic abilities converge. On the other hand, the latter can be seen as the object of negotiation among different interlocutors.

In seeing mitigation as a process, Caffi (2001, 2007: 256) distinguishes between different types of mitigation and different types of mitigating devices. Types of mitigations are divided into:

- Mitigation per se: it includes all the communicative actions aimed at “protecting” the interlocutor from negative perlocutionary effects (i.e., altruistic moves) and all the communicative actions aimed at reducing the speaker’s responsibilities (i.e., self-serving moves, see also Fraser 1980).
- Nonnatural mitigation, further divided into two subclasses: *mitigazioni lenitive*, in which the speaker-interlocutor relationship is mostly relevant and which mainly operate on directive speech acts, and *mitigazioni temperatrici*, where the speaker-content relationship is mostly relevant and which mainly operate on representative speech acts.

Types of mitigating devices are divided into:

- *Bushes*: they are aimed at reducing the level of specificity of the propositional content.⁴ Examples are approximators (e.g., “a bit,” “a sort of”), omission signals (e.g., “etc.,” “and so on”), euphemisms and nominal periphrases (e.g., “a bit of x + DIM”, “this and that”), fillers (e.g., “well,” “let’s say”), litotes, and understatements.
- *Hedges*: they are aimed at reducing the degree of evidentiality and assertivity of the illocutionary force. Examples are metapragmatic devices (e.g., “I don’t know . . .”), disclaimers, cautious premises, and markers of the preparatory conditions of the speech act (e.g., “If I understood correctly . . .”), attenuations of the interlocutor’s call to do something or to believe in something (e.g., “maybe,” “a tiny bit”), and modalizers of the epistemic commitment (e.g., “perhaps.” “I’d say,” “probably”).
- *Shields*: they are aimed at reducing or removing one or more aspects of the *instance d’énonciation* (Benveniste 1970). Examples are deictic shields (or “*nonego* devices,” e.g., footing shifts, quotes, impersonal subjects) and spatiotemporal shields (or “*non-hic* devices” and “*non-nunc* devices,” e.g., strategic uses of verbal past tenses and inclusive enallages).

In seeing mitigation as a product or as an effect, Caffi (2001: 452) presents a series of conversational macro-strategies such as semantic strategies (e.g., eventualization), metacommunicative strategies (e.g., fictionalization), sequential strategies (e.g., strategic turn-taking and strategic topic shifts and changes), and co-constructional strategies (e.g., stylistic actions on the speaking register aimed at increasing or decreasing the level of shared intimacy).

A very interesting type of co-constructional strategy of mitigation is empathetic attunement, which is defined by Caffi (2001: 218) as an operation of cognitive and emotive coordination enacted by the speakers of how they perceive each other and of what their interactional goals are. By attuning with each other, the communicative actors mutually verify the interpretation they should give to their exchange (i.e., cognitive operations) and mutually reduce their perceived distances (i.e., emotive operations). The author develops this concept from Stern (1985) and communication

⁴Lakoff (1974) included this type of devices among hedges.

accommodation theory, in particular Giles et al. (1991). Two kinds of attunement are hypothesized: “thematic attunement,” a strategy wherewith the speaker helps understanding her point (e.g., with reformulations), and “stylistic attunement,” a set of convergent strategies both on the topic and on the formal aspects of the conversation the speakers adopt in order to attempt a mutual approach to each other (e.g., by decreasing the level of formality and indirectness). The device of empathetic distance/proximity is proposed by Caffi as a manner of identification of the linguistic markedness of these two types of attunement.

However, why are mitigating devices and empathetic attunement relevant in the study of conflict talk? As I will show in the analysis that follows, sensitive issues are often presented with numerous kinds of mitigating strategies in order to avoid negative perlocutionary effects. On the contrary, the expression of contrasting stances can strategically present aggravating strategies (Merlini Barbaresi 2009) and generally emotive strategies opposite to mitigation as a manner of reinforcement of the status of distance and disagreement.

9.3 Key Concepts from the Conversational Analytic Framework

9.3.1 *The Idea of Emotive Stance (Ochs 1986)*

Emotive or affective stance has different definitions throughout linguistic literature. Ochs (1986: 410) defines it as “a mood, attitude, feeling and disposition, as well as degrees of emotional intensity *vis-à-vis* some focus of concern.” More recently, Du Bois (2007: 169) generally defines stance taking as “a public act by a social actor, achieved dialogically through overt communicative means (language, gesture and other symbolic forms), through which social actors simultaneously evaluate objects, position subjects (themselves and others) and align with other subjects, with respect to any salient dimension of the socio-cultural field.” Stivers (2008) uses the term stance to describe the affective treatment by a given speaker of the events she is speaking of.

Some authors postulate more forms of stance, of which the affective-emotive is one of the possible types. Ochs (1986) distinguishes between affective and epistemic stance, and she highlights the indexical nature of each of them. Goodwin (2007) distinguishes between five different types of stance: instrumental, cooperative, epistemic, moral, and affective. He imagines all of these stances manifested through verbal and mostly nonverbal strategies and devices, such as intonation, body posture, prosody, and gesture.

Other researches on the expressive modalities of emotive or affective stance are found, among others, in Goodwin et al. (2012), Niemelä (2010), Jaffe (2009), Englebretson (2007), and Kärkkäinen (2003, 2006). More recently, stance styles have begun to be regarded as intersubjective phenomena (Kärkkäinen 2003),

responsive to the interactional requirements and contexts in which the speakers interact. The *focus* has hence moved from the individual speaker toward a more co-constructive approach, the same approach my analysis aims at fitting into.

9.3.2 *The Concept of Affiliation (Stivers 2008)*

Affiliation is understood as that series of sequential actions in the context of a communicative exchange aimed at supporting or approving the speaker's emotive stance, this last being made explicit by the speaker herself in her conversational turns (Couper-Kuhlen 2012: 113).⁵

In his research on conversational storytelling, Stivers (2008) distinguished between two different types of reception adapted to the interlocutor: alignment and affiliation. Whereas the former indicates all the communicative tokens linked with the interlocutor's role (e.g., proper turn-taking or feedbacks on the understanding, like "mmh mmh," "a-ha," "yes"), the latter is the explicit endorsement of the speaker's affective orientation made evident by means of assessments congruous with those expressed by the speaker herself. Contrasting short replies, withholdings, and follow-up questions which appear in the conversational segment that follows the speaker's explanation or presentation of her emotive stance and generally all the communicative tokens which do not endorse the speaker's emotive stance are considered non-affiliative instead, together with those communicative tokens which are openly discordant with the speaker's affective stance (and that are thus based on a different and contrasting stance).

Even though affiliation is considered a preferred action in the communicative exchanges, responses are never intrinsically affiliative, but they become such depending on their position in the dialogue: for example, nodding is understood as a type of affiliative response if it occurs right after the speaker's presentation of her emotive stance, but it is viewed as non-affiliative if it occurs at the end of the speaker's storytelling sequence (Couper-Kuhlen 2012; Stivers 2008).

Lindström and Sorjonen (2012) consider complaint stories and trouble talks the conversational contexts where affiliative replies are more often preferred and exhibited (see also Selting 2010—*inter alia*). The relevance of affiliation as a fruitful practice in therapeutic contexts has been especially underlined by Ruusuvuori (2005, 2007, 2013).

Affiliative and disaffiliative types of responses constitute a resourceful aspect of emotive communication, though their modalities of presentation in the communicative exchanges are yet in need of further exploration. As I will show in the following paragraphs, the degree of convergence of affiliation elicitation and of

⁵Affiliation in this sense is a conversational category similar (but not isomorphic) to the psychological-affective dimension of affiliation mentioned by authors such as Gough (1957) or Russell (1991).

disaffiliative replies with the possible emotive stances presented by the interactants in a conflict talk may be measurable with the different types of emotive devices that those interactants employ.

9.4 Data

The *corpus* I examined is a thread of 155 posts in Italian from the generalist *forum Postare.it*, which is divided into 38 boards featuring topics ranging from health to comics, from cooking to philosophy. It is a widely attended message board by Italian users remarkably heterogeneous in age, sex, and gender.

The use of nicknames by the users—or, at all events, the impossibility to recognize their true identity—guarantees the privacy of the people who participated to these discussions behind the screens of their computers. The threads I selected for this analysis are publicly visible on the website without any need of registration to the message board. The interactions were faithfully reported, including their typing, orthographic, and grammar mistakes. The transcription system I used follows the Internet message boards standards used by Langlotz (2010).

9.5 Methods and Objectives

Central in my analysis is above all the concept of emotive communication I previously presented: the inner state of the interactants, particularly inaccessible also due to the kind of exchange they are protagonists of, was not considered. I instead focused on the various effects of *approach* and *withdrawal* (Frijda 1998) and of interest and disinterest inferable from the interactant's communicative choices. The persuasive and strategic importance of these choices in the acts of co-construction of meaning will also be shown, as well as the implications of those choices for the management of the communicative exchange.

The presentation of the various strategies adopted by the users and their articulation in a wide range of dimensions has been considered sufficient thus far in order to explore the conflictive emotions in this kind of interaction. While detecting the linguistic markers in the text, the following dimensions have been taken into account: linguistic, in its pragmatics, semantic, syntactic, stylistic, and rhetorical aspects; discursive, in particular dealing with metacommunicative, contextual, and co-textual strategies; psychological, evaluated mainly by means of the emotive devices by Caffi and Janney (1994) and by means of the markers of empathetic proximity and mitigation by Caffi (2001, 2007) and boosters and markers of linguistic aggravation (Merlini Barbaresi 2009); and sequential strategies and strategies of presentation and reception of the affective message in their different phases, evaluated mainly thanks to conversational analytic concepts (i.e.,

disaffiliative replies, possible misalignments, sequences of introduction and exit from complaint stories, assessments. See Table 9.1).

Mutatis mutandis, perspectives and analytical tools typical of oral dialogic exchanges were employed. I made particular use of the integrated pragmatic methodologies used by Caffi (2001, 2007) and of those used to analyze affectivity in conversational storytelling by Selting (2010) and Couper-Kuhlen (2012). The type of methodology I adopted, however, differs from those generally employed in the analysis of oral conversations at least with respect to the following aspects:

- The interpretation of proxemic and kinesic resources was replaced, where possible, by an interpretation of alternative visual means, in particular the contextualized use of emoticons.
- Interpretations of prosodic, phonological, and tonal aspects of communicative exchanges were omitted, due to the obvious limitations of the object of analysis.
- Sequential and turn-taking aspects were detected in a simplified and reduced manner. Namely, (1) I consider each user's post as a complex turn composed by several turn constructional units (see Selting 2000). (2) The division into different lines in the transcripts does not correspond to different turns, but to single turn constructional units, and it faithfully reproduces the number of times each user started a new paragraph in the post. (3) Because of the structural difference of the tokens I analyzed with regard to face-to-face synchronous interactions, signals of interruption and overlap are absent, whereas those of alignment and disalignment are only shown when clearly evident in the turns.

The following analysis has, moreover, a triple objective:

- The proposal of a macro-connection of the methodologies of integrated pragmatic analysis (Caffi 2001, 2007) with those belonging to the research on affectivity in conversational storytelling (e.g., Selting 2010; Couper-Kuhlen 2012)
- The proposal of a micro-connection of linguistic, visual, psychological, and metacommunicative tools of analysis and some hypotheses about their possible correlations
- The presentation of possible starting points for potential future research on the management of conflict talk as well as the co-construction of affectivity in digital communicative interactions

9.6 Analysis of a Thread

The reflections I present are based on a fervent cross talk among a guy whose nickname is "Calcolatore83" and twenty-nine other members of the message board *Postare.it*. Some of them provided contribution to the discussion as real counterparts, while others appeared only as cynical and detached commenters. This thread is placed in the board *Relazioni e sentimenti* (in English: "Relationships and

Table 9.1 A caption of the word-by-word analysis of the conflict talk (from line 1a to 1c) with English translations

1. Line no.	2. Type of outcome (contextual and co-textual description)	3. Syntactic, morphological, and lexical devices	4. Type of illocutionary acts	5. Rhetorical and stylistic devices	6. Metacom-municative and metadis-cursive devices	7. Nonverbal devices	8. Mitigating devices and markers of reinforcement and aggravation	9. Emotive devices and inferable empathetic distances	10. Sequential aspects: phases of reception of the affective message in the storytelling
1a	Calcolatore83 opens the thread and greets the users of the <i>forum</i>	Expressive acts		<i>Capitatio benevolentiae</i>				+Proximity (social): <i>amici</i> , i.e., “friends” –Specificity: <i>una dozzina di giorni</i> , i.e., “a dozen days”	
1b	Calcolatore83 warns the users of the forum and the prospective readers of his thread	Expressive and directive acts		Repetition with empathic purposes, e.g., <i>dall’inizio... inizio</i> (i.e., “since the beginning” “beginning”)	Metadiscursive resources: Calcolatore83 invites the readers not to read his thread if not interested			+Volitionality: <i>voglio</i> i.e., <i>I want</i> , +evidentiality <i>cominto che</i> , i.e., “confident that” –Specificity <i>della vicenda</i> i.e., “of the events”	

1c	Calcolatore83 invites the users to check his old thread regarding this same topic for those who are not yet acquainted with his problem	Representative acts		Co-textual anchor: <i>c'è l'altro post</i> (i.e., "there's the other thread," meaning the readers are kindly invited to check his previous discussion to catch up with the topic)	Reinforcing device: <i>tutta la storia</i> (i.e., "the whole story")	
----	---	---------------------	--	---	--	--

The second column on the left describes the type of outcome carried out in each post of the thread and possibly the topic of reference. Columns from 3 to 7 include a presentation of the various linguistic means employed in each post, as well as the possible visual and metacommunicative means. Columns 8–9 present the occurrences of the six emotive devices by Caffi and Janney (1994) as well as the occurring markers of reinforcement and mitigation and those of empathetic proximity (Caffi 2001, 2007; Merlini Barbarese 2009). Finally, the tenth column retraces the steps of presentation and reception of the user's affective messages by means of conversational analytic definitions

feelings”) and counts over 6,830 views. It was opened on January 9, 2012, and it was closed by the moderator with the nickname of “Pannocchia” on February 23, 2012. The title chosen by Calcolatore83 for this thread is *Dopo 10 anni torna l'ex (Parte 2)* (i.e., “After 10 years the ex comes back (Part 2)”): after an initial, first outburst presented in a previous thread, the user updates the *forum* members about his personal life events—a girl he had been in a relationship with 10 years before has come back to his life. This fact represents for him a cause of curiosity and attraction but also a potential interference to his current relationship. Due to this last aspect, the user appears into a state of confusion and, at times, sense of guilt. He self-discloses several times and often seeks for the *forum* members’ understanding and empathy.

Probably encouraged by the partial interest and support he had obtained in the previous discussion, Calcolatore83 opens a new thread to inform about the latest developments of his personal long-standing problem, but this time he mainly receives annoyed, angry, incredulous, sarcastic, and noncooperative replies, very few expressions of empathy and even some insults by the interacting users, visibly irritated by his continuous mood swings and indecisions.

The analysis was divided into the following sections:

- A section based on thematic macro-levels, which is useful to present the main stages of the polylogue.
- A section based on analytical micro-levels: an extract of the discussion was selected, and its main linguistic and emotive means have been identified.
- A section of comparative analysis where the detected emotive devices and the disaffiliative replies were matched.

9.6.1 *Macroanalysis and Plot of the Interaction*

The conflict talk is organized as follows:

- *Phase 1*: opening (lines 1a–3b). Calcolatore83 opens the thread after a brief introduction in which he refers to his previous discussion about the same subject. He also invites the users to avoid “unnecessary offences.”
- *Phase 2*: *Continuo a ripetermi che dovrei troncare, ma . . .* (i.e., “I keep on telling myself I should break up, but . . .” lines 4a–7c). Calcolatore83 reinforces his emotive stance after receiving the first cold answers from the board’s members.⁶ After some premises aimed at getting him empathetic listening, he tells another part of his story, and he explains the emotional involvement this caused him. There are different reactions to this post: an affiliative one by Caracas (who contributes in creating a cooperative behavior although the user doesn’t appear to be set in Calcolatore83’s emotive stance) and a non-affiliative reply by Opunzia,

⁶This is a typical behavior in trouble talk, see, for example, Peräkylä and Sorjonen (2012).

who suggests Calcolatore83 to interrupt his current relationship. Calcolatore83 thanks Caracas and represents his conflicting and confused emotive stance.

- *Phase 3: Ma che abbiamo fatto di male noi?* (i.e., “But what did we do wrong?” lines 8a–25b). The disaffiliative replies increase, with the exception of Sendoh’s answer which is characterized by some sort of cooperative sarcasm. The users now directly attack and criticize Calcolatore83, and they sometimes show their irritation through sidetalks. After these reactions, Calcolatore83 highlights once more his sense of guilt and his confusion in order to obtain the users’ attention.
- *Phase 4: Una persona così autolesionista e autocommiserante...* (i.e., “Such a self-defeating and self-commiserating person...” lines 26a–143e). Calcolatore83 makes numerous attempts to obtain the users’ support and suggestions, now using persuasive techniques and then updating the telling of his personal affairs. He fails many times to receive any attention, and his posts get rare affiliative replies, many disaffiliative and evidently annoyed responses, some attacks, and several teasings.
- *Phase 5: È il caso di chiudere questo post!* (i.e., “This thread should be closed!” lines 144a–155b). Calcolatore83 gives up his search for empathetic listening and support, and he states with a certain resentment that he wants to end the discussion. This decision is greeted with sarcastic relief by many users. Moderator Pannocchia concludes the thread allowing Calcolatore83 to open another one on the same issues yet remarking that the possibility of being criticized is always present in Internet message boards.

The listed section considers the interpersonal orientation followed by most of the participants to the discussion. In Table 9.2, a caption of the interaction mainly containing disaffiliative and sarcastic replies is shown.⁷


9.6.2 *Distribution and Analysis of the Emotive Devices in Relation to the Users’ Disaffiliative Replies*

As it is shown in Fig. 9.1, the lack of affiliation in the interlocutor’s responses displays a concentration of emotive devices characterized by a trend which strongly contrasts the emotive stance of Calcolatore83. The internal homogeneity of these replies and the replies discrepancy with the evocative effects of Calcolatore83’s outcomes suggest that the users adopted and shared a different emotive stance, opposite to that of Calcolatore83.

Calcolatore83 is insecure, confused, and constantly torn between his attraction to his ex-girlfriend and the sense of guilt against his current relationship. He seems also split between the urge to express his consciously unethical fantasies and the desire

⁷The complete discussion (in Italian) is publicly available at <http://www.postare.it/showthread.php?s=1d08e3acb2813a2d469591bb6292af90&threadid=260285&perpage=10&pagenumber=1>.

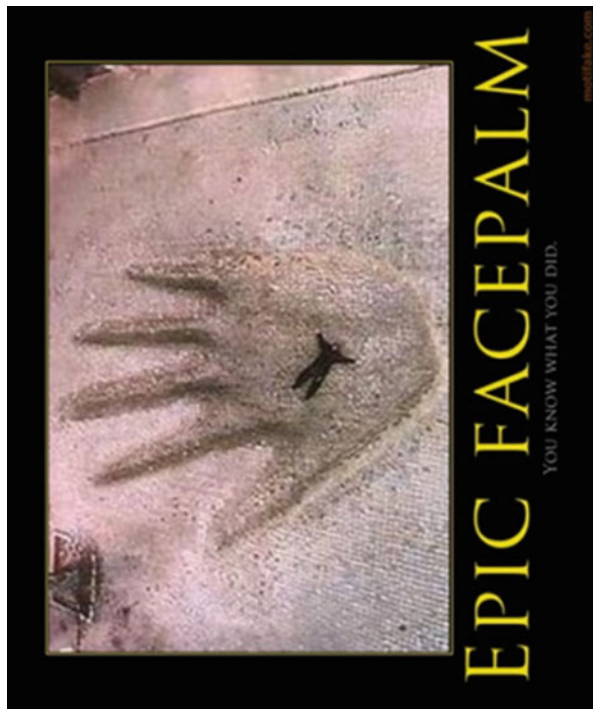
Table 9.2 A caption of the analyzed thread (from line 8a to 14b) with English translations

8a	<i>Pannocchia</i>
8b	<i>Ripetiamo ... con la tua fidanzata non può andare tutto benissimo. Tu non comprendi questo</i>
8c	We repeat ... not everything can't be perfect with your girlfriend. That's why you don't understand <i>Se andasse tutto veramente come dici tu, non ti troveresti in questa situazione</i> If everything went the way you say, you wouldn't find yourself in this situation
	<i>Quindi ... fermati un attimo e cerca di capire dove il tuo rapporto fa cilecca. Trova il problema e 1) lo risolvi 2) la lasci</i> So ... hang on for a while and try to understand where you're failing at your relationship. Find the problem and (1) solve it (2) you break up with her
9a	<i>Opunzia</i>
9b	... sì, però tu continui a pensare all'altra  ... yeah, but you keep thinking of the other one <i>edit: concordo in pieno con pannocchia</i> edit: I totally agree with pannocchia
11a	<i>Tennessee</i> <i>aspetto il :megasurfacepalm: dello Zio</i> I'm waiting for Zio's hyper facepalm
12a	<i>Mirrilla</i>
12b	<i>Qual è la domanda?</i> What's your question? <i>Cioè precisamente cosa vuoi sapere da noi?</i> That's to say, what do you want to know from us exactly?

13a Zio

Quote: *aspetto il :megasuperfacepalm: dello Zio*

13b



 *ti basta?*
enough?

(continued)

Table 9.2 (continued)

14a	§Euriclea§ <i>Io ho sempre seguito il post precedente, ma non sono mai intervenuta ma ora che (non so con quale coraggio) lo hai riaperto te lo devo dire</i>
14b	<i>Io ti avrei dato un calcio tra le gambe .. del tipo un centesimo per ogni tuo pensiero... ecco...stesso sistema ma a suon di calci</i> I have followed the previous discussion, but I have never intervened But now that you re-opened it (I don't know with what sort of courage) I must say this to you I would have kicked you between your legs ... like, a penny for each of your thoughts ... there ... same system but translated with kicks <i>hai scritto tante di quelle volte che nonostante tutto con la tua ragazza va tutto benissimo che probabilmente con tutti i calci le pa*le sarebbero marcite e ti sarebbe passata la smania sessuale nei confronti dell'altra</i> You wrote so many times that despite everything it is going great with your girlfriend that probably with all my kicks your balls would have rotten and you would have recovered from your sexual frenzy for the other one

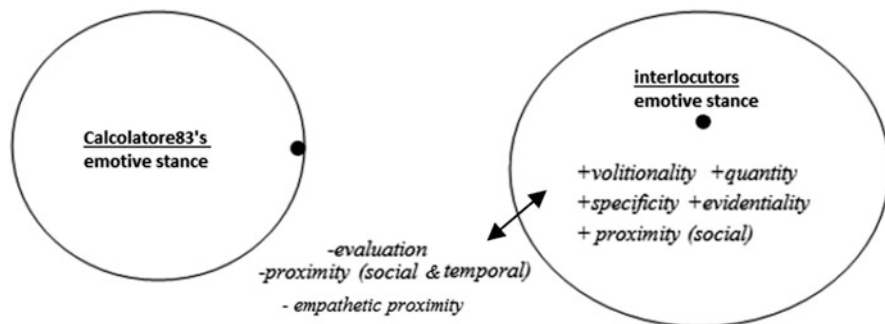


Fig. 9.1 Distribution of the emotive devices in the disaffiliative replies of Calcolatore83's interlocutors. The devices of negative evaluation and of negative social, temporal, and empathetic proximity are referred to as Calcolatore83's stance. The devices of positive volitionality, quantity, specificity, evidentiality, and proximity are used as a manner of reinforcement of the interlocutor's contrasting stance

of a *captatio benevolentiae* from his interlocutors. The user is often unassertive and uncertain and floats between positive evaluations and demonstrations of affective proximity toward his love story (examples are phrases such as *ci troviamo a meraviglia*, in English "we are doing awesome"; *va tutto benissimo*, in English "everything is going great") and evaluations and displays of proximity now positive and then negative toward the protagonist of his flirt (e.g., *poverina*, in English "poor little thing"; *le ho risposto subito*, i.e., "I replied immediately"; contrasting with phrases such as *relazione clandestina*, in English "a clandestine relationship"; *quest'altra storia*, i.e., "this other story").

Calcolatore83's interlocutors, often annoyed and bored by his indecisions, show their lack of affiliation with two different behaviors: on the one hand, they devalue and detach from the content of Calcolatore83's emotive stance using devices of negative evaluation and distance (e.g., *mania sessuale*, i.e., "sexual frenzy", or empathetic deixis such as *dell'altra*, i.e., "of the other one," meaning the other girl). On the other hand, they constitute and strengthen their own emotive stance, in such a way that it appears internally coherent in its manners and sometimes internally cohesive in its contents. The interlocutor's stance is mainly reinforced by emotive devices of social proximity (e.g., *noi*, in English "us"), expressions of epistemic certainty (e.g., *evidente*, *semplicemente*, i.e., "obvious," "simply"), signs of assertiveness (e.g., *te lo devo dire*, i.e., "I have to tell you"), and displays of intensity such as the use of exclamations, dysphemisms, hyperbolic images, and emoticons with emphatic expressions.

The interlocutor's lack of affiliation with the contents of Calcolatore83's affectively strong messages involves strategic and coherent choices. These choices are shown presumably after the recognition of Calcolatore83's emotive stance, a type of behavior partly predicted by Caffi and Janney (1994).

On the one hand, it is possible to identify in Calcolatore83's hearer-centered posts a prevalence of emotive devices of request of agreement and

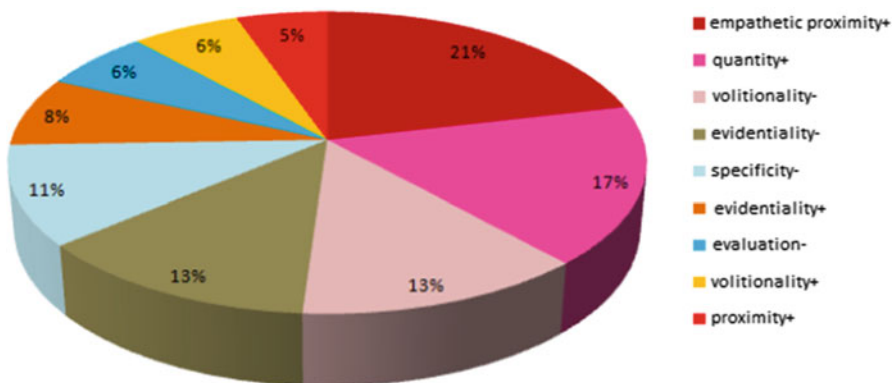


Fig. 9.2 Distribution of the emotive devices in Calcaltore83's hearer-centered posts. Prevailing are the devices of empathetic proximity, quantity, and low evidentiality and volitionality, as indexes of the user affiliation elicitation

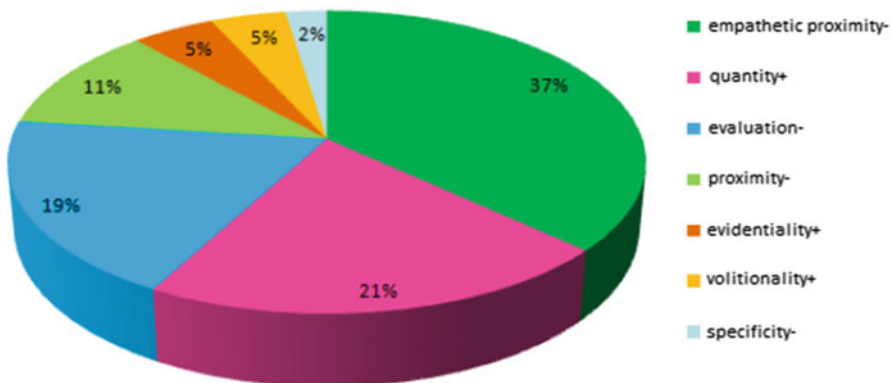


Fig. 9.3 Distribution of the emotive devices in the disaffiliative replies that Calcaltore83 receives. The users clearly and intensely distance themselves from Calcaltore83's stance and requests of affiliation: devices of empathetic distance, quantity, and negative evaluation are prevalent

approach elicitation *à la* Frijda (1998) (26 % of the total devices, divided into *+empathetic closeness*, 21 %, and *+proximity*, 5 %), juxtaposed by emotive devices of uncertainty and low assertiveness (26 % of the total, divided into *-evidentiality*, 13 %, and *-volitionality*, 13 %), and followed by devices of intensity (*+quantity*, 17 %) and devices of vagueness (*-specificity*, 11 %), as it is shown in Fig. 9.2.

On the other hand, in the user's disaffiliative replies, devices of strong emotive distance from Calcaltore83 are prevalent (*-empathetic closeness*, 37 %, and *-proximity*, 11 %, that being an amount of 48 %, almost half of the total devices), followed by devices of intensity (*+quantity*, 21 %) that can be interpreted as a sign of strong intemperance and *aggravation* (Merlini Barbaresi 2009) and devices of negative evaluation (whose assessed object is, clearly, Calcaltore83 himself: *-evaluation*, 19 %), as it is shown in Fig. 9.3.

The coloring criteria used for the charts in Figs. 9.2 and 9.3 are the same: the two main contrasting stances are also evident from the discrepancy both in quantity and in quality of the emotive devices and markers respectively employed.

9.7 Conclusions

To give interpretations often remain conjectural, probabilistic, abductive, and of exploratory nature, in particular for the humble and circumscribed goal and the sake of brevity of this work. Although it is not possible to generalize, I briefly resubmit below the results I obtained and thus try to give some concluding remarks.

In the online conflict talk I took into account, the user *Calcolatore83*'s emotive stance (confused, repetitive, and ambiguous) obtains a second contrasting stance from his interlocutors in the message board, who act as completely detached from his concerns. *Calcolatore83*'s *emotive stance* seems built through illocutionary, structural, semantic, sequential, stylistic, and rhetorical means characterized by insecurity, a low level of assertiveness, and contradictory evaluations, while the *emotive stance* of his interlocutors seem to display an opposite opinion, characterized by epistemic certainty, assertiveness, and a considerable distance from the semantic contents of *Calcolatore83*'s disclosures. This divergence is evident in the distribution of the emotive devices present in the interlocutor's outcomes, very often characterized by devices presenting negative polarities opposite to those emerging from *Calcolatore83*'s posts (e.g., devices of negative evaluation and distance, whereas *Calcolatore83* expresses, despite his insecurity, positive evaluations of his own story and displays of proximity). The user's outcomes also present devices with positive polarities (in particular, devices of assertiveness, evidentiality, and quantity), these last representing indexes of a second, divergent, and internally cohesive emotive stance: this aspect of internal cohesion is inferable, for example, from the markers of social proximity referred to as the interlocutors themselves.

While waiting for further and more deepened research results, the resource of the emotive devices by Caffi and Janney (1994), possibly joined with the analytical tools of integrated pragmatics (Caffi 2001, 2007) and to those of conversation analysis relating to affectivity in storytelling (e.g., Selting 2010; Fox et al. 2013; Peräkylä and Ruusuvuori 2013; inter alia), is a potentially fruitful heuristic for the prospects of research on emotive communication and also a potential way to connect different methodologies on communicative research all together.

Nevertheless, numerous explanations and more insights on the theoretical front are yet much needed: for example, it would be relevant to see if speakers use boosters and reinforcing devices opposite to the three more broadly known types of mitigators (i.e., bushes, hedges, and shields) and in what interactional contexts or for what purposes they are mainly employed. Also, the connection between the idea of emotive stance and that of the emotive devices should be further clarified. The objects, types, and objectives of the emotive devices should be more specifically investigated, too. Moreover, the analysis of the emotive activities of co-orientation in talk-in-interaction should be deepened by means of more interdisciplinary research, for instance, by seeing if the displays of affiliation and

disaffiliation can be considered the perlocutionary counterparts of the strategies of empathetic attunement. Finally, differences and analogies between CMC exchanges and face-to-face interactions with regard to emotions and affectivity should be further explored.

References

- Abercrombie D (1967) *Elements of general phonetics*. Edinburgh University Press, Edinburgh
- Angouri J, Tseliga T (2010) “You Have No Idea What You are Talking About!” From e-disagreement to e-impoliteness in two online fora. *J Politeness Res* 6:57–82
- Antonelli G (2007) *L’italiano nella società della comunicazione*. Il Mulino, Bologna
- Arndt H, Janney RW (1987) *InterGrammar. Toward an integrative model of verbal, prosodic and kinesic choices in speech*. Mouton de Gruyter, Berlin
- Benveniste E (1970) *L’appareil formel de l’énonciation*. *Langages*, vol 17. Rist. In: Id., *Problèmes de linguistique générale*, II, 1974, pp 79–88. *Problemi di linguistica generale II*, Il Saggiatore, Milano, 1985, pp 96–106
- Blum-Kulka S (1987) Indirectness and politeness in requests: Same or different? *J Pragmatics* 11:147–160
- Brown P, Levinson S [1978] (1987) *Politeness. Some universals in language usage*. Cambridge University Press, Cambridge
- Caffi C (2007) *On mitigation*. Emerald, Bingley
- Caffi C (2001) *La Mitigazione. Un approccio pragmatico alla comunicazione nei contesti terapeutici*. LIT, Münster, Hamburg, London
- Caffi C, Janney RW (1994) Toward a pragmatics of emotive communication. *J Pragmatics* 22: 325–373
- Chmiel A, Sienkiewicz J, Thelwall M, Paltoglou G, Buckley K (2011) Collective emotions online and their influence on community life. *PLoS One* 6(7):e22207
- Clark HH (1996) *Using language*. Cambridge University Press, Cambridge
- Couper-Kuhlen E (2012) In: Peräkylä A, Sorjonen ML (eds) *Exploring affiliation in the reception of conversational complaint stories*. pp 113–146
- Du Bois JW (2007) In: Englebretson R (ed) *The stance triangle*. pp 140–182
- Englebretson R (ed) (2007) *Stancetaking in discourse: subjectivity, evaluation, interaction*. John Benjamins, Amsterdam
- Fabri M, Moore DJ, Hobbs DJ (2005) Empathy and enjoyment in instant messaging. In: *Proceedings of 19th British HCI group annual conference (HCI2005)*, Edinburgh, UK, pp 4–9
- Fiorentino G (in press) *Forme di scrittura in rete: dal web 1.0 al web 2.0*
- Fox B, Thompson S, Ford C, Couper-Kuhlen E (2013) *Handbook of conversation analysis* (Stivers T, Sidnell J (eds.)). Wiley-Blackwell, New York
- Fraser B (1980) Conversational mitigation. *J Pragmatics* 4:341–350
- Frijda NH (1998) The laws of emotion. In: Jenkins JM, Oatley K, Stein N (eds) *Human emotions: a reader* (Gumperz JJ, Levinson SC (eds.)). Blackwell Publishers, Malden, MA, pp 271–287
- Giles H, Coupland J, Coupland N (eds) (1991) *Contexts of accommodation: developments in applied sociolinguistics*. Cambridge University Press, Cambridge
- Giles H, Scherer KR, Talor DM (1979) In: Scherer K, Giles H (eds) *Speech markers in social interaction*. pp 343–381
- Gill AJ, Gergle D, French RM, Oberlander J (2008) Emotion rating from short blog texts. *Proceedings of the ACM conference on human factors in computing systems (CHI 2008)*, pp 1121–1124
- Goodwin C (2007) Participation, stance, and affect in the organization of activities. *Discourse Soc* 18:53–73

- Goodwin MH, Cekaite A, Goodwin C (2012) In: Peräkylä A, Sorjonen ML (eds) *Emotion as stance*. pp 16–41
- Goffman E (1971) *The territories of the self. Relations in public*, pp 28–61
- Gough HH (ed) (1957) *Manual for the California psychological inventory*. Consulting Psychologist Press, Palo Alto, CA
- Grimshaw AD (ed) (1990) *Conflict talk. Sociolinguistic investigations of arguments in conversations*. Cambridge University Press, Cambridge
- Gumperz JJ (1982) *Discourse strategies*. Cambridge University Press, Cambridge
- Haakana M (2012) In: Peräkylä A, Sorjonen ML (eds) *Laughter in conversation: the case of “Fake Laughter.”* pp 174–194
- Hancock JT, Landrigan C, Silver C (2007) *Expressing emotion in text-based communication. Proceedings of ACM conference on human factors in computing systems (CHI 2007)*. pp 929–932
- Hölker K (1988) *Zur Analyse von Markern*. Franz Steiner, Stuttgart
- Hübler A (1987) In: Dirven R, Fried V (eds) *Communication and expressivity*. pp 357–380
- Jaffe AM (ed) (2009) *Stance: sociolinguistic perspectives*. Oxford University Press, Oxford
- Kärkkäinen E (2006) *Stance taking in conversation: from subjectivity to intersubjectivity*. *Text Talk* 26:699–731
- Kärkkäinen E (ed) (2003) *Epistemic stance in english conversation: a description of its interactional functions, with a focus on “I think”*. John Benjamins, Amsterdam
- Kleinke S (2008) *Emotional commitment in public political internet message boards*. *J Lang Soc Psychol* 27:409–421
- Labov W (1984) In: Schiffrin D (ed) *Intensity*. pp 43–70
- Labov W, Fanshel D (1977) *Therapeutic discourse: psychotherapy as conversation*. Academic, New York
- Lakoff R (1974) *Remarks on this and that in papers from the tenth regional meeting of the Chicago Linguistic Society*. University of Chicago Press, Chicago, pp 345–356
- Langlotz A (2010) In: Locher MA, Graham SL (eds) *Social cognition*. pp 167–202
- Langlotz A, Locher MA (2012) *Ways of communicating emotional stance in online disagreements*. *J Pragmatics* 44:1591–1606
- Lindström A, Sorjonen ML (2012) In: Sidnell J, Stivers T (eds) *Affiliation in conversation*. pp 250–369
- Locher M, Graham SL (2010) *Handbook of pragmatics, vol Interpersonal pragmatics*. De Gruyter Mouton, Berlin, NY
- Lyons J (1977) *Semantics, vol 2*. Cambridge University Press, Cambridge
- Malinowski B (1923) In: Ogden K, Richards IA (eds) *The problem of meaning in primitive languages*. pp 296–336
- Merlini Barbaresi L (2009) In: Gili Fivela B, Bazzanella C (eds) *Linguaggio intemperante e linguaggio temperato. Ovvero intensificazione arrogante e attenuazione cortese*. pp 59–78
- Marty A (1908) *Untersuchungen zur Grundlegung der allgemeinen Grammatik und Sprachphilosophie*. Halle an der Salle, Niemeyer
- Marwick AE, Boyd D (2011) *I tweet honestly, I tweet passionately: Twitter users, context collapse, and the imagined audience*. *New Media Soc* 13:1–20
- Niemelä M (2010) *The reporting space in conversational storytelling: orchestrating all semiotic channels for taking a stance*. *J Pragmatics* 42:3258–3270
- Och E (1996) *Linguistic resources for socializing humanity. Rethinking linguistic relativity*. University Press, Cambridge, pp 407–437
- Orletti F (ed) (2004) *Scrittura e nuovi media. Dalle conversazioni in rete alla Web usability*. Carocci, Roma
- Osgood CE, Suci GJ, Tannenbaum PH (1957) *The measurement of meaning*. University of Illinois Press, Urbana, Chicago, London
- Peräkylä A, Ruusuvuori J (2006) *Facial expression in an assessment*. In: Knoblauch HJ, Raab J, Soeffne H-G, Schnettler B (eds) *Video-analysis: methodology and methods. Qualitative audiovisual data analysis in sociology*. Peter Lang, Frankfurt

- Peräkylä A, Sorjonen ML (2012) *Emotion in interaction*. Oxford University Press, Oxford
- Pistolessi E (2002) Flame e coinvolgimento in IRC (Internet relay chat). In: Bazzanella C, Kobau P (eds) *Passioni, emozioni, affetti*. McGraw-Hill, Milano, pp 261–277
- Provine RR, Spencer RJ, Mandell DL (2007) Emotional expression online emoticons punctuate website text messages. *J Lang Soc Psychol* 26:299–307
- Rodham K, Gavin J, Miles M (2007) I hear, I listen and I care: a qualitative investigation into the function of a self-harm message board. *Suicide Life Threat Behav* 37:422–430
- Russell JA (1991) Culture and the categorization of emotions. *Psychol Bull* 110:426–450
- Ruusuvuori J (2013) In: Sidnell J, Stivers T (eds) *Emotion, affect and conversation*. pp 330–349
- Ruusuvuori J (2007) Managing affect: integrating empathy and problem solving in two types of health care consultation. *Discourse Stud* 9:597–622
- Ruusuvuori J (2005) “Empathy” and “sympathy” in action: attending to patients’ troubles in Finnish homeopathic and general practice consultations. *Soc Psychol Q* 68:204–222
- Schiffrin D (1987) *Discourse markers*. Cambridge University Press, Cambridge
- Selting M (2010) Affectivity in conversational storytelling. *Pragmatics* 20:229–277
- Selting M (2000) The construction of units in conversational talk. *Lang Soc* 29:477–517
- Selting M (1994) Emphatic speech style: with special focus on the prosodic signalling of heightened emotive involvement in conversation. *J Pragmatics* 22:375–408
- Stern DN [1985] (1987) The interpersonal world of the infant. In: Ammaniti M (ed) *Il mondo interpersonale del bambino*. Basic Books, New York, Bollati Boringhieri, Torino
- Stivers T (2008) Stance, alignment and affiliation during story telling: when nodding is a token of preliminary affiliation. *Res Lang Soc Interact* 41:29–55
- Tannen D [1989] (2007) *Talking voices. Repetition, dialogue, and imagery*. In: *Conversational discourse*. Cambridge University Press, Cambridge
- Tavosanis M (2011) *L’italiano del web*. Carocci, Roma
- Thelwall M, Buckley K, Paltoglou G (2011) Sentiment in Twitter events. *J Am Soc Inform Sci Technol* 62:406–418
- Thelwall M, Wilkinson D, Uppal S (2010) Data mining emotion in social network communication: Gender differences in MySpace. *J Am Soc Inform Sci Technol* 21:190–199
- Volek B (1987) *Emotive signs in language and semantic functioning of derived nouns in Russian*. John Benjamins Publishing Company, Amsterdam, PA
- Watzlawick P, Helmick Beavin J, Jackson DD (1967) *Pragmatics of human communication. A study of interactional patterns, pathologies and paradoxes*. Norton & Co., New York
- Yus F (2013) *Cyberpragmatics. Internet-mediated communication in context*. John Benjamins, Amsterdam

Part III
Communication of Aggression
and Aggressive Communication

Chapter 10

Giving Voice to Silence: A Study of State Violence in Bolzaneto Prison during the Genoa G8 Summit

Adriano Zamperini and Marialuisa Menegatto

10.1 Introduction

The 2001 Group of Eight Summit (G8) in Genoa was held from 16 to 22 July. Hundreds of groups and organizations combined in the Genoa Social Forum (Gsf) to organize the collective protest around the summit that brings together the state and government heads of the eight richest countries of the world. In preparation for the events, the town was transformed into a fortress. To prevent disorder and violent clashes, the Italian government imposed safety measures. Massive police guards at each gate controlled access to the city; the airport, port, and railway stations of the city were closed to free circulation; helicopters flew over the urban spaces at low altitude, and above all, to prevent the protest from disturbing the meeting, the measures focused on keeping demonstrators out of the summit area. High iron grating 5 m high and barriers were installed to protect the so-called red zone (the area in which the summit actually took place) from the rest of the city. The deployment of forces of law and order recalled a proper state of assault.

Despite the display of force and the climate of tension and fear, the first Migrants International March on 19 July 2001, which included approximately 50,000 people, was pacifically formed without incident. However, the situation changed the following day due to what the mass media described as the provocations of the Black Bloc, followed by indiscriminate and violent police responses. On the morning of 20 July, Black Bloc members were left unchecked while damaging

A. Zamperini (✉)

Department of Philosophy, Sociology, Pedagogy and Applied Psychology,
University of Padua, Via Venezia 8, Padova 35131, Italy
e-mail: adriano.zamperini@unipd.it

M. Menegatto

Department of Philosophy, Pedagogy and Psychology, University of Verona, Verona, Italy
e-mail: marialuisa.menegatto@univr.it

or destroying banks, shops, and the city prison. For the whole day, the dynamics of events followed this pattern: after the Black Bloc raids, the police replied indiscriminately, attacking the crowd of peaceful demonstrators, including doctors, photographers, and journalists. Against the police charge, groups of demonstrators reacted by throwing stones and raising barricades. The city became a war zone, and the symbolic framework of protest was transformed into a terror zone. Ethnographic accounts referred to events as the “battle of Genoa” (Juris 2005). Police officers used tanks, and during one clash, at around 5:25 pm in Piazza Alimonda, a jeep of carabinieri (a policelike Italian corps) and its occupants were attacked by demonstrators. One of the carabinieri inside opened fire, killing 23-year-old Carlo Giuliani. On Saturday, 21 July, the day of the great concluding march organized by the Gsf, the script continued: small groups intruded among the pacific demonstrators provoking clashes and devastation. The police officers charged the demonstrators. There were beatings, numerous episodes of police “hunting down the man” (Zamperini and Menegatto 2011). Just before midnight, a special unit of the Italian police, composed of 300 agents, broke into the Diaz School, where the Gsf, the media center for journalists, and a dormitory for demonstrators had been set up, searching for weapons and bombs.

The behavior of the police was particularly brutal. An English journalist, Mark Covell, who at that moment was in front of the school gate, was overwhelmed by the platoon of police. He was subjected to a violent beating and was almost killed. Inside the school, where most of the people were sleeping, the police attacked furiously, treating the unarmed people with ferocity and destroying computers. The final result of the “bloody” operation was around 69 injured, 3 of them critically, 1 in a coma, and 93 arrested, accused of criminal behavior, resisting arrest, and unlawful possession of firearms. Among the arrested, 75 were taken to the Bolzaneto prison. In the days following the events, various eyewitnesses spoke of demonstrators’ maltreatment in the Bolzaneto prison, a center that had been set up for identifying the arrested protesters. Most would claim that they were abused and maltreated.

The approximate results of 2 days of clashes are revealed by some statistics: 253 arrested, 606 injured, 6,200 tear gas bombs fired by police, 20 pistol shots, 50 billion lire of damage, and 1 death (Parlamento Italiano). According to [Amnesty International](#) in Genoa, during the G8 there occurred the most serious abuse of human rights in a Western country since the Second World War.

The main trials involved the Diaz School and the Bolzaneto prison. Those carrying out the trials have been systematically blocked by the Italian authorities, and public institutions have in fact isolated the public prosecutors charged with conducting the investigations. If on the one hand the behavior of the public institutions has contributed to making even more precarious the compact of trust between citizens and the state, on the other hand the trials have had the benefit of being the only place where communication between the concerned parties is providing a forum for revealing an incontrovertible truth. This is especially true in the case of Bolzaneto prison. In fact, while there is an abundance of visual and audio materials documenting what happened in the streets of Genoa and at the Diaz School at the hands of the police, with respect to Bolzaneto prison there

are no pictures, video clips, or photos of the events that took place within the prison walls. Because of this lack of evidence, Bolzaneto prison has risked being perceived as an exaggerated event, the fruit of the demonstrators' imagination, or a one-sided version of events. Starting from the testimonies in one trial – narratives submitted to a rigid judicial framework (Landowsky 1989) – in this chapter we analyze the process of reconstructing the narratives (Bruner 1990; Jackson 1990; Sherwin 1994) produced during the judicial communication, which is fundamental to reconstructing the events that occurred in Bolzaneto prison.

10.2 Bolzaneto: From Barracks to Prison

During the course of the summit, Bolzaneto was part of Italian Prime Minister Berlusconi's government strategies that aimed at guaranteeing the right to peaceful protest, as well as maintaining public order and security. It was decided that two freshman offices would be set up to be used for the temporary detention of arrested demonstrators during the operations of public policing, one at Bolzaneto, the other at Forte San Giuliano. However, following the urban clashes and the death of Carlo Giuliani, who was shot by a carabinieri on 20 July, the police headquarters decided to close Forte San Giuliano to avoid another source of tension. Thus, Bolzaneto remained the only place devoted to receiving and holding all those arrested for the remainder of the summit.

Bolzaneto is a northern district of Genoa that contains the Nino Bixio Barracks, the location of the mobile VI Department of the state police, and two of its buildings were intended to serve as a "temporary prison." Like branches of prisons of Pavia, Vercelli, Voghera, and Alessandria, it should have been a place where arrested demonstrators were taken for verification of identification documents, fingerprinting, medical examinations, notification of all rights protected by law, and subsequent transport to prison. All these actions should have been carried out quickly. For this reason the first interview with a lawyer was expressly forbidden. No detainees were allowed to have contact with the outside world until they were transferred to their prison destinations. Thus Bolzaneto became an *off-limits zone*, a black hole (Portanova 2008) where inside hundreds of arrested demonstrators disappeared for hours or days at a time. Following the closure of Forte San Giuliano, the Bolzaneto structure became too small to accommodate all the demonstrators. The police presence was not sufficient to deal with this emergency. They had to find an expedient solution to the problem of surveilling an increasing number of detainees.

This type of safety strategy not only caused a lot of organizational difficulties but also changed the relationship between policemen and arrested demonstrators. Policemen became guards and demonstrators became prisoners, so that after only 3 days the Bolzaneto barracks was transformed into a real prison, a system that began to take on a life of its own. The result was that ordinary policemen who should have protected citizens' human rights were unprepared and unable to face

the situation. In fact, the atmosphere created in prison was at the root of much of the unjust behavior and evil actions (Zimbardo 2007). Many demonstrators arrested by the police were subjected to a reign of violence and abuse. Detainees suffered fractured ribs, concussions, head injuries, verbal abuse, and ill treatment. They were forced to run the gauntlet between rows of guards swinging clubs, subjected to sleep deprivation, and tortured with burning cigarettes. In addition, they were compelled to strip naked and remain standing for up to 30 h. They were then either denied any sort of medical treatment, the right to an attorney, and contact with family and were made to wait in conditions of deprivation of basic needs such as warmth, liquid, and elimination, and women were denied basic feminine hygiene products. Detainees were also forced to sing fascist songs (Zamperini and Menegatto 2011). The suffering and humiliation was all the more severe because demonstrators did not understand what was happening. Most of them did not even know why they were there because many were foreigners, for example, from Germany, France, Britain, Spain, Austria, the USA, and a number of other countries, and they did not speak Italian. They included journalists, students, artists, and representatives of organizations such as Human Rights Watch or Indymedia journalists. This segment from a public hearing explains the situation well: “But we tried to . . . well to tell him that we had nothing to do with what had happened, we did not understand why we were there, why we had been beaten . . .” (S.G.B.).

10.3 The Black Hole and the Breaking of the Community Silence

The metaphor of the black hole explains well the workings of Bolzaneto prison. Nothing escapes from a black hole. In fact, silence has shrouded Bolzaneto for a long time. Several reasons explain the deafening silence.

At the macro level, government institutions, some political parties, the police force, and the mainstream mass media said that there had been no human rights violations, consistently rejecting the demonstrators’ accounts to protect and defend the Italian state from slanderous accusations. But this was a political strategy designed to avoid facing up to what really occurred restricted to minimizing the facts and blaming demonstrators (Noelle-Neumann 1993; Zamperini 2010). Such misinformation and cover-up have created a culture of denial (Cohen 2001). According to Stanley Cohen (Cohen 2001), denial includes the spheres of *cognition*, not acknowledging and rejecting the validity of the facts, what actually happened; *emotion*, not feeling in order not to be disturbed; *morality*, not recognizing wrongness or taking any responsibility for events; and *action*, not taking active steps in response to the facts, for example, remaining indifferent to atrocities by not taking action against it. Also, when somebody has sometimes discussed G8, the communication and the use of the language have shown excessive attention to avoiding the use of words that may be disturbing for a democratic state such as,

for example, “torture,” “state-sponsored violence,” “Bolzaneto prison,” or “Diaz School,” and the consequences were that “what happened inside Bolzaneto was not really violence”!

A second reason may be that the voices of those present in Bolzaneto have had great difficulty for a long time in securing a space to talk about their experience. Some theories focus on the fact that silence is an adaptive answer or a coping mechanism that enables survival, in the face of a traumatic event, especially one that is repressive for those concerned (Comas-Diaz et al. 1998), so that they may get on with their lives. People prefer not to talk about their experiences, which serves as a means of emotional avoidance to protect themselves. This tendency causes not only the isolation of the individual victim but of the whole victim community, reducing the opportunities for them to meet each other (Lykes 1994), even to the point of breaking up relationships.

Another reason for the silence is that a few months after the Genoa G8, in particular starting in September 2001, the attention paid to the events of the G8 summit rapidly declined. This trend was confirmed in October, and it disappeared entirely shortly thereafter. As several scholars have noted (Andretta et al. 2002), the terrorist attacks on the Twin Towers and the Pentagon and the war in Afghanistan overshadowed and distracted attention from what happened in Genoa, changing the agenda and, in the eyes of the media, reducing the significance of the Genoa events.

When some bold individuals found the courage to speak up about Bolzaneto, they received no acknowledgement or legitimacy. Some have paid a high price for their testimony, for instance, Marco Poggi, a prison nurse that some officers at Bolzaneto called the “infamous one of Bolzaneto” for having “blurted out” what happened. After his testimony he was advised to leave his job because it was no longer safe, and he received anonymous threats with words like “you’ll pay for it”; in addition, he was sued for slander. Poggi was the first person who felt the need to break the circle of silence. He said that the violence occurred behind the walls of the barracks, took place in an enclosed and protected space in an environment that promised impunity. Only those who saw it could say what had happened. Only those who were there could confirm the authenticity of the story. But for a long time, Poggi was alone, ostracized by institutions and colleagues alike.

Other events that allowed the silence surrounding Bolzaneto to be broken are the formation of the Legal Team Italia and the Truth and Justice Committee for Genoa.

The first entity was formed on 4 August 2001 with the aim of guaranteeing defense to all persons who have suffered violent or unjust imprisonment, facilitating the work of around 150 lawyers in the reconstruction of events through videos, photos, audio recordings, and testimony, and establishing a chronology of events. The Truth and Justice Committee for Genoa was founded in July 2002 on the initiative of a group of persons, victims, and witnesses of abuses committed by the police during the G8 summit. The purpose of the committee is to promote knowledge about the facts of Genoa and to raise funds for legal activities; in addition, it is a cosponsor of the petition “Never again another Genoa.” The petition

calls for legislative action on the theme of torture, nonviolent training for officers in the service of public order, and the identification of policemen through the use of numbers on their uniforms.

10.4 The Juridical Conversation of the Trial and the Legal Narrative of Courtroom Interaction

A trial can be compared to a stage where the actors involved are summoned by the law to confront each other and give their own versions of the facts, ranging from prosecution to defense. A trial represents a framework with its own “grammar” that makes it possible to produce and compare different stories about events, leading to validation or falsification and creating a coherent narrative through a rigorous legal process (Landowsky 1989), forcing it to conform to the rules of history. As regards Bolzaneto, nobody could see what was happening inside. No one except those directly involved has been able to observe the interaction between police and the demonstrators. For this, the narratives produced during the trial work take on a central role compensating for the lack of independent testimonials. In fact, the simultaneous presence of the parties that characterizes the process of justice allows a new form of encounter between those who have committed and those who have suffered the crime for a construction of meaning through the intervention of a third party (Drew 1985).

In a juridical narrative, we can observe an ordered sequence of events that unfold in temporal and chronological dimensions in such a way that they form a unified sequence in which there is a beginning, a middle, and an end in order to obtain a coherent and unified story. Finally, according to Landowsky (1989), if we analyze the strategic activity from the point of view of semiotic narrative, we find two types of relationships: contradiction and negotiation. The purposes of the first relationship, comprising the strategy of contradiction, are opposite, and the comparison takes the form of a struggle in which we arrive at an asymmetric outcome. The second is the trading, whose aim is the balancing of a contract. The problem is certainly important when the accused are institutional figures. In fact, in our case, we are faced with a situation of conflict where the representation of the events is generated in the course of political conflict and is determined by antagonistic relations between groups of policemen and demonstrators (Zamperini et al. 2012). The two groups related two different and contrasting accounts of the events and portrayed two very different images of reality. On the one side, policemen described the facts as merely natural moments of tension and disorder due to the great and unexpected number of demonstrators and to disorganization. On the other side, demonstrators reported that deliberate acts of physical and psychological cruelty were perpetrated against them. In other words, participants arrived at the trial with two highly incompatible narratives of the events. The noncorrelation between the available evidence on the two sides is an impediment to determining the nature of the incident. When this

happens in a trial, usually only one of the two stories may be officially accepted as being a “true” reconstruction of reality, whereas the other is generally rejected as being false (Scheppelle 1989).

At the same time the guilty have three basic strategies for trying to cancel or at least minimize the degrading effects of the crime committed and the consequences of it. The first is a sociological strategy through which they seek to show that behind the charge are hidden particular games or interests. The second is a psychological strategy in which the accused seek to discredit the accusers, calling into question their mental competency (Boltanski 1990). Finally, the third strategy is of a semantic nature and consists in defining events in a completely different way from how it is presented by the prosecution (Giglioli et al. 1997).

This last step is crucial because in a trial the facts are the outcome of the mediation of different narratives, and this strategy has been systematically implemented by the members of the police force. Indeed, it was difficult to resort to the other strategies because it was hard to find the narratives that showed a particular interest in the acts of violence and impossible to call into question the mental competency of hundreds of people. Thus, interrogation is the main tool for obtaining information to reconstruct what happened, so as to establish the facts, determine the circumstances, and identify the actors. The underlying logic of this process of investigation requires a set of rules and implications that lead to the greatest possible efficiency, the so-called principle of cooperation (Grice 1975), which is based on four maxims: (1) quality: the contribution should be true and supported by adequate evidence; (2) quantity: the contribution should be given exactly to the extent required; (3) relationship: the contribution must be relevant and important; (4) manner: the contribution should be clear, brief, unambiguous, and orderly. According to Grice, during a conversation, the participants should be forced to give their contribution so that the conversation flows well.

However, this principle clashes with the mechanisms of “saving face,” namely, the personal need for witnesses and defendants to avoid negative responsibility. Goffman (1963), when referring to Durkheim (1965), mentions a double articulation – negative and positive – in relation to the ritual designed to save face. If the ritual is negative, it involves staying away, avoidance, and interdiction; if it is positive, the gap between donor and receiver becomes closer through offerings and gifts of various kinds. Interrogation in particular is a step in forcing the defendant to answer questions appropriately and coherently, limiting freedom of action because the only action that is allowed is to provide a consistent and relevant answer. In a regular trial, during a hearing, the participants are obliged to answer the questions and to tell the truth, while in the modern legal system witnesses must swear or affirm that their testimony will be true. The trial becomes a symbolic arena wider than a juridical-legal trial because it involves devoting considerable attention to the rituals of presentation of oneself and to strategies for the avoidance of the charges brought. During a trial the public prosecutor and the lawyers, in order to obtain information and reconstruct the facts, appeal to the principle of cooperation within a coercive framework. In other words, witnesses and defendants are forced to act in a cooperative manner (Penman 1987).

10.5 Bolzaneto: A Case Study

10.5.1 Overview

In the present study, we explore the process of reconstructing narratives (Bruner 1990; Jackson 1990; Sherwin 1994; Penman 1990) as an interactive practice between people asked to reconstruct what happened inside Bolzaneto. Although it is difficult to arrive at a single definition (Ryan 2007), in this analysis we make use of *storytelling* as a general term to indicate the manner of reconstructing and selecting past events (Ewick and Silbey 1995; Ochs 1997; Ochs and Capps 2001). According to the theory of Bennet and Feldman (1981), storytelling can be considered a way to reconstruct events and actions in a narrative form and the interactions in a courtroom, a kind of framework through the same narrative structure as the one individuals use in everyday conversations. As a result, elements such as consistency, coherence, and linguistic characteristics are taken into consideration (Bennet and Feldman 1981; Conley and Conley 2009; Cotteril 2003; Ehrlich 2010). However, this does not mean that a witness is equivalent to a narrative. In fact, in a courtroom, no kind of story is allowed (Almog 2001; Burt 2009), but in the present study we use the concept of narrative as a heuristic device that can facilitate the understanding and analysis of interactions in a courtroom (Bennet and Feldman 1981). In fact, as argued by Ewick and Silbey (1995), narratives can play a role in research as an object of investigation, for instance when studying narrative production, or a method of investigation as a means to analyze psychological or sociological factors, or, finally, as a product of investigation, for example when the results of a study are used for the production of accounts.

As discussed earlier, narratives are produced first in an interactive way since they are induced by lawyers' questions; second, they are created in a cooperative manner by witnesses. Narratives, like other social practices, are organized by the context in which they are activated, even at the constitutive level. This is particularly relevant for legal narratives. Actually, legal doctrine, namely, the way in which a trial or hearings are structured, and the system of rules that govern interactions combine to form legal narratives (Almog 2001; Atkinson and Drew 1979; Beach 1985). More specifically, the system of rules states when, what, and what kind of stories must be produced. In this respect, three aspects must be specified. First, the juridical narration is produced mainly through question and answer. Witnesses are usually not free to tell the facts but must adhere to lawyers' questioning and cross examination, in which they are subjected to a relatively fixed sequence of rounds of question and answer. During this examination the lawyers' control over witness narratives is crucial for the construction of a legally acceptable story (O'Barr 1982). Second, the witness narratives should take, as far as possible, independent and objective evidence (such as a medical report), including it consistently in their story. Finally, not all details are considered legally relevant. Thus, to comply with legal standards, witnesses should avoid, or be encouraged to avoid, certain parts of their stories (Conley and O'Barr 1990).

In this section we will concentrate on the forms of interaction between police and arrested persons that took place inside Bolzaneto prison starting from the narrative structure of the legal trial and hearings that enabled what was said to come to light, creating visibility where silence and darkness had been imposed previously. As was already stated, the trial was the first time that victims and perpetrators were called together publicly to testify, and this “narrative construction” (Darley 1999) remains the only binding material for understanding what happened.

10.5.2 Theory of Social Delegitimization

One of the more noted sociopsychological theories to explain the legitimacy of violence is *social delegitimization* of Daniel Bar-Tal (1989). In fact, to exercise an abuse of power inside a state or government institution, a “civic discharge” is often required (Zamperini 2001), a social process by which the figures of alterity are transformed into enemies and are, thus, eligible for moral exclusion (Staub 1987; Opatov 1990). Moral exclusion is a process by which individuals or groups are placed outside the normal moral community’s boundaries, perceived as beyond the boundaries within which norms and values are applied, guided by criteria of equity and justice, becoming a sort of nonentity to exploit or to be no longer worthy consideration. It is the position occupied with respect to the moral boundaries that determine different treatments aimed at opinion and social action. Consequently, anybody may become a target of violence and prevarication, with the difference being that when damage is inflicted from within, it is more likely to be considered an unjust action, following which there may be legitimate requests for repair, while when the target comes from outside the group, it is much more likely that any form of aggression against the outsiders is acceptable and appropriate, and no violation of human rights is perceived.

Specifically, in Bart-Tal’s conception, delegitimization is a process characterized by extreme moral exclusion that results in a willingness to harm someone. In addition, from the point of view of the narrative paradigm, delegitimization can be a narrative that appears in public discourse as well as in cultural and educational products (Bekerman and Maoz 2005; Savage 2007), becoming a fundamental part of the culture of conflict.

According to Bar-Tal’s theory, delegitimization is defined as a cognitive process of categorizing a group or groups into extremely negative social categories that exclude it or them from the sphere of human groups that act within the limits of acceptable norms or values so as to make physical or psychological expulsion socially plausible and morally acceptable, to commit acts of violence, or to consider maltreatment as having been deserved. Bar-Tal (1990) proposed five main processes of delegitimization; they represent the *rhetorical strategies* of categorization that authorize inflicting harm on the delegitimized group. Through *dehumanization* the group is labeled as nonhuman, and in discourse the main linguistic labels used refer to an inferior race or to the animal world and the supernatural, for instance demons,

devils, or monsters. *Trait characterization* aims at describing a group according to personal elements that are extremely negative and therefore unacceptable in a given society. The rhetorical attribution tends to portray the components of a group as transgressors of fundamental norms and deviants. *Political labels* are used to denote a political group that threatens the constituent values of a collective, and for this reason the members of that group are totally rejected because they are dangerous. A fourth process of delegitimization is known as *outcasting* and involves rhetorical categorization into groups that violate the pivotal shared social norms, referring to social actors who are not organized, such as murderers, thieves, terrorists, or maniacs. Finally, delegitimization by *group comparison* occurs when a highly negative label is assigned to a group; the label is symbolic and represents the most undesirable group of a certain culture, such as vandals or barbarians. In general, while some groups are categorized negatively but continue to be part of the society, delegitimization is a social phenomenon that indicates that a particular group exists only outside the boundaries of commonly accepted groups.

10.5.3 Method: Data and Tools

The data of our study are defined by demonstrators' testimony and public hearings given during the first trial in Genoa from 12 October 2005 to 14 July 2008. The Bolzaneto trial produced $n = 179$ public hearings with an audio record of 468 h of juridical conversations; $n = 361$ witness reports were collected, and $n = 208$ demonstrators spoke during the trial, classified as 162 males and 46 females (average age 25.8; age range 16–51). The transcripts of witness textual data were processed using Atlas-ti, a software analysis program.

10.5.4 Results

The analysis, which focused on the narratives produced by $n = 208$ demonstrators, representing 57 % of the entire sample of witnesses present at the trial, shows that references to phenomena of delegitimization constitute $n = 1,183$ out of a total of 208 testimonials, with an average of 5.7 for each demonstrator.

The use of *political labels* occurs most frequently, with a total of 367 mentions, in particular for the environmental delegitimization with 162 mentions in 208 cases. *Trait characterization* follows with a total of 341 mentions, while the remaining labels were 203 *outcasting*, 174 *group comparison*, and 98 *dehumanization*. In the present study, next to the classic assumption exposed by Bar-Tal's theory, two new forms of delegitimization have emerged from this analysis: the first consists in making the delegitimized group assume a specific behavior; in the second, the delegitimizing group uses rhetoric that refers to hostile symbolic frames. This new conception of delegitimization leads us to differentiate and define three types

Table 10.1 Strategies of social delegitimization

Strategies of social delegitimization	Frequency	Total	%
Political label			
Definitional	136	367	31
Behavioral	69		
Environmental	162		
Trait characterization			
Definitional	267	341	29
Behavioral	8		
Environmental	66		
Outcasting			
Definitional	150	203	17
Behavioral	19		
Environmental	34		
Group comparison			
Definitional	137	174	15
Behavioral	32		
Environmental	5		
Dehumanization			
Definitional	73	98	8
Behavioral	12		
Environmental	13		
Total	1,183	1,183	100

of delegitimization: one is called *definitional*, as Bart-Tal's model suggests, and in addition there are *behavioral* and *environmental* delegitimization (Zamperini and Menegatto 2011; Menegatto and Zamperini 2012). So for each of the five processes of extreme categorization in Bar-Tal's conception, we obtained a three-part delegitimization based on different modes of interaction between police and demonstrators (Table 10.1).

10.5.4.1 Political Label

Definitional

Protesters' narratives show how these linguistic methods register the prisoners as politically belonging to the left or extreme left. Those stopped are the target of delegitimization through the frequent use of terms such as *communist* or *anarchist* or describing the political danger in terms of *terrorist* or *black bloc* or even as a representative of a social center like *Casarini's son*.

Behavioral

In this specific case, demonstrators are forced to sing fascist songs or assume postures that hearken back to fascist ideology: “*We were forced to line up and make a fascist sign and sing ‘faccetta nera’ and Many of us were forced to raise our right arm in a ‘roman salute.’*”

Environmental

The environment is used as a delegitimization tool through the creation of a framework with political meanings. In this form of delegitimization we find the police force, which distributed hymns and songs in praise of fascist or dictatorial systems: “*With the cell phone ringtones they (police force n.d.r.) made us hear songs of the regime ‘faccetta nera’*”; and “*Some agents (. . .) hummed ‘One, two, three, cheers for Pinochet.’*”

10.5.4.2 Trait Characterization

Definitional

In this category the qualifying strategies used refer to the use of terms aimed at casting dishonor on people, such as *bastard*, *piece of shit*, or *motherfucker*.

Behavioral

In this case, personal discrediting is accomplished by the same delegitimized person. The demonstrators are forced to use the same language as the delegitimizing police force, for example: “*This person was close to the grating and had to . . . he was obliged to repeat that he sucked, that he was shit.*”

Environmental

In this category, delegitimization manifests itself in gestures that, directly or indirectly, places the demonstrators in a context that communicates negative characteristics. For example, a prisoner is suddenly spat at in his face or an agent comes up close to a small group of prisoners and farts with indifference.

10.5.4.3 Outcasting

Definitional

In this case, delegitimizing strategies draw from the broad classification of *outcast* that has been used throughout human history, that is, individuals that various communities exile to the margins of society or even reject from the circles of “us.” In the language of the delegitimized, these demonstrators are considered outsiders for socio-racial reasons by terms like *Jews* or *gypsies* or by sexual orientation like *gay*, *prostitute*, or *lesbian* or by social conduct like *revolutionaries*, *murderers*, or *drug addicts*. This language aims at giving demonstrators an outsider identity.

Behavioral

In this category, demonstrators assume a series of behavioral types that usually characterize the outsider. From a narrative: “*After she vomited (. . .) she asked for a rag, she asked for something to clean up with, and they kept saying ‘no’, even telling her ‘now clean it up with your tongue, we don’t care if you have made a mess’; or ‘I was ordered to collect the garbage and put it out . . . to collect the garbage.’*”

Environmental

History has taught us that outsiders can be physically cut off from the community of respectable people and confined to special places. Since the nineteenth century it has been possible to have symbolic resources to create an environment that expresses social exclusion. From the account of a witness a smiling agent emerges to welcome the demonstrators saying, “*Welcome to Auschwitz,*” or with a chilling tone he declares that “*The Nazis are thinking of you.*”

10.5.4.4 Group Comparison

Definitional

Here delegitimization is organized in ascribing to the outgroup the essence of negativity, before a virtuous ingroup. For example, a kefia worn by a demonstrator becomes the pretext for verbal attack: “*A keffiyeh was found. I was asked if that was a symbol of the Arabs and with subsequent insults to the Arab world.*” Women are addressed as *housewives of shit*.

Behavioral

At the behavioral level, the group comparison is shown through extolling the police force and order, that is, the social category to which the pole of positivity is attributed: *“When we were waiting by the walls of the corridor they made us say: ‘Long live the penitentiary police’; they made us repeat it.”*

Environmental

The ingroup’s (police force) superiority compared to the outgroup (demonstrators) is not only displayed in a qualitative way (good vs. evil) but also in terms of quantity (many vs. few). One of the methods adopted by the police force was to shout out throughout the area questions like: *“Carlo Giuliani, where is Carlo Giuliani?”* or singing songs like *“One less, you are one less,”* referring to the tragic killing of a young boy.

10.5.4.5 Dehumanization

Definitional

This category relegates demonstrators to the animal kingdom, stripping them of human attributes. The police force refers to them as *pigs, monkeys, ticks, and rats*. An example from the reports: *“We were very cold because there was no glass in the window of the cell and they distributed blankets that smelled of disinfectant and there were not enough to cover everyone, and then we stayed close to one another to cover us all. In this situation I perfectly remember a comment coming from the door: ‘look at them, how they get close, like animals.’”*

Behavioral

The police officers kept up the procedure of expelling the demonstrators from the human race, forcing them into action. For example, the order was given to one prisoner to *“jump chamois!”* Another prisoner had his shoelaces removed and thrown to the ground with the order: *“Now bend over and pick them up and get down on all fours and bark.”*

Environmental

The denial of humanity is completed by inserting into the prison environment elements that suggest an animal refuge. The officers simulated animal sounds to instill fear in the detainees: *“They imitated dogs to scare us”* or *“Sing a song ‘on the old farm.’”*

In the contexts in which abuses occur at the collective level, there is normally a climate of impunity that protects those who commit acts of violence. In addition, dominant groups devote themselves to ignoring or silencing the injustice and suffering produced because their admission is perceived as too dangerous or destabilizing. As regards the events that occurred during the Genoa G8 Summit, with the exception of some sporadic initiatives, the trial was the only public meeting space, an instrument through which it was possible to break the “wall of silence” built around the events. Only in the courtroom has a comparison and an assessment of what really happened been possible, providing detailed documentation of what happened inside the prison and acknowledging the truth about the abuses. This trial has exceeded its traditional mandate with respect to victims’ sociopsychological needs. In fact, in highlighting abuses and violent acts, justice has taken a leading role in the healing process for victims of the violence.

10.6 Implications of the Findings: Therapeutic Jurisprudence and Victims of State Violence

Therapeutic jurisprudence – the role of the law as a therapeutic agent – focuses on how the law can facilitate achieving psychological well-being. From this perspective, legal rules and procedures and the roles of legal actors (such as lawyers and judges) constitute social forces that, whether intentional or not, can produce therapeutic or antitherapeutic consequences (Petrucci et al. 2003). Therapeutic jurisprudence seeks to utilize psychological and social science research to explore these questions empirically. Although therapeutic jurisprudence has been criticized for not offering a clear and limited definition of the term *therapeutic* (Melton 1994), the decision not to offer a narrow definition of therapeutic enables researchers to use a wide range of concepts that characterize individual and social well-being. Thus, what is meant by *therapeutic* far exceeds the opposite of ill health.

Therapeutic jurisprudence, which began in the late 1980s and grew to become a truly interdisciplinary enterprise, is consonant with restorative justice (Braithwaite 2002; Scheff 1998), community justice (Bazemore and Schiff 2001), social justice (Montada and Lerner 1996), and procedural justice (Thibaut and Walker 1978; Tyler 1992). Therapeutic jurisprudence emphasizes the importance of how victims perceive the legal process, allowing a merging with victimology research (Wright 1996), and pays attention to how the legal system responds to victims (Wexler 1990). Therapeutic jurisprudence also emphasizes the importance of empowering citizens to tell their story in an environment that supports dignity, trust, and respect. A key feature of therapeutic jurisprudence is that it provides an invitation to consider the legal process from a multitude of perspectives that are often overlooked. In many cases, this means considering those who previously had no voice, such as victims of domestic violence (Winick 2000).

On the basis of this research, we suggest that therapeutic jurisprudence can also be extended to cases of state-sponsored violence. As has been widely documented (Krug et al. 2002; Minow 1998), the victims of such violence have their basic human needs compromised: their sense of security, positive identity, self-efficacy, and social bonds. In fact, as we have demonstrated in the case of the G8 Summit in Genoa, violence was exercised within a sociopolitical context that sought, through processes of delegitimization, to represent the victims as a foreign body to the moral and legal community. In these situations, suffering occurs on two levels: the physical body and psychosocial recognition (Zamperini and Menegatto 2013). Even unconsciously, often the victims feel they are “wrong,” as if what they experienced was somehow deserved because of personal characteristics or actions taken.

For this reason, the victims expect justice to restore their rights and ensure fair compensation, but above all, recognition of the oppressed group is also important (Garapon 2002). In fact, the victims of collective and political violence often experience a sense of isolation within their community, even among their friends and family, and they feel that their dignity has been violated by the institutions that have delegitimized and labeled them as “deviants” or “criminals.” It follows that the victims have a desperate dual need: (a) to regain their dignity and honor through collective reparations in which their private truth is validated so that it becomes a public truth; (b) to obtain acknowledgement of the wrong, the suffering, and to ensure that it is assumed as a part of social memory. It is important that both collective reparations and acknowledgement take place in a common space having a common function where collectivity converges in the same direction: perpetrators and victims must shape a common memory or a common element for the construction of historical truth. To that end, the victims may find in the legal and institutional sphere (juridical or innovative as in the South African TRC) a place for a process of reparation and acknowledgement. Moreover, a trial may serve to foster an expression of a plurality of different points of view, coming to a common construction of meaning thanks to a third subject above the parties (Garapon 2002).

During a trial, victims have the opportunity to show themselves and tell their story. This process frees the victims from the status of inferiority to which they have been relegated, regaining the right to speak and to be heard. They pass from a state of passivity to a state of active agent. Such a perspective lends great importance to their testimony, which is not limited to the reconstruction of the facts but becomes a tangible proof that the victims’ words are again active and taken into consideration. The testimony becomes the psychological rescue of a life wrecked by historical events, and the narrative is the primary tool for the reconstruction of the identity. Through their testimony, victims regain consciousness and an active role in the construction of the historical truth of their experience. Obtaining the perception of the effectiveness and importance of the *storytelling*, the experience of the narrative becomes an object of communication. With this transformation, the person is in a position to lighten the weight of his or her own pain, previously endured in solitude. And if the listener is an official audience as in a legal trial, the audience creates the conditions whereby the violent experience of the victim is inserted into a political narrative that can give meaning and illumination to the events and give voice to the silent and obscured violence.

Then the trial tends to gain value as a witness, transforming the space as a common space to celebrate memory. Then, during the trial, those who participated as witnesses in the definition of what happened inside the Bolzaneto barracks contributed to the creation of a social memory, ready to be shared with the rest of the community thanks to a process of transmission of communication. Through a juridical process, every victim, through special words, memories, and revealed emotions, has “worked” to generate a common and collective heritage starting point for the respect of civil and human rights. Thus, for the victim, therapeutic jurisprudence consists of a psychopolitical rebirth.

For these reasons, in the case of the Genoa G8, the common space has been well represented by the Bolzaneto trial, which was the first institutional and collective stage at which all the actors of those events were summoned. The trial was more than a juridical convocation becoming a sort of “civic call,” able to bring back within a community’s bounds the theme of justice and suffering, to reintegrate the victims in the group of citizenship, and to bring to light what was obscured. For this reason we have called it the “book of memory” (Zamperini and Menegatto 2011).

10.7 Conclusion

Political violence during a protest event is always reconstructed in the public sphere by two adversarial coalitions: public order and social movements. While violence in the streets is limited to police and demonstrators, the discursive battle is fought mainly by speakers such as politicians, journalists, judges, and so on. In democratic systems, this battle is a battle for public opinion, a battle for imposing one narrative as the dominant interpretation of events. In these scenarios, the legal conversation that occurs during a trial is a “social force” involved in this narrative reconstruction of violence. The legal conversation thus can help in illuminating the dark side of violence, forcing the different actors involved to contribute to a shared narrative as much as possible. The court thus becomes an arena that creates social memory. A memory of violence that interrogates the political institutions of the state, because it defines who did what, orients attitudes with respect to what happened, and shows the way we look to the future. In the case of the G8, the memory of the institutions and the memory of the demonstrators are still in conflict. Therefore, through the “social force” of the trial the many testimonies given during the hearings can become the heritage of the community and lay the foundations for the construction of a shared memory.

In addition, the conversational analysis of the process of Bolzaneto shows how therapeutic jurisprudence and social psychology can work in tandem. Well-being is a fertile ground in social psychology, with a multitude of directions in which to proceed, depending on who is being considered: an individual, a family, a community, or society at large. As can be seen, the “who” of well-being considered here have been victims of state violence in Bolzaneto. Therapeutic jurisprudence provides a framework that is useful to social psychology for the integration of well-

being into the legal arena, and it draws attention to constructs such as dignity, trust, and respect. In societies where state violence is manifested, a climate of impunity applies that protects the perpetrators, based on silence and denial. Often the victims are deprived of their right to speak, to communicate what happened, or to denounce the perpetrators of the violence. The sense of helplessness, the anxiety aroused by the delegitimization, and the psychological burden of silence represent suffering that also needs to be addressed, especially in a public and supportive environment. The trial surrounding Bolzaneto, even with all its limitations, has been able to provide victims with a kind of therapeutic environment. Beyond the strictly legal outcome, the process has allowed the victims to reverse the delegitimization they suffered; regaining that social recognition is essential to the subjective well-being of every person.

Similar results are found in social practices in order to restore a climate of peace after collective and political conflicts (Sapio and Zamperini 2007). In particular, there are similarities with the psychological processes activated by commissions for truth and reconciliation both in the case of intrastate conflicts, such as that of South Africa (Allan and Allan 2000), and in the case of conflicts of community and of the racist violence that took place in the city of Greensboro, North Carolina, in the USA (Magarrel and Wesley 2008). These commissions start with an assumption, that it is possible to help people by allowing them to tell their own stories, obtaining in this way public acknowledgement. Those who have worked with victims of collective violence relate how the story of the trauma suffered can be transformed through testimony, thereby passing from a story of shame and humiliation to a narrative of dignity (Sironi 2007). Therefore, empowerment – that is, to regain a sense of power and control over one's life, to recover social membership, and to receive a sense of positive identity and acknowledgement – becomes the basis for embarking on an appropriate healing path.

It is possible to trace some indicators of the transition from a state of powerlessness to a state of empowerment even within the demonstrators of the Genoa G8. In fact, research that we conducted in 2010 with $n = 167$ demonstrators (Zamperini and Menegatto 2011) – 2 years after the trials ended – showed a decrease of emotions related to the delegitimization (shame and self-blame) and the rise of assertive emotions (pride and indignation) that orientated the collective demand of apologies and admission of guilt by the Italian institutions.

References

- Allan A, Allan MM (2000) The South African truth and reconciliation commission as a therapeutic tool. *Behav Sci Law* 18:459–477
- Almog S (2001) As I read I weep: in praise of judicial narrative. *Okla City U L Rev* 26:471–501
Amnesty International. <http://www.amnesty.org>
- Andretta M, della Porta D, Mosca L, Reiter H (2002) Global, nonglobal, new global. La protesta contro il G8 a Genova. Laterza, Roma-Bari

- Atkinson JM, Drew P (1979) *Order in court: the organization of verbal interaction in judicial settings*. Macmillan, London
- Bar-Tal D (1989) Delegitimization: the extreme case of stereotyping and prejudice. In: Bar-Tal D, Graumann C, Kruglansky AW, Stroebe W (eds) *Stereotyping and prejudice: changing conceptions*. Springer, New York, pp 169–182
- Bar-Tal D (1990) Causes and consequences of delegitimization: models of conflict and ethnocentrism. *J Soc Issue* 46:65–81
- Bazemore G, Schiff M (2001) *Restoring community justice*. Anderson Publishing Co, Cincinnati, OH
- Beach R (1985) The use of rhetorical strategies in narrative and analytic modes. *American Educational Research Association*, Chicago, IL
- Bekerman Z, Maoz I (2005) Troubles with identity: obstacles to coexistence education in conflict ridden societies. *Identity* 5:341–358
- Bennet L, Feldman M (1981) *Reconstructing reality in the courtroom*. Rutgers University Press, New Brunswick, NJ
- Boltanski L (1990) *L'amour et la justice comme compétences. Trois essais de sociologie de l'action*. Métailié, Paris
- Braithwaite J (2002) Restorative justice and therapeutic jurisprudence. *Crim L Bull* 38:244–262
- Bruner J (1990) *Acts of meaning*. Harvard University Press, Cambridge
- Burt MN (2009) The importance of storytelling at all stages of capital case, 77. *UMKC Law Rev* 4:877–910
- Cohen S (2001) *State of denial. Knowing about atrocities and suffering*. Polity Press, Cambridge
- Comas-Diaz L, Lykes MB, Alarcon RD (1998) Ethnic conflict and the psychology of liberation in Guatemala, Peru and Puerto Rico. *Am Psychol* 53:778–792
- Conley RH, Conley JM (2009) Stories from the jury room: how jurors use narrative to process evidence. *Stud Law Polit Soc* 49:25–56
- Conley JM, O'Barr WM (1990) *Rules versus relationships: the ethnography of legal discourse*. University of Chicago Press, Chicago, IL
- Cotteril J (2003) *Language and power in court: a linguistic analysis of the O.J. Simpson trial*. Palgrave MacMillan, New York
- Darley JM (1999) Methods for the study of evil-doing actions. *Pers Soc Psychol Rev* 3:269–275
- Drew P (1985) Pre-trial justice – a study of case settlement in magistrates courts – Baldwin. *J Sociol* 19:626–628
- Durkheim E (1965) *The elementary forms of religious life*. The Free Press, New York (1912)
- Ehrlich S (2010) Courtroom discourse. In: Wodak R, Johnstone B, Kerswill P (eds) *The SAGE handbook of sociolinguistic*. Sage, London, pp 361–375
- Ewick P, Silbey SS (1995) Subversive stories and hegemonic tales: toward a sociology of narrative. *L Soc Rev* 29:197–226
- Garapon A (2002) *Des crimes qu'on ne peut ni punir ni pardonner. Pour une justice internationale*. Odile Jacob, Paris
- Giglioli P, Cavicchioli S, Fele G (1997) *Rituali di degradazione: anatomia del processo Cusani*. Il Mulino, Bologna
- Goffman E (1963) *Behavior in public places. Notes on the social organization of gatherings*. Free Press, New York
- Grice HP (1975) Logic and conversation. In: Cole P, Morgan JL (eds) *Syntax and semantic, vol 3, speech acts*. Academic, New York, pp 41–58
- Jackson BS (1990) Narratives theories and legal discourse. In: Nash C (ed) *Narrative in culture. The uses of storytelling in the sciences, philosophy, and literature*. Routledge, London, pp 23–50
- Juris J (2005) Violence performed and imagined: militant action, the black bloc and the mass media in Genoa. *Crit Anthropol* 25:413–432
- Krug EG, Dahlberg LL, Mercy JA, Zwi AB, Lozano R (2002) *World report on violence and health*. World Health Organization, Geneva
- Landowsky E (1989) *La société réfléchie. Essais de socio-sémiotique*. Seuil, Paris

- Lykes MB (1994) Terror, silencing and children: international, multidisciplinary collaboration with guatemalan maya communities. *Soc Sci Med* 38:542–543
- Magarrel L, Wesley J (2008) *Learning from Greensboro: truth and reconciliation in the United States*. University of Pennsylvania Press, Philadelphia, PA
- Melton G (1994) Therapy through law. *Contemp Psychol* 39:215–216
- Menegatto M, Zamperini A (2012) Gender and state violence at the G8 in Genoa. *Riv Sessuol* 36:133–140
- Minow M (1998) *Between vengeance and forgiveness: facing history after genocide and mass violence*. Beacon Press, Boston, MA
- Montada L, Lerner MJ (eds) (1996) *Current societal concerns about justice*. Plenum, New York
- Noelle-Neumann E (1993) *The spiral of silence: public opinion – our social skin*. University of Chicago Press, Chicago, IL
- O’Barr WM (1982) *Linguistic evidence: language, power, and strategy in the courtroom*. Academic, New York
- Ochs E (1997) Narrative. In: van Dijk TA (ed) *Handbook of discourse: a multidisciplinary introduction*. Sage, London, pp 185–206
- Ochs E, Capps L (2001) *Living narrative: creating lives in everyday storytelling*. Harvard University Press, Cambridge
- Opatov S (1990) Moral exclusion and injustice: an introduction. *J Soc Issue* 46:1–20
- Parlamento Italiano. <http://wai.camera.it>
- Penman R (1987) Discourse in courts: cooperation, coercion, and coherence. *Discourse Process* 10:201–218
- Penman R (1990) Facework and politeness: multiple goals in courtroom discourse. *J Lang Soc Psychol* 9:15–38
- Petrucci CJ, Winick BJ, Wexler DB (2003) Therapeutic jurisprudence: an invitation to social scientists. In: Carson D, Bull R (eds) *Handbook of psychology in legal contexts*, 2nd edn. John Wiley & Sons, Chichester, pp 579–601
- Portanova M (2008) *Inferno Bolzaneto*. Melampo, Milano
- Ryan ML (2007) Toward a definition of narrative. In: Herman D (ed) *The Cambridge companion to narrative*. Cambridge University Press, Cambridge, pp 22–36
- Sapio A, Zamperini A (2007) Peace psychology: theory and practice. In: Galtung J, Webel C (eds) *Handbook of peace and conflict studies*. Routledge, London, pp 265–278
- Savage R (2007) Disease incarnate: biopolitical discourse and genocidal dehumanization in the age of modernity. *J Hist Sociol* 20:404–440
- Scheff TJ (1998) Therapeutic jurisprudence forum. Community conferences: shame and anger in therapeutic jurisprudence. *Riv Juridic Univ Puerto Rico* 67:97–119
- Scheppele KL (1989) Telling stories. *Mich Law Rev* 87:2073–2098
- Sherwin RK (1994) Law frames: historical truth and narrative necessity in a criminal case. *Stanford Law Rev* 47:39–83
- Sironi F (2007) *Psychopathologie des violences collectives. Essai de psychologie géopolitique clinique*. Odile Jacob, Paris
- Staub E (1987) Moral Exclusion and Extreme Destructiveness: Personal Goal Theory, Differential Evaluation. Moral equilibration and steps along the continuum of destruction. Paper presented at American psychological association meeting, New York
- Thibaut J, Walker L (1978) A theory of procedure. *CA L Rev* 66:541–566
- Tyler T (1992) The psychological consequences of judicial procedures: implications for civil commitment hearings. *SMU L Rev* 46:433–445
- Wexler DB (1990) *Therapeutic jurisprudence: the law as a therapeutic agent*. Carolina Academic Press, Durham
- Winick BJ (2000) Applying the law therapeutically in domestic violence cases. *U Mo K C L Rev* 69:33–91
- Wright M (1996) *Justice for victims and offenders: a restorative response to crime*. Waterside Press, Winchester, VA

- Zamperini A (2001) *Psicologia dell'inerzia e della solidarietà. Lo spettatore di fronte alle atrocità collettive*. Einaudi, Torino
- Zamperini A (2010) *L'ostracismo. Essere esclusi, respinti, ignorati*. Einaudi, Torino
- Zamperini A, Menegatto M (2011) *Cittadinanza ferita e trauma psicopolitico*. Liguori, Napoli
- Zamperini A, Menegatto M (2013) *La violenza collettiva e il G8 di Genova. Trauma psicopolitico e terapia sociale della testimonianza*. *Psicoter Sci* 47:423–442
- Zamperini A, Menegatto M, Travaglino GA, Nulman E (2012) *Social representations of protest and police after the Genoa G8 summit: a qualitative analysis of activist accounts of events*. *Pap Soc Represent* 21:15.1–15.30
- Zimbardo PG (2007) *The lucifer effect: understanding how good people turn evil*. Rider, London

Chapter 11

The Rhetoric of Conflict Inside and Outside the Stadium: The Case Study of an Italian Football Cheer Group

Rosa Scardigno, Maria Luisa Giancaspro, Amelia Manuti,
and Giuseppe Mininni

11.1 Theoretical Background

11.1.1 *Intergroup Conflict and Social Identity in Sport Organized Support*

Within the sports domain, organized support is a widespread phenomenon across nations. Most recently, sociologists and psychologists have devoted much attention to it because of its potential in engendering conflict and violence (Stott and Reicher 1998). Over the decades, several explanations about the origin of these so-called riots have been given (Stott et al. 2001).

In accordance with the group mind hypothesis (Le Bon 1895, trans. 1947), open conflict is seen as irrational and normless, as a natural consequence of being committed to large groups, which shuts off individual rational control. In this overview, minds are not confined to the boundaries of individuals, and some groups can present specific features that are not apparent in their members (Theiner et al. 2010).

On the other hand, the so-called individualistic tradition (Allport 1924) has claimed that group behaviors may derive from participants' common traits (this is the case, for instance, with violent and antisocial personalities). In this perspective (Dunning 1994), the football crowd conflict and violence should be related to the convergence of "roughly" socialized persons (Suttles 1968). Among the latest approaches, the Leicester school (Dunning et al. 1988, 1991) has claimed that violence can be related to the presence of these typologies of individuals in the

R. Scardigno • M.L. Giancaspro • A. Manuti (✉) • G. Mininni
Department of Education, Psychology, Communication,
University of Bari Aldo Moro, Piazza Umberto I, Bari I 70121, Italy
e-mail: rosa.scardigno@uniba.it; amelia.manuti@uniba.it

football context more than in other ones. Thus, applying the self-categorization theory (Turner et al. 1987), in crowds, individuals do not simply lose control, rather they act in accordance with their contextually specified and relevant social identification (Reicher 1984, 1987, 1996; Drury and Reicher 2000; Potter and Reicher 1987).

A more psychological account (Kerr 1994) assumes that “hooligans” are individuals subjected to excessive arousal, which may arise from perceived lacks in their everyday life. In this vein, more recent developments of the theory of fan aggression in sports (Wann and Grieve 2005) incorporates the spectator’s need for a positive social identity and his or her degree of identification with the team as being critical in setting the stage for aggression. It is argued that spectator aggression is more likely to occur in persons high in identification with the team relative to those low in identification and that this aggression is a direct result of the human desire to create and maintain a positive social identity. That is, highly loyal fans, namely members of cheer groups, because of their inability to disassociate themselves from unsuccessful groups in an attempt to protect their identity, will resort to the aggressive strategy of shouting at or derogating others to try to regain their positive social identity.

Even if these approaches have enabled researchers to better explain the riot phenomenon, they have attracted some criticisms: several studies could neither individuate attributes predicting riot participation nor clarify the relations between violence and crowd behaviors (Stott and Reicher 1998).

In addition, some ethnographic accounts emphasize the importance of impression management and possible differentiation of fans, as in the case of the “carnivalesque” conduct of Scottish fans in the 1990 Italian World Cup versus English hooligans (Giulianotti 1991). In other words, fans can vary their conduct and behavior in accordance with their adherence to different “discourses” and forms of action (Giulianotti and Robertson 2004). As a consequence, beyond these theoretical and, certainly, subjective perspectives, social and contextual factors and dynamics should be considered when talking about social conflict among fans. Therefore, a more situated perspective is advocated so as to better investigate cheer groups as cultural and discursive constructions.

11.1.2 Cheer Groups as Cultural and Discursive Forms of Life

In the field of cultural (Cole 1998) and discursive psychology (Harré and Gillet 1994; Mininni 1995), the social, cultural, historical, and political context in which people live is fluid and malleable, based on the discursive practices and actions of all social groups. In accordance with socio-constructivism, both private and public discourses (including mass media) play an important role in forming and informing ingroups and outgroups (Kohl 2011; Power and Peterson 2011) and, therefore, are an integral part of the dialogical construction of both subjective and social positioning and social relations (Power 2011; O’Sullivan Lago 2011). As a consequence, the

construction of inter- and intragroup interactions and the nature of cheering fans' subjective and collective identities can be culturally and discursively conveyed.

In this domain, social conflict can find some cultural and discursive basis in the process of dehumanization, which is an extreme form of depersonalization where individuals and their groups can lie outside the boundaries of both humanity and morality (Tileagă 2007, 2012). In its several forms, dehumanization denies the emotive and moral qualities of persons. As a consequence, it can make injustice normal and acceptable, activate forms of moral disengagement, and, at the extreme, represent a clue for possible violent behaviors, such as expulsion and atrocity.

The construction of intergroup and dehumanizing processes can be acted by some kinds of metaphor, which is an essential and performative discursive tool for the social construction of reality (Volpato 2011, 2012):

- “Animalization” entails presenting others as irrational, immature, rough, uneducated, behaving in a primitive way. This kind of metaphor is related to some forms of western culture; in particular, it was used in the colonial period to emphasize the myth of the white man by delegitimizing opponents (Jahoda 1999). Further studies are needed to deepen both the negative features and the positive values (power and status) of animal metaphors.
- “Demonization” involves the depiction of human beings as “monsters” because of some excess, blemish, or deformity. This kind of metaphor started in the medieval period (e.g., the giant became a savage figure), whereas in the Christian era the essence of wickedness was represented by the devil (Le Bras-Chopard 2000).
- “Biologization” makes use of metaphors about disease and purity and transforms others into viruses, pestilences, cancers, tumors, and pollution. The genocides of the last century found their origins in this metaphor and in the resulting practices of cleansing, suppression, and purification (Savage 2007).
- “Mechanization” involves presenting others as mechanical organisms who are unable to feel emotions, detached, passive, without curiosity, compassion, or imagination. Taking into account both the medical and the technological domain, this metaphor found its greatest application in the Tayloristic idea of work.
- “Objectification” is where individuals are considered as objects, tools, or goods, in the worst cases as slaves. Indeed, several dimensions characterize this metaphor, but the most dangerous is instrumentality: the person becomes “useful” because of some traits. As a consequence, in contrast to the other types, this involves approaching (rather than leaving) the person.

Classic psychosocial studies (Allport 1954; Kelman 1973; Opatow 1990) identified dehumanizing dynamics as a possible basis for social disqualification, atrocities, and extreme violence, expedited by the negation of the humanity and the moral properties of the group.

All of these phenomena can be constructed via discourse: Bruner (1990) asserted that what people say is just as important as what they do. In this vein, it could be argued that understanding of intergroup conflict in sports could be fruitfully framed from a cultural and discursive perspective.

11.2 Case Study

11.2.1 Participants, Aims, and Methodology

The aim of the present study is to investigate the discursive construction of cheer culture, specifically focusing on the dynamics of ingroup and outgroup identity construction. The main assumption of this study is that the analysis of these constructs may help researchers in investigating the basis of social conflict. To this end, we encountered some exponents of a national organized football cheer group, named *Drughi* (fans of Juventus F.C.), and we listened to some public discourses on their radio program (named “*3 stelle sulla pelle*” = “Three stars on the skin”). Even the title of the program is intended to fulfill a polemical and provocative aim: several years ago, the Juventus Football Club was legally punished, so two championships were rescinded. In other words, even if Juventus actually had two stars, the fans claim that the team had three stars earned “on the field” and imprinted on their skin (Fig. 11.1).

Four episodes of the program were analyzed, two after two winning matches and two after two losing matches. In all cases, the matches were played against what fans perceived as bitter rivals.

The episodes were recorded on Mondays – so after each match, usually played on either Saturday or Sunday – and lasted approximately 1 h. The first one was broadcast following a loss to the Inter team (3 November 2012), the second one following a loss to the Milan team (25 November 25 2012), the third one following a win over the Milan team (21 April 21 2013), and the fourth one following a win over the Turin team (28 April 28 2013).



Fig. 11.1 Flyer of the radio program analyzed

Each episode starts with a theme song celebrating the Juventus team. Then the program is animated through comments, tales, jokes, and gags played by four people: the presenter, a journalist who is a fan of the team; the president of the Drughi, a man from northern Italy; one of the main representatives of the group, named “Ciccio”, from southern Italy; and another fan, named the “Lawyer,” from northern Italy. During the program, other guests can be contacted, for example, the head of the Drughi group from Rome. The program ends with the same opening theme song.

The texts were transcribed and analyzed using qualitative content analysis and diatextual analysis. In the first approach, we focused on the “what” of communication, concentrating specifically on the topics of the Drughi culture. In the second approach, diatextual analysis allowed us to investigate the “how” of communication. Actually, diatextual analysis is a special kind of discourse analysis that aims at identifying and understanding the strict nexus between identity and discourses/texts.

Diatextual analysis is based on the assumption that sense does not reside permanently within texts; rather, it permeates them as a result of the conjunct action of the speakers, who negotiate the frame of the situation (*stake*) in which they are actively involved. It tries to glean the meaning of a discourse by answering three basic questions that organize the interpretative procedures of the SAM model since they suggest looking for a series of markers that identify the *subjectivity*, *argumentation*, and *modality* of discourses (Mininni 1992, 2003a, 2011, 2013).

The first question (*who says that?*) aims at clarifying the way the text treats its subjects by weaving complex links with the image the speaker elaborates of him/herself and of the addressee.

1. Agency markers include all textual units showing whether the speaker is a source or goal of some action. Both discursive and grammatical tools are used to position oneself as an “actor” (who makes choices, elaborates programs, makes decisions) or “patient” (who is subjected to external powers without any responsibility): first-person pronouns state the speaker as the subject; modal verbs indicate that the person has no choice (a person *must do* something); in addition, passive verbs and marking someone as an object rather than as a subject hide the agency.
2. Affectivity markers reveal the emotional dimension of texts. In particular, the emotional extent of discourses can be identified using the following markers (Caffi and Janney 1994):
 - (a) Evaluation markers, which position statements on the axis good/bad, positive/negative, and so on.
 - (b) Proximity markers, which metaphorize the emotion through the physic, social, or temporal gap.
 - (c) Specificity markers, which modify the reference world (e.g., affective distance can be shown by dealing with specific arguments in a general way).
 - (d) Evidentials, which regulate the credibility and the authority of what is said using modals and subjective/objective verbs.
 - (e) Volition markers, which modify the self-identification level with respect to partners, for example, through a declarative, interrogative, or imperative tone.

- (f) Quantity markers, which intensify or weaken the meaning of a sentence by the degree of adjectives and adverbs, repetitions, and phonological lengthening.
3. Embrayage and débrayage markers make it possible to reveal whether or not the speaker is involved in discursive acts by enhancing or clouding any reference to “I-here-now” (Mininni 2003b).

The second question (*why does he/she say that?*) indicates an axis of semiotic pertinence that allows the discourse to articulate arguments and to give voice to the reasons and aims of what is said. Stake markers reveal the aims and interests animating the text; story markers focus on scenes, characters, and models of action; and network markers point out the system of *logoi* and *antilogoi* activated within the multiple narrative and argumentative programs, as we will see in a slide with a semiotic square.

The third question (*how does he/she say that?*) focuses on the articulation of the “*dictum*” and the “*modus*” of discourse according to which the meaning is shaped. Modality can be traced through the following markers:

- (a) Meta-discursive markers, which includes comments and a reformulation index. Meta-discourse refers to the nonpropositional features of discourses that enable listeners or readers to organize the text in a more coherent way and to capture the personality and credibility of the speaker. In particular, *textual* meta-discourse includes logical connectives (“and,” “so,” “as a consequence,” and so on), *frame* markers (“in sum,” “in few words,” “our aim is,” and so on), *endophoric* markers (“aforementioned,” “we will see later”), *evidential* markers (“x claims that . . .”), and *gloss practices* (“that is,” “in other words”). *Interpersonal* meta-discourse is related to the relation between the speakers; it makes use of *weakening* markers (“it could be,” “maybe,” “it’s possible”), *potentiating* markers (“actually,” “clearly”), *attitude* markers (“I agree”), *personal* markers (“I,” “me,” “our”), and *relational* markers (“honestly”) (Crismore et al. 1993).
- (b) *Genre* markers, which include any reference to the kind of text and any intertextual allusions. Diatextual analysis tries to capture how a certain text, subjected to situational restrictions coming from a particular genre, is enclosed in several discursive practices. In other words, in a changing scenario, the same words can transform their meaning.
- (c) Opacity markers: rhetorical figures, metaphors, and so on. The rhetorical analysis of discourses aims to capture the actual positioning of speakers in the universe of beliefs, attitudes, and moral duties in the context of a specific culture.

More specifically, so as to fulfill our research aims, we decided to focus on the reoccurrence of some peculiar discursive markers, such as agency and metaphors, to investigate how cheering culture is shaped through discourse and may generate conflict.

11.3 Results

11.3.1 Analyzing Drughi Culture

Taking into account the definition of organizational culture given by Schein (1992), Drughi culture can be identified by the presence of artifacts, declared values, and shared implicit assumptions.

The main artifact can be represented by their logo: as shown in Fig. 11.2, the picture recalls the four protagonists of the book (and film) *A Clockwork Orange*, a very eloquent symbol of free and cruel violence. The black background and the colors of the Italian flag evoke a nationalistic and extremist political orientation. In addition, as attested by the words of the protagonists of the radio program (e.g., “we will surely win . . . the cheer match”/“*vinceremo sicuramente . . . la partita del tifo*,” episode 2¹), the group worked very hard to carry out the choreographies, thereby manifesting their utmost passion and engagement: for them it seemed like another match to be played.

As for the declared values, the first value of the group is internal cohesion: the fans define a hierarchy in which the football team occupies the last position (e.g., “*first there is the group, the stadium curve, then taken together, Juventus comes*”/“*prima viene il gruppo, la curva, poi insieme viene la Juve*,” episode 1).

The second value is illegality (e.g., “*In particular I greet all the distrusted. Obviously we are close to those who were detained and arrested yesterday [. . .] I always make the same recommendation the same ‘guys, please, don’t steal . . . little’*”/“*in particolare saluto tutti i diffidati. Siamo vicini ovviamente anche ai fermati e agli*



Fig. 11.2 Logo of Drughi culture

¹The four episodes were numbered in accordance with the chronological order of broadcasting.

arrestati di ieri [...] io faccio sempre la stessa raccomandazione 'ragazzi, mi raccomando, non rubate... poco,'” episodes 4 and 1). The third one is the use of violence, also acted out by girls (e.g., “*when I was in Florence there were other girls, the idea of getting into fights came out /quando sono andata a Firenze c'erano delle ragazze, l'idea di fare a botte magari usciva fuori*”, episode 1). Another declared value is pride: the reference to a “mythic” dimension, represented by a specific place (the bench) and time (the Juventus birthday) acts as a reminder of shared roots (e.g., “*Our roots are on that bench, on the 1st of November 1897, we celebrate so many years of history and we have to feel very proud/Le nostre radici sono su quella panchina, con il 1 Novembre del 1897, sono tanti e tanti anni di storia e dobbiamo essere profondamente orgogliosi*”, episode 3).

As shown by the preceding examples, the declared values are discursively constructed both through space- and time-specific references and deictic markers and through more general attitudes, for example, through the present tense in the following extracts.

As for the shared implicit assumptions, implicit references to sexism and racism have been discovered. Sexism is demonstrated by striking the objective features of females: women are just considered for their physical attributes (e.g., “*the single positive feature of that stadium are the beautiful girls we could see on the gallery sometimes/l'unica nota positiva di quello stadio sono qualche gnocca che ogni tanto si vede in tribuna,*” episode 2). Racism is represented in several forms, for example, through a north–south opposition (e.g., “*He's not Italian; he's from Bari [...] we do not know yet if he plays as an Italian or as a non-European player/lui non è italiano, lui è di Bari [...] ancora non si sa è se giochi come oriundo o come extra comunitario,*” episode 1). This is rather unexpected, as one of the protagonists of the program comes from southern Italy. Thus, this could be explained as a strategy to simply depersonalize the player.

Another value is related to the “sacred”: this attribute is given not only to the cheer – several times the protagonists joke about divorcing their wives – but also to places, for example, the stadium (e.g., “*they turned out to lose the inviolability of the stadium/sono riusciti a perdere l'invulnerabilità dello stadio*”, episode 1).

A particular value is the “code of silence,” as shown in the following example:

Extract 1:

A: *Somebody talks about the choirs that came from the north curve, so (...)*

B: *I haven't heard them*

A: *But they probably just made them by themselves!*” (episode 3)

Estratto 1:

A: *Si parla di cori che provenivano dalla nord, quindi (...)*

B: *io non li ho sentiti*

A: *ma se li saranno fatti da soli!* (episodio n. 3)

Here in the first line the new was presented as uncertain, so that B could deny it and A, in this escalation, could even reverse the situation.

11.3.2 *Sketching the Ingroup and Outgroup Identity*

Going beyond the values, how were both ingroup and outgroup identities discursively constructed and conveyed?

As expected, an analysis of the corpus of data collected has revealed a marked use of embrayage markers dealing with the group. The speakers often make use of the first-person plural, and the reference is to Drughì by default, so that when the first-person plural is used to refer to another ingroup identity, the speaker needs to better specify the interlocutor (e.g., “*Because we have told the truth and we have defined Zeman as a “disabled,” we have received 15,000 € for a fine, eh? By “We” I mean as society/Noi per aver dato, detto la verità, dato dell’«incapace» a Zeman abbiamo preso quindici mila euro di multa eh?! «Noi» intendo come società*”, episode 1).

The construction of embrayage is also conveyed by the use of a particular vocabulary, such as, for instance, nicknames, that have contributed to creating a familiar and informal atmosphere (e.g., “*Mimmo the grandfather/Mimmo, il nonno, “Little Fabio/Fabietto”*”).

In line with the most traditional social categorization, research studies (e.g., the magisterial examples by Tajfel, Doise, Deschamps) underlining that while comparing groups’ actions the ingroup tends to activate a need for a positive specification of the self as compared with the outgroup, the data collected confirmed this peculiarity. The focus of “power” on the ingroup showing the need for a positive image of the self dictated a bias against the outgroup as observed in the minimal groups paradigm. More specifically, so as to mark this positive conception, speakers appeal to their magical influence on players’ performance. Discursively speaking, the occurrence of specific rhetorical strategies of celebration, such as, for instance, omnipotence (e.g., “*There are no limits for us white-blacks/Non ci sono limiti per noi bianconeri,*” episode 2), self-glorification (e.g., “*Let’s clap our hands to all of the south curve, since they took part in an extraordinary way so that everything would go well, but then the photo in the newspaper doesn’t do justice to all the work that has been done/bisogna fare un applauso sempre a tutta la curva sud che ha partecipato veramente in maniera straordinaria affinché tutto andasse bene, poi la foto sul giornale non dà giustizia a tutto il lavoro che è stato fatto,*” episode 3), magical influence (e.g., “*when the players entered the stadium, they absolutely had to have a visual adrenaline rush, and the eleven played a perfect match/quando i giocatori entravano in campo dovevano assolutamente avere questa botta visiva che ha dato poi l’adrenalina e gli undici hanno poi dimostrato in campo e hanno fatto una partita perfetta,*” episode 2), and mythopoesis (e.g., “*Gigi Buffon said . . ./Gigi Buffon ha detto . . .*” “*as the President said/come ha detto il Presidente*”) aimed at building up the ingroup as the best.

As for the outgroup, in opposition with what observed with reference to the ingroup, they have been mostly presented by use of débrayage strategies, for instance through the shift from the second to the third plural person (thus aiming at discursively “labeling” the speakers) and, again, to the second one (e.g., “*you have to know that the history of Turin is also made up of episodes, as they say, glorious,*

but remember above all the last twenty years of your history. You're virgins and immaculate/Allora sappiate che la storia del Torino è fatta sì anche di episodi come dicono loro gloriosi, ma ricordatevi soprattutto gli ultimi vent'anni della vostra storia. Siete vergini e immacolati" (episode 4).

Another strategy is delegitimation aimed at ridiculing (e.g., *"when Torino won the last derby, there were no personal computers. It was the age of the club/quando il Torino vinse l'ultimo derby non c'erano assolutamente neanche i computer. Era l'età della clava,*" episode 4) and infrahumanizing (e.g., *"the next time it will be necessary to take a little drawing/la prossima volta bisognerà fare un disegno,"* episode 3). Most times, the speakers use very vulgar and aggressive words (e.g., *"when Zanetti kicks the bucket/quando Zanetti creperà,"* episode 4), mostly made up of offensive epithets (e.g., *"rotten teeth Moratti/denti marci Moratti"* and *"Cassano the snarler/Cassano il ciccione,"* episode 1).

As discussed earlier, infrahumanization could be seen as a softer version of dehumanization: in a comparison between outgroup and ingroup, the members of the former are "deprived of complete humanness by attributing to them fewer uniquely human characteristics, such as secondary emotions (e.g., love, contempt) than to ingroup members" (Pereira et al. 2009, 336).

In the texts analyzed, as already highlighted by some examples, dehumanization took mainly the following forms:

- *Animalization*. This strategy is used by comparing the outgroup supporters with animals, most specifically with mice and pigs, who live in the stadium as their natural context (e.g., *"just four people could go rid the field of the mice [. . .] they can stay in their natural context, since they are pigs, aren't they?/potrebbero andare in quattro a pulire il campo dei sorci [. . .] per farle stare nel loro contesto naturale visto che sono dei maiali no?"*, episode n. 3).
- *Biologization*. Such metaphors were applied not simply to the outgroup but also to their acts, in this case their choreography (e.g., *"They made a huge choreography and we couldn't understand if it was menstrual blood, if it was a big ass, if it was a very big hemorrhoid/Hanno fatto un'enorme coreografia e noi non riuscivamo a capire se si trattava di sangue mestruale, se si trattava di grande culo, se si trattava di una grossissima emorroide,"* episode 4).
- *Mechanization*. This strategy aims at depicting the outgroup as "poor" just because of the colors they support (e.g., *"these poor devils [. . .] this despicable gang [. . .] these poor creatures in black and blue stripes/questi poveri diavoli [. . .] questa banda spregevole [. . .] queste povere creature a strisce nere e blu,"* episode 1).
- *Objectification*. This strategy presents "others" as vegetables to be cut off (e.g., *"the weed has to be cut off from the roots, listen to me/le erbacce vanno tolte alla radice, datemi retta,"* episode n. 4).
- *Demonization*. This stylistic option could be used both in relation to a single person or to a whole group (e.g., *"Cassano the obese/Cassano il ciccione,"* episode 1).

11.3.3 *Celebrating Victories and Exorcising Defeats*

When celebrating victory, the discourses are based on rhetorical strategies that aim to celebrate the ingroup: this means that the victory of the football team is also a victory outside the stadium.

The argumentative strategies that mark the victory “outside the stadium” and underline the strength of the ingroup are represented by ostentation (e.g., “*We beg forgiveness of the championship for killing it a thousand times/Chiediamo perdono al campionato perché l’abbiamo ammazzato ripetutamente,*” episode 4) and self-glorification (e.g., “*We can be really the sixth gear, that x factor that gives players the chance to get to the ball even when you think you cannot/ noi riusciamo ad essere veramente quella marcia in più, quel qualcosa in più che da la possibilità ai giocatori di raggiungere la palla anche quando pensi di non farcela,*” episode 2). Such options aim to convey the idea that cheer groups are fundamental in contributing to the victories of the team on the battlefield. In the following example, the metaphors of killing, of teachings, and of the motor emphasize the absolute, dangerous, and unreachable qualities of the ingroup.

When the radio program concentrates on the team’s defeat, arguing about why it lost, the most frequent argumentative strategy adopted by the speakers is to give a rational justification, thereby limiting the importance of the event. At the same time, they seem to be looking for a scapegoat: as a consequence, an internal distinction is made between “true” fans and “occasional” ones. After the defeat, many references hint at fear and worry, emotions that are not contemplated in victory situations.

The various argumentative strategies used by the speakers are aimed at justifying the defeat as a natural step in the football season and then as nothing to worry about; they also make use of irony to minimize the event, belittling the other teams and asserting their team’s dominance by appealing to its magnanimity; to minimize the defeat, the supporters use the argumentative strategy of authorization because they make an appeal to the established and strong tradition of the team, underlying the fact that its value remains intact despite the trivial defeat.

In relation to the two episodes dealing with defeat, the need for justification is more evident following the match against Inter, whereas in the case of Milan discourse is focused more on the future and on efforts to move beyond the situation (e.g., “*On Saturday we’re going to get a ton of goals, we’ll be happy and we’ll have forgotten how we lost yesterday/Sabato poi ehm riusciremo a fare una goleada, come insomma da pronostico ehm saremo tutti ben felici e avremo già dimenticato di come abbiamo perso ieri,*” episode 2). In the Italian championship, even if there is a really strong rivalry with both Milan and Inter, the match against Inter is commonly defined as the “Italian derby.” This might be why a loss to that team is harder to take and, as a consequence, requires much more targeted argumentative strategies and rational argumentation, whereas Milan is one of the steadiest teams, so defeat is accepted as something that can happen.

11.3.4 *Metaphors of Cheer Group Identity*

A very important piece of the rhetorical construction of the Drughì identity is formed by the use of analogical frames, as vivid examples of how speakers construct ingroup and outgroup identities. Different rhetorical scripts hint at different domains, ranging from the classic weather metaphors – recalling points in the rankings (e.g., “*A cold snap has hit all of Italy, look!/un'ondata di gelo in tutta Italia, guarda [. . .] Milan minus 18 not to mention Inter. You know neither weather forecast mentions it/Milan meno diciotto ma senza parlare dell'Inter ormai. Sai che non lo mettiamo nemmeno più nelle previsioni del tempo,*” episode 4) – to the culinary world (e.g., “*They've got three real pears/abbiamo rifilato tre autentiche pere,*” episode 2), from the body (e.g., “*The South (curve) gives us emotions, it is the beating heart/la Sud regala emozioni è in particolare il cuore pulsante,*” episode 2) to the animal domain (e.g., “*we are 11 lions/noi siamo 11 leoni,*” episode 2), up to the most classic war reference (e.g., “*It seemed like war, they've taken care of each corner where there could have been danger/Sembrava di essere in guerra, hanno presidiato tutti gli angoli da dove poteva venire qualche pericolo,*” episode 4). This evidence confirms the importance of the pathemic side of discourse in the domain of soccer cheers, in order to involve and commit the ingroup, as well as to get its members excited.

11.4 **Concluding Remarks**

The results of this explorative study allow us to draw some interesting conclusions and to investigate a quite new field of research. Actually, the peculiar typology of participants involved has made it possible to investigate the nature of cheer culture and the core of its basic assumptions. As a “lifestyle” sports help to organize the aggressive and destructive motivations that are built into human sociality and redirect them into a framework of competition, aiming at self-acknowledgement in the public discursive sphere. Nonetheless, the renegotiation of the meaning attached to conflict that is acted out in sports does not eliminate the use of strategies that seek to inhumanize the enemy; rather, it frames them within a discursive regime that is aimed at mere spectacularization. Such a matrix of meaning tends to balance ethical weaknesses with aesthetic bravery.

In accordance with the transversal sensitivity offered by cultural-discursive psychology as well as with the main psychological approaches dealing with social identity theory, the organizational culture and the construction of relations between ingroup and outgroup, the images of identities and relations are (co-)constructed, shared, and negotiated in a discursive and rhetorical way. Starting from the identification with artifacts, references to declared values, and shared implicit assumptions, the protagonists of the radio program “Three stars on the skin” construct a “hard” ingroup identity. The extreme and masculine values, the current use of embrayage

markers – in particular the first-person plural – the emotionally charged lexicon, and the argumentative strategies converge in the picture of the ingroup as exclusive, as the best one and as an object of true devotion. In opposition, as expected, the outgroup is denoted by *débrayage* markers, *infrahumanizing* and *dehumanizing* both lexical and argumentative choices. In addition, the metaphors, which derive from several domains, contribute to the construction of a strong pathemic impact, to psych up the audience, and to make the radio program an interlocutory diatext.

Furthermore, the present study leaves us with some questions that will require further research, involving, for instance, other cheer groups belonging to other domains of sports or other football teams. Moreover, we need to understand the difference in cheering between these kinds of groups and nonorganized cheer participation. Finally, it would be interesting to compare how media depict and represent these groups and how the groups perceive themselves.

In view of the above our conclusion and future research agenda concerns the role of psychology in this frame. Actually sport psychology may mediate conflicts by helping designing first of all communication campaigns which will be aimed at diffusing a non violent culture of sport and team support. The two hearts of cheer: the organized and the non organized components should communicate more and better since they have the same objective that is to support their own team. Then specific training interventions could be designed to make these cheer groups aware about the identity of the outgroup whose presence is not a challenge to one's own rather it is an interlocutor, the other who legitimizes the self as dialogical interpretation of communication.

References

- Allport F (1924) *Social psychology*. Houghton Mifflin, Boston, MA
- Allport GW (1954) *The nature of prejudice*. Addison-Wesley, New York
- Bruner J (1990) *Acts of meaning*. Cambridge University Press, Cambridge
- Cole M (1998) Can cultural psychology help us think about diversity? *Mind Cult Act* 5(4):291–304
- Caffi C, Janney RJ (1994) Toward a pragmatic of emotive communication. *J Pragmat* 22:326–360
- Crismore A, Markannen R, Steffenson M (1993) Metadiscourse in persuasive writing: a study of texts written in American and Finnish University students. *Writ Commun* 10(1):39–71
- Drury J, Reicher S (2000) Collective action and psychological change: the emergence of new social identities. *Br J Soc Psychol* 39(4):579–604
- Dunning E (1994) The social roots of football hooliganism. A reply to the critics of the Leicester school. In: Giulianotti R, Bonney N, Hepworth M (eds) *Football, violence and social identity*. Routledge, London, pp 128–157
- Dunning E, Murphy P, Waddington I (1991) Antropological versus sociological approaches to the study of soccer hooliganism: some critical notes. *Sociol Rev* 39(3):459–478
- Dunning E, Murphy P, Williams J (1988) *The roots of football hooliganism: an historical and sociological study*. Routledge, London
- Giulianotti R (1991) Scotland's Tartan Army in Italy: the case for the Carnavalesque'. *Sociol Rev* 39(3):503–527
- Giulianotti R, Robertson R (2004) The globalization of football: a study in the glocalization of the 'serious life'. *Br J Sociol* 55(4):545–568

- Harré R, Gillet G (1994) *The discursive mind*. Sage, London
- Jahoda G (1999) *Images of savages: ancient roots of modern prejudice in western culture*. Routledge, London
- Kelman HC (1973) Violence without moral restraint: reflection on the dehumanization of victims and victimizers. *J Soc Issue* 29(4):25–61
- Kerr J (1994) *Understanding soccer Hooliganism*. Open University Press, Buckingham
- Kohl D (2011) The presentation of “Self” and “Other” in Nazi Propaganda. *Psychol Soc* 4(1):7–26
- Le Bon G (1895) *The crowd. A study of the popular mind* (trans. 1947). Sparkling Books, New York
- Le Bras-Chopard A (2000) *Le zoo des philosophes. De la bestialisation à l’exclusion*. Plon, Paris
- Mininni G (1992) *Diatesti. Per una psicosemiotica della vita quotidiana*. Liguori, Napoli
- Mininni G (1995) *Discorsiva mente. Profilo di Psicosemiotica*. Edizioni ScientificheItaliane, Napoli
- Mininni G (2003a) *Il discorso come forma di vita*. Guida, Napoli
- Mininni G (2003b) L’approccio psicosemiotico: testi e immagini. In: Mantovani G, Spagnoli A (eds) *Metodi qualitativi in psicologia sociale*. Il Mulino, Bologna, pp 159–198
- Mininni G (2011) *Teste e Testi*. Aracne, Roma
- Mininni G (2013) *Psicologia culturale e discorsiva*. Franco Angeli, Roma
- O’Sullivan Lago R (2011) The possibility of resolving conflict through dialogue? Continuing from power. *Psychol Soc* 4(1):67–70
- Opatow S (1990) Moral exclusion and injustice: an introduction. *J Soc Issue* 46:1–20
- Pereira C, Vala J, Leyens JP (2009) From infra-humanization to discrimination: the mediation of symbolic threat needs egalitarian norms. *J Exp Soc Psychol* 45:336–344
- Potter J, Reicher S (1987) Discourses of community and conflict: the organization of social categories in accounts of a ‘riot’. *Br J Soc Psychol* 26:25–40
- Power S, Peterson L (2011) Theorizing propaganda: extending Kohl. *Psychol Soc* 4(1):27–30
- Power SA (2011) Towards a dialogical model of conflict resolution. *Psychol Soc* 4(1):53–66
- Reicher S (1984) Social influence in the crowd: attitudinal and behavioral effects of deindividuation in conditions of high and low group salience. *Br J Soc Psychol* 23(4):341–350
- Reicher S (1987) Crowd behavior as social action. In: Turner J, Hogg M, Oakes P, Reicher S, Wetherell M (eds) *Rediscovering the social group*. Blackwell, Oxford
- Reicher S (1996) The ‘battle of Westminster’: developing the social identity model of crowd behaviour in order to explain the initiation and development of collective conflict. *Eur J Soc Psychol* 26:115–134
- Savage R (2007) “Disease incarnate”: biopolitical discourse and genocidal dehumanisation in the age of modernity. *J Hist Sociol* 20(3):404–440
- Schein EH (1992) *Organizational culture and leadership*, 2nd edn. Jossey-Bass, San Francisco, CA
- Stott C, Reicher S (1998) How conflict escalates: the inter-group dynamics of collective football crowd ‘violence’. *Sociology* 32(2):353–377
- Stott C, Hutchinson P, Drury J (2001) Hooligans’ abroad? Inter-group dynamics, social identity and participation in collective ‘disorder’ at the 1998 world cup finals. *Br J Soc Psychol* 40(3):359–384
- Suttles G (1968) *The social order of the slum: ethnicity and territory in the inner city*. University of Chicago Press, Chicago, IL
- Theiner G, Allen C, Goldstone R (2010) Recognizing group cognition. *Cogn Syst Res* 11(4):378–395
- Tileagă C (2007) Ideologies of moral exclusion: a critical discursive reframing of depersonalization, delegitimization and dehumanization. *Br J Soc Psychol* 46(4):717–737
- Tileagă C (2012) The right measure of guilt: moral reasoning, transgression and the social construction of moral meanings. *Discourse Commun* 6(2):203–222
- Turner JC, Hogg M, Oakes P, Reicher S, Wetherell M (1987) *Rediscovering the social group: a self-categorization theory*. Basil Blackwell, Oxford
- Volpato C (2011) *Deumanizzazione. Come si legittima la violenza*. Laterza, Roma-Bari

Volpato C (2012) La negazione dell'umanità: i percorsi della deumanizzazione. *Riv Int Filos Psicol* 3(1):96–109

Wann DL, Grieve FG (2005) Biased evaluations of in-group and out-group spectator behavior at sporting events: the importance of team identification and threats to social identity. *J Soc Psychol* 145:531–545

Chapter 12

Some Puzzles of Politeness and Impoliteness Within a Formal Semantics of Offensive Language

Carl Vogel

12.1 Introduction

Some forms of conflict involve participants intending or experiencing psychological offence, as opposed to more physically impinging threat. Offensiveness may be modulated by behaviors of politeness and impoliteness. An upsetting message may be tempered by polite delivery, or a message that is literally positive may be reversed through impolite conveyance (see D’Errico and Poggi 2014). Among the features that comprise delivery are forms of linguistic packaging. The means, cross-linguistically, of creating emphasis within propositional content of a message, partly with reference to information states of interlocutors, have been discussed as “information packaging” (Vallduví and Engdahl 1996). Herein, attention is drawn to the fact that attitudinal orientation to both message content and interlocutors is also potentially packaged linguistically.

Linguistic manifestations of politeness and impoliteness present some puzzles about human behavior. A number of these puzzles are detailed in the next part of this paper. However, the main puzzle of linguistic politeness and impoliteness is an existential one. That these linguistic flourishes exist at all is at odds with other linguistic behaviors which tend to minimize communicative effort.

Aspects of linguistic politeness and impoliteness are well explained from the point of view of semantic theory, and may ultimately be integrated into a multi-dimensional account of meaning in the context of interaction (Bunt et al. 2012). Such an approach may, as here, be sympathetic to the idea that perceptions of (im)politeness may differ among interlocutors in particular situations (Davies et al. 2011). However, perspectivalism is not obviously more necessary in accounts of

C. Vogel (✉)

Computational Linguistics Group, Centre for Computing and Language Studies,
School of Computer Science and Statistics, Trinity College Dublin, Dublin 2, Ireland
e-mail: vogel@tcd.ie

(im)politeness than of propositional content, for which it is also necessary (Healey and Vogel 1994). This chapter details an attempt to use a semantic analysis of linguistic (im)politeness to explain some of the puzzles of (im)politeness. A number are resolved by noting the semantic function of politeness in signaling attitudinal orientation.

The semantic theory of impoliteness described in the third part of this paper is intended to partially specify what is meant by expressions regarded as polite or impolite. A truth-conditional approach to semantics articulates the constraints that the world has to satisfy in order for relevant expressions to be correctly interpreted as true or false, polite or impolite. Because of the situated element of (im)politeness, articulating the constraints on the world which have to be in place in order for an utterance of a sentence to be evaluated as, for example, true and impolite also provides a reasonable characterization of what makes non-linguistic behaviors in the same situations also be deemed impolite, and therefore a semantic theory of the sort described here provides an effective interface to pragmatic theory, as well. Addressing some of the puzzles of politeness and impoliteness through the semantic theory provides an evaluation of the theory.

The semantic theory invoked here is informed by the view that politeness and impoliteness behaviors are manifestations of offence management, offence rooted in disgust (Vogel 2014a,b).¹ This etiological account of politeness and impoliteness has the advantage of offering an explanation for the existential puzzle of linguistic (im)politeness: the energy required for these locutions is significant, and unlike other forms of complexity in language, for example, referential descriptions, which become less involved and more phonologically reduced with each reference in conversation, the language of (im)politeness does not appear to undergo reduction on the same scale. The explanation, which anchors (im)politeness in disgust management, holds that it is the desire not to be seen as disgusting which prompts politeness and the perception of disgust which draws out impoliteness. This theory is feasible because of the nature of the human disgust response: disgust triggers are generally felt as irreversibly polluting²; the response, which has physical oral-facial realizations connected to expulsion reflexes, includes subsequent avoidance even after a single exposure; the disgust response generalizes from primary sources to associates in relatively unfettered contagion; associates may be conceptual as well as tangible. Because the disgust response, once triggered by an encounter, leads to

¹This is an evolutionary account that anchors politeness in selective processes associated with disgust. Alternative accounts, from the perspective of socio-cultural development, also exist (Bax 2012).

²Attempts to reverse pollution often involve symbolic treatments that are more vigorous than any physical cleansing that would be strictly necessary.

immediate and future avoidance,³ and since people generally dread ostracism, it is in the interest of social agents to avoid triggering disgust. Politeness mechanisms are therefore adaptive mechanisms that mitigate disgust. Impoliteness is explained as well by the experience of (generalized) disgust triggered by the target of impoliteness and the desire the impolite speaker has for this view of the target to be shared by witnesses.⁴ The persistent effort involved in the language of (im)politeness makes sense in light of the profound power of the disgust response.

The paper proceeds by describing some of the other puzzles of linguistic (im)politeness (Sect. 12.2). In Sect. 12.3 a formal semantic theory of (im)politeness is detailed, and Sect. 12.4 addresses the puzzles within this framework. The semantics makes use of events as ontological primitives, but assumes that these can be examined at varying levels of granularity. That is, the semantics does not require an extensive ontological inventory, but rather specifies constraints that must hold on events that figure into acts of (im)politeness, linguistic or otherwise.

12.2 Puzzles of (Im)politeness

(Im)politeness behaviors in general present some puzzles: as indicated above, the biggest puzzle is that linguistic politeness is so persistently ornate. An explanation has already been provided for this, and the framework for semantic analysis of (im)politeness that arises from formalizing aspects of offence management can also elucidate other curiosities surrounding (im)politeness. The issues described in this section are possibly surprising in contrast to other aspects of language or first glances at how the semantics of politeness might function.

12.2.1 *Linguistic or Extra-Linguistic*

One puzzle is whether (im)politeness is a feature of language at all, or rather only a feature of behavior more generally. To focus on linguistic (im)politeness is to attend to dimensions of offence mitigation and accentuation that are achieved through language, apart from dimensions of offence management that are conducted via other channels of communication. Some have essentially argued that there is no place for a linguistic theory of (im)politeness, since perceptions of such are subjectively reached and given the claim that linguistic forms do not have politeness

³The effects of proximity have been argued to discriminate disgust and fear (Kolnai 1929a): separation by a small distance and secure barriers mitigates fear, but with the same intervening distance and iron bars, disgust is not diminished.

⁴Impoliteness contains the risk for the speaker that the act will be viewed by witnesses without sympathy for the speaker's view but in favor of the perspective of the impoliteness target.

or impoliteness inherently in their meaning (Watts 2003). Others have argued that (im)politeness is inherent in the meaning of some forms of language (Culpeper 2011, p. 118). That an expression is ambiguous or conveys layers of meaning is not sufficient to dismiss the importance of semantics in the interpretation of (im)polite language, just as semantic theory provides critical illumination of the interpretation of metaphorical language (see Van Genabith 2001; Vogel 2001).

12.2.2 *Excess*

For example, although politeness is a positive behavior by default, excessive politeness, obsequiousness, is not esteemed (see (12.1))⁵; thus, politeness is not monotonically additive.

- (1) Please, allow me to most humbly and with supreme deference offer to your superlatively splendid self the information that the time is 12:30.

12.2.3 *Circular Spectrum*

It has been noted that in some contexts, language that might otherwise be understood as rude is actually part of the normal code for interaction (Bousfield 2008), and is therefore less reasonable to understand as impolite. Additionally, in some of the contexts that include interactants who are exceedingly familiar (such as family), it can be deemed impolite to use the language of politeness: an example like (12.2.a), in contrast to (12.2.b), may well be less appropriate for communication with one's parents than for a colleague or client.

- (2) a. Please, let me know what time you'd like me to arrive.
- b. Let me know what time you'd like me to arrive.

Note that this is distinct from the possibility of mock (im)politeness, in which an ironic interpretation is forced. Rather, the same utterance, without ironic intent, may obtain a politeness–polarity switch depending on the context. Therefore, while it is natural to think of a mono-dimensional scale between politeness and impoliteness, it is worth considering whether a more appropriate conceptualization is not a spectrum, but a color wheel.⁶

⁵Conversely, impoliteness, when used as a tool of iconoclasm, may earn esteem (Van Kleef et al. 2011), in the sense that comics like Groucho Marx are appreciated. This may be connected to the macabre attraction that disgust triggers have: as pointed out by others (Kolnai 1929a), the oral-facial expulsion reflex on the verge of vomiting in response to a disgust trigger proceeds *as if* the substance had been ingested, even if it has not—the disgust response requires imagining contact.

⁶Perhaps a color spindle is a more appropriate conceptualization (cf. Gärdenfors 2000).

12.2.4 *Saussureanism*

Saussurean communicators use language with the same meaning-form mapping for interpretation as for production, and it seems that natural language in general is used in this bi-directional way. This is not a necessary truth about language, but one that could have evolved differently. With evolutionary models, it has been shown that Saussureans “win” over imitators and calculators under relevant initial-state assumptions (Hurford 1989). Nonetheless, people are not manifestly Saussurean with respect to offence and offensive language. A behavior may offend the speaker when produced by others, but the speaker may not intend nor notice offence felt by others when the speaker produces the same behavior.⁷ Thus, it seems that the language of (im)politeness is not completely Saussurean.

12.2.5 *Learning*

The poverty of the stimulus argument for the innateness of aspects of natural language syntax is based on observations surrounding language learning, which occurs quickly during development and in the absence of explicit language teaching and despite a relative dearth of negative examples.⁸ In contrast, the language of politeness appears to be explicitly taught (Greif and Gleason 1980; Gleason et al. 1984) and slowly learned. Given the arguments made above attempting to explain the existence of the language of (im)politeness through a need for disgust management in social settings, coupled with one-shot learning from any immediate encounter with a disgust trigger, one might expect the opposite, that linguistic (im)politeness would be learned quite quickly. However, the disgust response generalizes rapidly from the trigger to associates of the trigger, and it is in this period of association that “accidental” triggers emerge. Thus, the systematic behaviors

⁷See Babrius’ “The Two Wallets”:

Prometheus was a god, but of the first dynasty. He it was, they say, that fashioned man from earth, to be the master of the beasts. On man he hung, the story goes, two wallets filled with the faults of human kind; the one in front contained the faults of other men, the one behind the bearer’s own, and this was the larger wallet. That’s why it seems to me, men see the failings of each other very clearly, while unaware of those which are their own.

(Fable 66, 1–8, p82. Babrius and Phaedrus: Fables, trans. Ben Edwin Perry) 1965 Loeb Classical Library: Harvard University Press.

⁸It is bolstered by results from formal language theory which show that in the absence of negative examples, an infinite context-free languages cannot be learned “in the limit” (Gold 1967), thus lending support to the notion that some of the structures within language be innate rather than learned.

within a culture condition common associates, and therefore cultural specificity in disgust triggers, and, perhaps surprisingly, relatively slower assimilation of the resulting conventions.

12.2.6 Synonymy

While acknowledging that humans have a deep-seated urge to avoid synonymy in natural language and create discriminations in use conditions where none necessarily existed before, in order to avoid full truth-preserving and use-preserving intersubstitutability of expressions, one might use the struggle to differentiate senses as evidence of initial synonymy in language. People are not reliable in their judgements of relative (im)politeness of various expressions available to modulate the manner of presentation.⁹ Thus, it is possible to find cases in which (12.3.a) is not fully intersubstitutable with (12.3.b), and similarly for (12.4.a) and (12.4.b); however, it will be difficult to obtain reliable judgements for either pair that one is more (im)polite than the other. Where such expressions are not reliably ranked in (im)politeness by the same individual on different occasions, then one has an argument that relative to (im)politeness considerations they are synonymous expressions.

- (3) a. May I move by, please.
 b. Excuse me, please.
- (4) a. You're in my way.
 b. Move.

That equivalence classes of (im)politeness are available, resulting in synonymy, is compatible with the notion that gradedness is also available, and that some expressions are reliably and robustly deemed more polite than others. Extremes in the gradient are associated with taboo words. Interesting support for the notion of a circle of gradience (see Sect. 12.2.3) is in the fact that labels associated with extremely positive and negative concepts are both subject to taboo.

12.2.7 Taboo Objects Have Exceptional Anaphoric Potential

Within the discussion of taboo topics, it has been noticed long ago that suppressed objects of verbs of excretion are available as antecedents to anaphora, although

⁹Similar facts obtain with respect to quantifying determiners in natural language (Moxey and Sanford 1993).

suppressed objects of other verbs are not accessible (Tic Douloureux 1992).¹⁰ Contrast the potential antecedents (indicated with superscripts) of the pronoun *it* in the examples (12.5) and (12.6).¹¹

- (5) After the gun fired, it^{the gun/#the bullet} dropped to the ground.
 (6) After the dog defecated, it^{#the dog/the excrement} bore the missing ring.

The imagination is evidently more robust in constructing an unmentioned antecedent in the situation normally associated with disgust. This is consistent with the activity of imagination that is evidently rapid and vigorous in the human disgust response (see footnote 5) since the physical reactions proceed as if contact had occurred, even when they have not.¹²

12.2.8 *Metaphorical, Yet Metaphorically Inert*

(Im)politeness expressions may be non-literal, as in the examples of (12.7), but in yet another asymmetry between the language of politeness and impoliteness, politeness expressions are seldom themselves metaphorical vehicles. Compare (12.9.a) and (12.9.b): in these examples, expressions of (im)politeness (12.9.a) and (12.9.b) are used to construct non-literal sentences about how easy the new hires found use of the photocopier. While with sufficient effort, nearly every expression may be used in a novel, non-literal way, it appears easier to use the language of impoliteness to construct a metaphor (12.9.b) than it is to do so with the language of politeness (12.9.a). Thinking of the accessibility of metaphorical interpretation as akin to accessibility of antecedents to anaphors, this situation appears to be the complement of the blocking of the accessibility of indefinite noun-phrases as antecedents to pronouns by negation, as in the contrast between (12.10.a) and (12.10.b) (Kamp and Reyle 1993). In the case of accessibility of metaphorical interpretation in (12.9), the “negative” expression (impoliteness) allows the link while politeness blocks it.

- (7) a. Please, make yourself at home.
 b. Buzz off, you.
 (8) a. That manager/photocopier cradles new hires.
 b. That manager/photocopier vulcanizes new hires.

¹⁰This is akin to the claimed asymmetry in the quantity of labels for negative emotions in relation to named positive emotions (Kolnai 1929b).

¹¹The prefixed symbol (#) is used to indicate semantic infelicity.

¹²This does not include disgust reaction to smells, since smells do entail indirect ingestion of the disgust trigger.

- (9) a. #The new hires were made at home by the photocopier.
 b. The new hires were buzzed off by the photocopier.
- (10)a. I own a bicycle. It is green.
 b. I do not own a bicycle. #It is green.

12.2.9 *Impoliteness May Be Reflexive, But Not Politeness*

While accepting that the primary use of language is within thought, rather than for communication, one must concede that linguistic politeness is one dimension of language use that is quintessentially communicative. Because other agents are not involved in thinking for oneself, as opposed to thinking for speaking,¹³ pragmatics principles such as in (12.11)¹⁴ do not have force.

- (11)a. Putting oneself before others is impolite.
 b. Putting others before oneself is polite.

Thus, manifestations of (im)politeness expressions within thought for oneself are surprising. However, the greater curiosity is that private thought appears to asymmetrically favor impoliteness over politeness. I personally can easily enough imagine having thoughts that express negative politeness (12.12), but I am extremely unlikely to entertain a reflexive thought with positive politeness (12.13).

- (12)Idiot_{*i*}, why did I_{*i*} let the toast burn?
 (13)#Carl_{*i*}, please *t_i* enjoy this toast I_{*i*}'ve made.

12.2.10 *Summary*

A number of puzzles of (im)politeness have been indicated. Among these, an asymmetry is evident in which some possibilities are available within the language of impoliteness that lack counterparts among expressions of politeness. In analyzing these puzzles further, it is necessary to have a clear view of the semantics of linguistic forms that may be deployed to achieve (im)politeness effects. Accordingly, a formal semantics has recently been proposed for the analysis of linguistic (im)politeness (Vogel 2014a). The next section (Sect. 12.3) presents the core of this theory, and the section that follows the next (Sect. 12.4) demonstrates how the semantics makes sense of some of the puzzles of (im)politeness behaviors.

¹³See MacNeill (1997).

¹⁴See Vogel (2014a).

12.3 Semantics of (Im)politeness

The theory outlined in this section is developed more extensively elsewhere (Vogel 2014a). This presentation draws upon that account.

12.3.1 *The Semantic Model*

The denotations of linguistic expressions of politeness and impoliteness are events that are constrained by predications of relative offence as experienced by the speaker of those expressions in relation to the participants in the triggering events.¹⁵ Treating such events, which are accompanied by relations among their participants, as the denotation of expressions of (im)politeness allows for semantic and pragmatic analysis to make reference to a shared ontology. Pragmatic inference about an utterance will also make reference to the events that are in the denotation of the utterance.

It is useful to characterize relevant properties of event types as in (12.14), treating events as particulars which instantiate the type.¹⁶ An event token e of the type \hat{e} makes relevant relations among arguments true and has certain other properties as specified. Events have temporality, tense, aspect, and mode (*realis* or *irrealis*). Any event token e will fix values for these features, according to the level of granularity with which it is viewed.¹⁷ Events are taken to have proto-agents and proto-patients (see Dowty 1991). Syntactic person is used to characterize the fillers of these proto-roles, the event participants. Events have *use* and *cost* for each participant. A three-valued polarity system (with 1 representing positive value, 0 representing neutral value, and -1 representing negative value) may be deployed within simple calculus of use and cost to establish a *net* value. Table 12.1 provides a possible specification of the combination of values of *use* and *cost* to yield *net* values.

¹⁵See (12.27) and (12.28) for characterization of the sets of events denoted by impoliteness and politeness expressions, respectively. The next paragraphs explain the terms used by those characterizations. Taking denotations of expressions to be sets of events may be compared with treatments of modality that analyze propositions as sets of possible worlds (Kratzer 1981) or with situation theoretic approaches to meaning that take denotations to be sets of supporting situations (Cohen 2009).

¹⁶Feature-value matrices provide an effective visual organization of bundles of first-order descriptions. See Kasper and Rounds (1986) and Carpenter (1992) for details of some feature logics and equations on paths through feature-value structures.

¹⁷Formal models that enable variable granularity in analysis of events exist (Fernando 2013).

Table 12.1 A specification of *net* offence/affinity as a function of *use* and *cost*

Use	Cost		
	1	0	-1
1	1	1	0
0	1	0	-1
-1	0	-1	-1

$$(14) \hat{e} =_{\text{event}} \left[\begin{array}{l} \text{temporality: } \left[\begin{array}{l} \text{tense} \\ \text{aspect} \\ \text{mode} \end{array} \right] \\ \text{proto-agent: } 1\text{st} \vee 2\text{nd} \vee 3\text{rd} \\ \text{proto-patient: } 1\text{st} \vee 2\text{nd} \vee 3\text{rd} \\ \text{use } \left[\begin{array}{l} \text{person: } \text{polarity} \\ 1\text{st: } 1 \vee 0 \vee -1 \\ 2\text{nd: } 1 \vee 0 \vee -1 \\ 3\text{rd: } 1 \vee 0 \vee -1 \end{array} \right] \\ \text{cost } \left[\begin{array}{l} \text{person: } \text{polarity} \\ 1\text{st: } 1 \vee 0 \vee -1 \\ 2\text{nd: } 1 \vee 0 \vee -1 \\ 3\text{rd: } 1 \vee 0 \vee -1 \end{array} \right] \end{array} \right]$$

The relative offence ($\hat{\delta}$) experienced by an agent in response to an event resolves the disjunctions in (12.15).

$$(15) \hat{\delta}(e) =_{\text{attitude}} \left[\begin{array}{l} \text{before-}e: \left[\begin{array}{l} \text{person: } \text{polarity} \\ 1\text{st: } 1 \vee 0 \vee -1 \\ 2\text{nd: } 1 \vee 0 \vee -1 \\ 3\text{rd: } 1 \vee 0 \vee -1 \end{array} \right] \\ \text{during-}e: \left[\begin{array}{l} \text{person: } \text{polarity} \\ 1\text{st: } 1 \vee 0 \vee -1 \\ 2\text{nd: } 1 \vee 0 \vee -1 \\ 3\text{rd: } 1 \vee 0 \vee -1 \end{array} \right] \\ \text{after-}e: \left[\begin{array}{l} \text{person: } \text{polarity} \\ 1\text{st: } 1 \vee 0 \vee -1 \\ 2\text{nd: } 1 \vee 0 \vee -1 \\ 3\text{rd: } 1 \vee 0 \vee -1 \end{array} \right] \end{array} \right]$$

The model allows for sequencing of events and varying attitudes as sequences and events within them unfold. In a null context, speakers are assumed to maintain a default evaluation of the participants as in (12.16), in which by default the speaker has a positive self-evaluation which is equal to the evaluation of

all others.¹⁸ However, a reasonable alternative specification is available according to the temperament being modeled—for example, in (12.17) the speaker has a positive self-evaluation and lesser evaluations of others, but none are considered negatively.

$$(16)_{\text{attitude}} \left[\begin{array}{l} \text{person: polarity} \\ \text{1st:} \quad /1 \boxed{i} \\ \text{2nd:} \quad / \boxed{i} \\ \text{3rd:} \quad / \boxed{i} \end{array} \right]$$

$$(17)_{\text{attitude}} \left[\begin{array}{l} \text{person: polarity} \\ \text{1st:} \quad /1 \\ \text{2nd:} \quad /0 \\ \text{3rd:} \quad /0 \end{array} \right]$$

The propensity for triggers of disgust to spread to associates in a contagion is modeled with the sharing of the minimum value of the offence level as in (12.18).¹⁹ Because this is a worst-case formulation of contagion in which all are deemed contaminated within an event if any are, the principle, even expressed as a conditional, must be understood as dependent on additional parameters not specified here.

(18) Contamination spreads

$$(MIN \{ \boxed{j} \mid \exists \boxed{i}, \hat{\theta}(e): \text{after-}e:\text{person:} \boxed{i}:\text{polarity:} \boxed{j} \} = -1) \implies \\ \forall \boxed{k}, \hat{\theta}(e): \text{after-}e:\text{person:} \boxed{k}:\text{polarity} = -1$$

When extreme values emerge, follow-on reactions by the reflecting agent may be expected, depending on other constraints that are in effect at the time. As appropriate to the levels of offence associated with each participant in an event being commented upon (whether a dialogue event within an ongoing conversation or in the rest of the external world), linguistic behaviors may be anticipated.

12.3.2 Predictions

Predictions about the behavior of a speaker may be made on the basis of further instantiation of $\hat{\theta}$ and the \hat{e} on which it depends. Below, θ -polarity refers to the final value within $\hat{\theta}$ for the corresponding grammatical person or the proto-agent of the triggering event (e). The defaults (12.19)–(12.23) characterize reasonable

¹⁸A formal framework for default feature structures is available (Lascarides et al. 1996); the value to the right of the slash is defeasible. Co-indexing encodes value sharing.

¹⁹This formulation is hopefully more clear than that of Vogel (2014a); however, this one, too, involves polymorphic notation—here, in that the “=” of the antecedent is a test of equality while the one in the consequent is an assignment.

expectations of how a speaker might intend to react to the event e being spoken about in situations that are marked by divergence from the default evaluation of event participants (12.16) (cf. Altmann and Riška 1966).

- (19) If 1st-polarity < 2nd-polarity, expect politeness
- (20) If 1st-polarity > 2nd-polarity, expect impoliteness
- (21) If 1st-polarity < proto-agent(e)-polarity, expect politeness
- (22) If 1st-polarity > proto-agent(e)-polarity, expect impoliteness
- (23) If 1st-polarity = 2nd-polarity, expect politic behavior.

These principles may yield conflicting expectations. The first (12.19) is triggered when the speaker has a greater estimation of the person addressed than self-estimation, and an expectation of politeness is licensed. The penultimate principle (12.22) leads to an expectation of impoliteness that conflicts with the expectation arising from (12.19) when the speaker's self-estimation exceeds the speaker's estimation of the proto-agent of the event being commented upon, as the latter may be a third-party. The result may be a mixed signal towards deference to the addressee and disdain for the third party. The final principle (12.23) conveys the default that only marked behavior is appropriate to label polite or impolite, and therefore the expectation for unmarked situations of equal addresser and addressee polarity is merely *politic* behavior, following the terminology adopted by Watts (2003).

It may be useful to entertain additional principles in relation to extreme values, as in (12.24).

- (24)a. If 2nd-polarity < 0, expect impoliteness.
- b. If 2nd-polarity > 0, expect politeness.

However, given tendencies of pejoration in language change (Borkowska and Kleparski 2007), there may asymmetrically be evidence for (12.24.a) as a robust principle but not (12.24.b). Additionally, the applicability of all these principles may vary with culture, age, gender, personality, context, and so on.²⁰

12.3.3 Interpretation

From (im)polite language corresponding further specification of \hat{o} and \hat{e} may be inferred. Expressions of impoliteness support the inference that the speaker's response to an event e'' is consistent with the conjunction of (12.25.a) and (12.25.b). A neutral comment by the speaker supports the inference of (12.26.a) and (12.26.b).

²⁰On gender, it has been noted that over time a number of words related to women have obtained pejorative connotations with which they did not begin ("hussy" or "wench", for example), but that words related to men have sometimes gained improved connotations (e.g., "knight") (Fitch 2007).

- (25)a. $\hat{\delta}(e^{\clubsuit}) = \text{attitude} \left[\text{after-}e^{\clubsuit}: \left[\frac{\text{person: polarity}}{1\text{st: } > -1} \right] \right]$
 b. $\hat{\delta}(e^{\clubsuit}): \text{after-}e^{\clubsuit}: \text{person: 1st: polarity} > \hat{\delta}(e^{\clubsuit}): \text{after-}e^{\clubsuit}: \text{person: 2nd: polarity}$
- (26)a. $\hat{\delta}(e^{\dagger}) = \text{attitude} \left[\text{after-}e^{\dagger}: \left[\frac{\text{person: polarity}}{1\text{st: } 1 \vee 0 \vee -1} \right] \right]$
 b. $\hat{\delta}(e^{\dagger}): \text{after-}e^{\dagger}: \text{person: 1st: polarity} = \hat{\delta}(e^{\dagger}): \text{after-}e^{\dagger}: \text{person: 2nd: polarity}$

More generally, epithets of impoliteness may be taken to denote sets of events in which the speaker's ultimate self-estimation exceeds the speaker's estimation of others (12.27). In contrast, epithets of politeness express sets of events constrained by the opposed relation (12.28).

(27) $\lambda e. [\hat{\delta}(e): \text{after-}e: \text{person: 1st: polarity} > \hat{\delta}(e): \text{after-}e: \text{person: 2nd: polarity}]$

(28) $\lambda e. [\hat{\delta}(e): \text{after-}e: \text{person: 1st: polarity} < \hat{\delta}(e): \text{after-}e: \text{person: 2nd: polarity}]$

These denotations are not assumed to exhaust the meanings of corresponding expressions. The lexical semantics of any two expressions of (im)politeness may separate them further, conveying other vivid associations. However, on the current theory, with respect to (im)politeness, addressing someone as “fool” supports the same inference as addressing the person as “spittle,” in terms of the speaker's estimation of the addressee versus the speaker's self-estimation.²¹ One may wish to argue further that this is the same meaning contribution provided by acid delivery (D'Errico and Poggi 2014).

12.4 Puzzles Revisited

That the framework provides a theoretical explanation of both general and uniquely linguistic puzzles of (im)politeness demonstrates that the semantic framework offers a useful interface to pragmatics (Sect. 12.2.1). To dismiss the possibility of a theory of linguistic (im)politeness on the basis of the fact that virtually any expression may be used politely or impolitely extends to dismissing all of semantics, given that any expression may be used literally, metaphorically, or ironically. However, this dismissal seems wrong, since, for example, irony is parasitic on literal meaning. Thus, in the view adopted here, both linguistic and extra-linguistic aspects of (im)politeness may be understood in relation to the perception of agents participating in events. The constraints involved may or may not lead to linguistic articulations. One need not accept all of the postulates in Leech's theory of pragmatics (Leech 1983) to agree with his view that “if we approach meaning

²¹ A theory which supports infinite graduations of esteem or disgust is not antithetical to the present work, but is not central to the present cause, either.

from a point of view which combines semantics and pragmatics, the result can be a satisfactory explanation. . .” (p. 7). As pragmatics is about the contextualized use of language, holding, as this semantic theory does, the denotation of (im)politeness expressions to be sets of events (constrained as appropriate to the relevant perception of offence and disgust according to the agents involved) individuates the same locus of reasoning as is employed in pragmatic inferences about language use in context.

The framework posits structure on events additional to the usual descriptive content. These constraints relate, as indicated, to the relative offence-polarity with which various dialog participants and potentially third parties are viewed, and the relative use and cost of events for the agents. One consequence of this treatment is that equivalence classes of (im)politeness exist as individuated by those constraints: all of the expressions that yield the same values are effectively synonymous (Sect. 12.2.6) to the extent that they support the same inferences regarding the speaker’s estimation of the agents involved in the triggering event. As mentioned above, lexical semantics may supply means of differentiating expressions of (im)politeness, but relative to the inferences directly connected to (im)politeness alone, synonymy obtains.²² The current treatment of polarity does not provide a direct model of a cyclical politeness spectrum (Sect. 12.2.3).

Semanticists since Reichenbach are accustomed to thinking about distinctions among utterance time, event time, and reference time. It is also necessary to take into account the utterer’s stance on events, the use and cost, for themselves and others. Interpretation of communicative actions as polite or impolite constrain the possible stance values. Therefore, the possibility for politeness to be excessive (Sect. 12.2.2) is explained by the determination of inconsistent information between stance values known (or presumed by reasonable default) to be in place and those derived from utterance interpretation. During pragmatic evaluation, the self-contradiction implied by an obsequious utterance fails to earn esteem, just as with other unreliable or manifestly deceptive behaviors.

Using the defaults of interpretation expressed so far, a reflexive thought such as (12.12) (see Sect. 12.2.9) encodes a self-appraisal along the lines of (12.29).²³ This is at odds with the defeasible provisions of both of the default interpretation principles (12.16) and (12.17). The example involves self-reference at two points in time: the event of burning of the toast and the utterance event. There is also implicit reference to the event preceding the burning of the toast. It is coherent for the speaker to adopt the perspective of superiority of the pre-toast-burning-self over the post-toast-burning-self, thus leading to the expectation of impoliteness. Moreover, if principle (12.24.a) has force, impoliteness is independently expected for this example. Thus, reflexive impoliteness expressions like (12.12) are sensible.

²²An anonymous reviewer disagrees with some of this, but that is possibly with respect to a version of the theory in which it is claimed that the relevant expressions are fully intersubstitutable and undifferentiable on lexical-semantic grounds. This stronger view is not argued here.

²³The intended interpretation of the co-indexing (e.g., i) is of structure sharing, including between the resolution of the disjunction in the first person value with the second person value.

$$(29)a. \hat{\sigma}(e^\Delta) =_{\text{attitude}} \left[\begin{array}{l} \text{before-}e^\Delta: \left[\begin{array}{l} \text{person: polarity} \\ \text{1st: } (1 \vee 0) \boxed{i} \\ \text{2nd: } \boxed{i} \end{array} \right] \\ \text{during-}e^\Delta: \left[\begin{array}{l} \text{person: polarity} \\ \text{1st: } (0 \vee -1) \boxed{j} \\ \text{2nd: } \boxed{j} \end{array} \right] \\ \text{after-}e^\Delta: \left[\begin{array}{l} \text{person: polarity} \\ \text{1st: } (0 \vee -1) \boxed{k} \\ \text{2nd: } \boxed{k} \end{array} \right] \end{array} \right]$$

b. $\hat{\sigma}(e^\Delta):after-e^\Delta:person:1st:polarity < \hat{\sigma}(e^\Delta):before-e^\Delta:person:1st:polarity$

That the contrasting example (12.13) is less felicitous constitutes possible evidence that the principle of extreme positive values (12.24.b) does not have force. The infelicity of the reflexive politeness example (12.13) could be explained by the impossibility of signalling greater esteem than is already conveyed in the default discussed above (12.16) with respect to the situation being evaluated (12.30).²⁴ While at face value both of these expressions ((12.12) and (12.13)) are contrary to the prediction of politic behavior with equal first and second person polarity, in the case of the politeness reflexive, unlike the impoliteness example, a differentiation of polarities associated with the co-extensive first and second person roles is more difficult to achieve. The constraint expressed by the explicit language of politeness cannot be satisfied in this formulation (12.30.b).

$$(30)a. \hat{\sigma}(e^\heartsuit) =_{\text{attitude}} \left[\begin{array}{l} \text{before-}e^\heartsuit: \left[\begin{array}{l} \text{person: polarity} \\ \text{1st: } 1 \boxed{i} \\ \text{2nd: } \boxed{i} \end{array} \right] \\ \text{during-}e^\heartsuit: \left[\begin{array}{l} \text{person: polarity} \\ \text{1st: } 1 \boxed{j} \\ \text{2nd: } \boxed{j} \end{array} \right] \\ \text{after-}e^\heartsuit: \left[\begin{array}{l} \text{person: polarity} \\ \text{1st: } 1 \boxed{k} \\ \text{2nd: } \boxed{k} \end{array} \right] \end{array} \right]$$

b. $\# \hat{\sigma}(e^\heartsuit):after-e^\heartsuit:person:1st:polarity > \hat{\sigma}(e^\heartsuit):before-e^\heartsuit:person:1st:polarity$

²⁴This would be an explanation in the spirit of the observation within generalized quantifier theory that positive strong determiners as arguments to existential assertions amount to tautologies, thereby accounting for the oddity of non-contrastive and non-demonstrative uses of sentences like “there is the person at the door” (cf. “there is a person at the door”) (Barwise and Cooper 1981).

Notice that a further contrasting example (12.31), involving not the language of reflexive politeness, but politic commentary, is not problematic.

(31) I_i am enjoying this fine toast I_i 've made.

Thus, self-praise is not predicted to be prohibited on this account, but self-politeness is.

The puzzles of learning (Sect. 12.2.5) and the incompletely Saussurean nature (Sect. 12.2.4) of the language of (im)politeness are both partly explained by the complexity of the calculus involved in using politeness expressions, relative to other most aspects of language use. Cooperative use of definite noun phrases presupposes that the addresser is able to make a reasonable calculation of what entities might be shared in a common ground understanding of salient possible references. Even in the use of definite noun phrases, it has been shown that speakers tend to make use of a notion of common ground that is inclined to their own world view rather than to the information available to interlocutors (Horton and Keysar 1996), demonstrating that calculations of addressee states are not straightforward. The calculation necessary in the proposed framework is additional to that required by the easily learned aspects of language, and partly accounts for the complexity of learning involved in (im)politeness, sensitive to triggers that are established by convention in the chain of association and to the relative status of interlocutors. While the calculators in Hurford's simulations were the worst performers in competition with Saussureans and imitators in establishing strategies for communicating messages (Hurford 1989), it is open that the calculating strategy may be necessary in optimal packaging of the message.

This discussion has demonstrated some of the successes of the semantic theory of linguistic (im)politeness described here in explaining several of the puzzles of (im)politeness that have been noted.

12.5 Final Remarks

Not all of the puzzles of linguistic (im)politeness are addressed in this chapter, and there is more to say about each of the puzzles within this framework. However, the semantic theory outlined has been demonstrated to have some traction in explaining a number of the puzzles. It is demonstrated that some aspects of politeness are appropriately treated as linguistic and through formal semantics. The calculation involved in politeness management is too important for it to disregard views of interlocutors, and therefore a strictly Saussurean strategy for linguistic packaging may be inappropriate. The fact that politeness expression augmentation can create disgust at a point of excess, rather than managing disgust positively arises through conflicting constraints. Asymmetry in the relative availability of reflexive impoliteness over reflexive politeness is also modeled. The complexity of politeness also makes it no surprise that linguistic politeness learning is at a different pace to language learning, generally.

Theories of (im)politeness have tended to focus on pragmatic theory.²⁵ The literature on politeness and impoliteness properly accords a major role for context in the evaluation of the experience of (im)politeness (Allan and Burridge 2006). However, appeal to the norms of language use within communities of practice to determine what counts as polite or impolite implicitly relies on the possibility of straightforward semantic interpretation to decode content and levels of politeness (e.g., Schnurr et al. 2008). Following Goffman's analysis of human interactions in terms of "face" (Goffman 1956, 1967), much work has been guided by a view of politeness as "facework" (Brown and Levinson 1987). An alternative view (Locher and Watts 2008; Bousfield and Locher 2008; Culpeper 2008) gives priority to relation management more than to agents in themselves, and in this conception, (im)politeness is seen as "relational work." In exceedingly coarse-grained terms associated with networks of communicating agents, the "facework" approaches may be seen as giving primary attention to the nodes (the agents), and the "relational work" approaches may be seen as attending primarily to the links (relations between agents), while the analysis presented here addresses (im)politeness as the management of a fog of offence that might otherwise engulf the whole network of agents and their relations.

The view pursued here is thus not at odds with those prior conceptions, but rather contains a distinct focus. The overarching effort is an attempt to contribute both to explanation of the perception of acts as polite or impolite and to the specification of the ontology required by a formal truth-conditional semantics for linguistic (im)politeness. It is argued that speakers use polite forms in order to avoid invoking disgust and impolite forms when they do not mind disgusting others.²⁶ Specifying denotations of expressions of (im)politeness as sets of events provides a semantic ontology also presupposed by pragmatic analysis of the phenomena. It is claimed that this approach clarifies some of the puzzles of (im)politeness, but others remain for future analysis.

Acknowledgements This research is supported by Science Foundation Ireland (Grant 12/CE/I2267) as part of the Centre for Global Intelligent Content (www.cngl.ie) at Trinity College Dublin. I thank Irene Sheedy for helpful discussions on the topic of (im)politeness. I am grateful to an anonymous reviewer for very constructive criticisms of an earlier draft of this chapter.

²⁵Of course, some others also emphasize that forms of language associated with (im)politeness expressions are open to semantic interpretation (Culpeper 2011).

²⁶Recall footnote 4. It is held in this chapter that impoliteness forms are expected when the speaker feels greater levels of offence from others than self-offensiveness. Accordingly, if a speaker is disgusted, then one expects impoliteness. However, this is always at the risk of in turn disgusting not just the target of any such impoliteness, but also any witnesses to the communication. Thus, ultimately, impolite forms are used when the triggers of disgust for the speaker are so great that the risks of being deemed disgusting by others are outweighed.

References

- Allan K, Burrige K (2006) *Forbidden words: taboo and the censoring of language*. Cambridge University Press, Cambridge
- Altmann G, Riška A (1966) Towards a typology of courtesy in language. *Anthropol Linguist* 8(1):1–10. <http://www.jstor.org/stable/30029136>
- Barwise J, Cooper R (1981) Generalized quantifiers and natural language. *Linguist Philos* 4(2):159–219
- Bax M (2012) An evolutionary take on (im)politeness: three broad developments in the marking out of socio-proxemic space. In: Bax M, Kádár DZ (eds) *Understanding historical (im)politeness*. John Benjamins, Amsterdam, pp 251–277
- Borkowska P, Kleparski G (2007) It befalls words to fall down: pejoration as a type of semantic change. *Studia Anglica Resoviensia* 47(4):33–50
- Bousfield D (2008) *Impoliteness in interaction*. John Benjamins, Amsterdam
- Bousfield D, Locher M (eds) (2008) *Impoliteness in language: studies on its interplay with power in theory and practice*. Mouton de Gruyter, Berlin
- Brown P, Levinson S (1987) *Politeness: some universals in language usage*. Cambridge University Press, Cambridge
- Bunt H, Alexandersson J, Chae JW, Fang AC, Hasida K, Petukhova O, Popescu-Belis A, Traum D (2012) Iso 24617-2: a semantically-based standard for dialogue act annotation. In: *Proceedings LREC 2012*
- Carpenter B (1992) *The logic of typed feature structures*. Cambridge tracts in theoretical computer science, vol 32. Cambridge University Press, Cambridge
- Cohen A (2009) No alternative to alternatives. *J Semant* 26(1):1–48. doi:10.1093/jos/ffn005. <http://jos.oxfordjournals.org/content/26/1/1.abstract>
- Culpeper J (2008) Reflections on impoliteness, relational work and power. In: Bousfield D, Locher M (eds) *Impoliteness in language: studies on its interplay with power in theory and practice*. Mouton de Gruyter, Berlin, pp 17–44
- Culpeper J (2011) *Impoliteness: using language to cause offense*. Cambridge University Press, Cambridge
- Davies BL, Haugh M, Merrison AJ (eds) (2011) *Situated politeness*. Bloomsbury, London
- D'Errico F, Poggi I (2014) Acidity. The hidden face of conflictual and stressful situations. *Cogn Comput*. <http://link.springer.com/article/10.1007/s12559-014-9280-1>
- Dowty D (1991) Thematic proto-roles and argument selection. *Language* 67(3):547–619
- Fernando T (2013) Segmenting temporal intervals for tense and aspect. In: *Proceedings of the 13th meeting on the mathematics of language*. Association for Computational Linguistics, Stroudsburg, PA, pp 30–40
- Fitch WT (2007) An invisible hand. *Nature* 449:665–667
- Gärdenfors P (2000) *Conceptual spaces: the geometry of thought*. MIT Press, Cambridge
- Gleason JB, Perlmann R, Greif E (1984) What's the magic word: learning language through politeness routines. *Discourse Processes* 7(4):493–502
- Goffman E (1956) *The presentation of self in everyday life*. Doubleday, New York
- Goffman E (1967) On face-work. In: *Interaction ritual: essays in face-to-face behavior*. Transaction Publishers, New Brunswick, NJ, pp 5–45. Reprinted 2008
- Gold EM (1967) Language identification in the limit. *Inf Control* 16:447–474
- Greif EB, Gleason JB (1980) Hi, thanks, and goodbye: more routine information. *Lang Soc* 9(2):159–166
- Healey P, Vogel C (1994) A situation theoretic model of dialogue. In: Jokinen K (ed) *Pragmatics in dialogue management*, vol 71, pp 61–79. Gothenburg Papers in Theoretical Linguistics. Proceedings of the special session within the XIVth Scandinavian conference of linguistics and VIIIth conference of Nordic linguistics, 16–21 August 1993
- Horton WS, Keysar B (1996) When do speakers take into account common ground? *Cognition* 59:91–117

- Hurford J (1989) Biological evolution of the saussurean sign as a component of the language acquisition device. *Lingua* 77:187–222
- Kamp H, Reyle U (1993) *From discourse to logic*. Kluwer, Dordrecht, NL
- Kasper R, Rounds W (1986) A logical semantics for feature structures. In: Proceedings of the 24th annual meeting of the Association for Computational Linguistics. Columbia University, New York, NY, pp 257–266
- Kolnai A (2004 (1929a)) Disgust. In: Smith B, Kormsmeier C (eds) *On disgust*. Open Court Press, Chicago, pp 29–91. Translated by Elisabeth Gombrich, Elizabeth Kolnai and Barry Smith
- Kolnai A (2004 (1929b)) The standard modes of aversion: fear, disgust and hatred. In Smith B, Kormsmeier C (eds) *On disgust*. Open Court Press, Chicago, pp 93–108 Translated by Elisabeth Gombrich, Elizabeth Kolnai and Barry Smith
- Kratzer A (1981) The notional category of modality. In: Eikmeyer J, Rieser H (eds) *Words, worlds and contexts*. Mouton de Gruyter, Berlin, pp 38–74
- Lascarides A, Briscoe T, Asher N, Copestake A (1996) Order independent and persistent typed default unification. *Linguist Philos* 19(1):1–89
- Leech GN (1983) *Principles of pragmatics*. Longman, London
- Locher M, Watts R (2008) Relational work and impoliteness: negotiating norms of linguistic behaviour. In: Bousfield D, Locher M (eds) *Impoliteness in language: studies on its interplay with power in theory and practice*. Mouton de Gruyter, Berlin, pp 77–99
- MacNeill D (1997) Growth points cross linguistically. In: Nuys J, Pederson E (eds) *Language and conceptualization, language, culture & cognition*. Cambridge University Press, Cambridge, pp 190–212
- Moxey LM, Sanford AJ (1993) *Communicating quantities: a psychological perspective*. Lawrence Erlbaum, Hove, UK
- Schnurr S, Marra M, Holmes J (2008) Impoliteness as a means of contesting power relations in the workplace. In: Bousfield D, Locher M (eds) *Impoliteness in language: studies on its interplay with power in theory and practice*. Mouton de Gruyter, Berlin, pp 211–229
- Tic Douloureux PRN (1992) A note on one's privates. In: Zwicky A, Salus PH, Binnick RI, Vanek AL (eds) *Studies out in left field: defamatory essays presented to James D. McCawley*. John Benjamins, Philadelphia, pp 45–51. Originally published in 1971
- Vallduví E, Engdahl E (1996) The linguistic realization of information packaging. *Linguistics* 34(3):459–519
- Van Genabith J (2001) Metaphor, logic and type theory. *Metaphor Symb* 16(1 and 2):43–57
- Van Kleef GA, Homan A, Finkenauer C, Gündemir S, Stamkou E (2011) Breaking the rules to rise to power: how norm violators gain power in the eyes of others. *Soc Psychol Pers Sci* 2(5):500–507
- Vogel C (2001) Dynamic semantics for metaphor. *Metaphor Symb* 16(1 and 2):59–74
- Vogel C (2014a) Denoting offence. *Cogn Comput*. <http://link.springer.com/article/10.1007/s12559-014-9289-5>
- Vogel C (2014b) Intended and unintended offence. In: Bassis S, Esposito A, Morabito FC (eds) *Recent advances of neural network models and applications, smart innovation, systems and technologies*, vol 26. Springer International Publishing, Switzerland, pp 395–404
- Watts RJ (2003) *Politeness*. Cambridge University Press, Cambridge

Chapter 13

Direct and Indirect Verbal and Bodily Insults and Other Forms of Aggressive Communication

Isabella Poggi, Francesca D'Errico, and Laura Vincze

13.1 Introduction

An insult is the last move within the field of communication before resorting to physical aggression, yet it is definitely to be preferred over physically acting out: as Freud put it, insulting your enemy instead of throwing him a lance was the dawn of civilization.

But what is an insult? How is it done? How do we distinguish it from and how does it overlap with other communicative acts like curses, imprecations, or simply bad words?

This paper outlines a model of insult in terms of a sociocognitive view of multimodal communication, while setting it apart from other types of aggressive communicative acts, and finally proposing an analysis of insults in Italian political talk shows and social media.

Section 13.2 presents the sociocognitive reference model, Sect. 13.3 distinguishes insults from imprecations, curses, and bad words in terms of the presented framework, and Sects. 13.4 and 13.5 present cases of direct and indirect verbal and bodily insults in Italian TV and online political communication.

I. Poggi (✉)

Dipartimento di Filosofia Comunicazione e Spettacolo, Università Roma Tre,
Via Ostiense, 234 - 00146 Roma, Italy
e-mail: isabella.poggi@uniroma3.it

F. D'Errico
UniNettuno University, Rome, Italy

L. Vincze
Università di Macerata, Macerata MC, Italy

13.2 A SocioCognitive Model of Social Interaction, Emotions, and Communication

According to one model of mind, social interaction, and emotions in terms of goals and beliefs (Conte and Castelfranchi 1995; Miceli and Castelfranchi 1998; Poggi 2007), the life of any natural or artificial individual or collective system consists in pursuing *goals*: regulatory states that, when sensed to be discrepant from the actual state of the world, trigger action. To realize a goal, the system projects and performs a plan where each action aims at one goal and possibly to one or more supergoals, all in the end aiming at a final goal.

13.2.1 Power, Dependence, Adoption, Aggression, and Image

In this framework (Castelfranchi 2003), an Agent X has the “*power of*” *g* if X is likely to achieve goal *g*, thanks to world conditions (e.g., material resources) or X’s own skills and knowledge. *Evaluation* is a belief about the extent to which some object, event, or person has or provides one with the “*power of*” necessary to achieve some goal (Miceli and Castelfranchi 1989; 1998). We make a positive evaluation of something or someone if we think that thing or person can or will allow us to achieve some goal and a negative evaluation when either it or the person does not have the necessary resources to achieve the goal (negative evaluation of impotence) or has resources that risk thwarting some goals (negative evaluation of noxiousness). Therefore, systems – and typically humans – very often conceive of evaluations about objects, events, and other people according to various aesthetic, moral, utilitarian criteria (goals), judging people as good or bad, stupid or intelligent, beautiful or ugly.

If X lacks the skills or resources to achieve *g*, while another Agent Y possesses them, X *depends on* Y to achieve *g*. This *dependence* gives rise to the social devices of *adoption*, *influence*, and *aggression* (Conte and Castelfranchi 1995). If X *depends on* Y, then X can achieve *g* either if Y *adopts* X’s goal *g*, i.e., if Y puts her actions and resources to the service of X’s goal, or if X *aggresses* against Y, i.e., she thwarts some of Y’s goals, to seize Y’s resources. Thus, X and Y may have the goal of influencing each other: X may want to influence Y to adopt X’s goal *g*_X, and Y may want to influence X to pursue Y’s goal *g*_Y in exchange. But if X depends on Y, then Y has the “*power to influence*” X, and to exercise that power, X may threaten to use this power of *aggressing* against X, that is, of preventing X from achieving an important goal. This gives Y “*power over*” X.

Adoption multiplies people’s “*power of*” to achieve their goals, thanks to resource exchange; but to decide what goals of what people to adopt, we need to evaluate people and their capacity and willingness to reciprocate: we form an image of them. Our *image* (Castelfranchi 1988; Castelfranchi and Poggi 1990) is the set of evaluative and nonevaluative beliefs others have about us. We strive to present a

positive image of ourselves so that others may adopt our goals, thereby gaining more “*power of*.” The image we present is generally functional to the type of adoption we aim at: to hire me as a real-estate seller you evaluate me as to my extraversion or argumentation skills, to choose me as a friend, as to my affective qualities. In general, to be adopted, we must elicit a positive evaluation, i.e., show an image of power in some areas (people adopt our goals because they esteem us); in rare cases we obtain adoption by presenting an image of a lack of power (people help us out of compassion), while an image of noxious power or moral lack of power results in contempt, inducing others to reject us.

Since a third agent, Z, in choosing whether to adopt the goal of X or Y, compares their respective values, a new kind of power stems from a power comparison: to have “*more power than*” another. We all have a “*goal of image*” and a “*goal of positive image*” (goal of esteem), including the goal of being evaluated better than others, and all the goals against which we want to be evaluated positively by others make up part of our *goal of* (positive) *image*. We also have a *self-image* (evaluative and nonevaluative beliefs about ourselves), a *goal of self-image* (we want to know our actual worth, to decide which goals to pursue, leaving aside ones beyond our reach), and a goal of positive self-image, or *goal of self-esteem*, to feel confident about ourselves while pursuing goals. A person’s image and self-image are tightly connected since they determine each other, but a person’s adaptation mainly depends on his self-confidence, which mainly holds when his self-image is not too dependent on the image other people have of him. Preserving a good image and self-image is among the most important goals of a person, since it is a primary source of power and a means to gain adoption, so much so that to *discredit* a person is an aggressive act aimed at lowering his actual power.

To discredit someone or something (a person as well as a commercial product or an institution) means to spoil its image before some audience, and it is a deliberate or nondeliberate effect of some communicative action (Poggi and D’Errico 2012a; D’Errico et al. 2012). The object of discredit is some Target T, and discredit may be cast on T either by some action performed by T itself or, deliberately or not, by some communicative action by some Sender S that spoils T’s image before some Audience AU, showing that T is not so good, smart, powerful as he tries to look. Politicians on political talk shows often try to discredit their opponents by communicative acts of accusation, criticism, and insult performed verbally or through bodily gestures.

13.2.2 Emotions

In this framework, emotions are multifaceted subjective states, encompassing internal feelings and cognitive, physiological, expressive, or motivational aspects that are triggered any time an important adaptive goal of a biological system is, or is likely to be, achieved or thwarted (Castelfranchi 2000). They monitor the achievement or thwarting of important adaptive goals, like survival and wellbeing, the acquisition

of knowledge, and acquiring and keeping resources, but also the goals of equity, attachment and affiliation, image and self-image. Humans experience positive emotions for the achievement and negative ones for the thwarting of these goals; hence, emotions can be clustered in families depending on the type of goal they monitor (Poggi 2008). Anger, but also pity and sense of guilt, monitors the goal of equity and is triggered by an undeserved imbalance between the fortunes of people. Shame and pride are “*image emotions*” that monitor the goals of image and self-image: we feel shame when we think that what we are or do may cause others or ourselves to have a negative evaluation of us, and we feel pride for something that enhances our self-image (Poggi and D’Errico 2012b). Further, since we also have the goal of making up an image of others, to decide whether to have positive or negative or no social relationships with them, we feel “*others’ image emotions*”: admiration (Poggi and Zuccaro 2008), which confers highly positive evaluations on the other and possibly a goal to imitate him, or else pity and contempt, the former giving the other a negative evaluation of impotence and inducing us to help him (Castelfranchi 1988), the latter assigning to him an evaluation of ethical baseness and inducing us to reject any relationship with him.

13.2.3 *Communication*

Communication (Poggi 2007; Poggi and D’Errico 2012a) takes place when a System S (sender) has the goal of having another System A (addressee) come to have Belief K and, to achieve this goal, produces a signal s that is linked to K as its Meaning M, according to the rules of a Communication System CS. s is a behavior or a morphological trait produced by some organs of S’s body, to be perceived by A in some receptive modality (e.g., visual, auditory, olfactory, tactile). The unit of communication is the communicative act: a single signal or a set of signals that conveys, as its meaning, both a performative and a propositional content. The performative is the specific communicative goal of the sender –e.g., to state, confirm, approve, order, implore, ask, promise, wish – and all performatives can be grouped into one of four types: to provide information (generally expressed by declarative sentences), ask a question (interrogatives), request some action (imperatives), or express a desire (optatives). The propositional content is the object of the communicative goal: the information provided by a declarative sentence or asked by a question, the action requested by an imperative, the desire expressed by an optative. Any act of communication, through its performative or propositional content, provides information about the world (states or events concerning persons, objects, their time and space), the sender’s identity (age, gender, ethnicity, geographical and cultural roots, projected self), and the sender’s mind: the beliefs, goals, and emotions held during and about the present communicative act.

The signal side of a communicative act may be verbal or bodily behaviors (words, gestures, gaze, posture) or traits (like a blush), and the whole communicative act may be conveyed either by one *holophrastic* or more *articulated* signals. In the

former, a single signal has the meaning of both performative and propositional content, as with interjections in verbal languages (Poggi 2009) (e.g., “*wow!*” which means “I inform you I am pleasantly surprised with this event”), but also as “holophrastic” bodily signals (the gesture *palm down with fingers bending down* for “please come here”). In the latter, the whole meaning is fragmented into lexical items: a sentence resulting from the combination of words.

A distinction can be made between communication and expression based on the level of intentionality and awareness of the signal and on the type of meaning conveyed (Poggi 2004, 2005): in *communication* the sender has an individual conscious goal of providing some information that may concern the world or the sender’s mind or identity; in bare *expression*, in contrast, the goal of communicating is a biological, social, or unconscious goal, and the information conveyed concerns the sender’s mind (e.g., level of certainty of conveyed information or emotions being felt). Thus, while an aware expression of indignation is *communicative*, an involuntary blush of shame, a person’s regional accent, or an eyebrow raised in emphasis are *expressive* signals. Moreover, if an emotion leaks from our posture, voice, or facial expression, not because we deliberately want to communicate it but because we simply “give vent” to it, this is a “nonsocial,” “noncommunicative” signal, hence only *expression* – not a full-blown *communication* – of emotions (Poggi 2004, 2005, 2009).

Any verbal or bodily communicative act necessarily has a literal meaning, the one drawn from its lexicon and (for verbal acts) syntax, but it may also have an indirect meaning that the sender wants the addressee to catch by inference. For example, the indirect meaning of a question may be a criticism, one of a sad facial expression, an apology.

13.3 Insults and Other Types of Aggressive Communication

The world of aggressive communication is manifold. People may hurt each other by whole sentences or discourses, single words uttered here and there, or by gestures, posture, and gaze. What is hurt by these communicative weapons are not parts of our body but recesses of our soul. We are wounded by others’ hate, contempt, and indifference, and what is wounded is our deepest self.

Insults are but one example of aggressive communicative action and is distinguished from other ways of hitting others in communication, with which they insults are often combined. They are easily confused with curses, imprecations, and bad words, which, though being in some way aggressive, may stem from different emotions, arise from different antecedents, and carry a peculiar type of blow in social interaction.

According to the foregoing sociocognitive model, these forms of aggressive communicative action differ from two points of view:

1. Pragmatic structure

Curses, imprecations, and insults are whole communicative acts. Bad words, in contrast, are single words, conveying only a part of a communicative act – for instance, a noun like *shit*, an adjective like *idiot* (which can well be used as an insult, but only given certain conditions), a verb like *fuck* or *damn*, a noun phrase like *the hell* – that all convey meanings subject to communicative taboo.

2. Antecedents: facts and emotions

Curses, imprecations, insults, and bad words also differ in the types of events that can precede and cause them and in the emotions expressed by them and triggered by those events.

13.3.1 *Curse*

A curse is a communicative act through which a sender communicates to a target that he wants some very bad event to happen to him. This may be expressed in two ways:

1. An imperative communicative act, by which the sender orders the target to do a somewhat self-defeating action to himself, as in

(1) *Go to hell!*

(2) *Fuck off!*

2. An optative communicative act (Poggi 2009): an act in which the sender, though addressing the target, makes appeal to a third entity – god, fate, fortune, to which the sender attributes the power to make events occur or not – to cause a very bad event to happen to the target, as in

(3) *May you sink at sea!*

Or in the biblical curse by Noah to Canaan (Genesis, 9:25):

(4) “*Cursed be Canaan! The lowest of slaves will he be to his brothers.*”

Here the sender wants something bad to happen to the target and, to express this desire, uses a modal verb of power (*may*) in an optative mode: a typical phrasing of an appeal to a third omnipotent entity considered as being able to fulfill the sender’s desire.

A particular kind of optative curse, so-called elliptical optative curses, are elliptical with respect to verbs or nouns mentioning a curse or bad things. Expressions like (5) in some Italian dialects

(5) *La potta de to’ ma’*
(your mother’s vagina)

or

(6) *L’anima de li mortacci tua*
(the soul of your dead ancestors)

may be considered elliptical with respect to words mentioning the cursing intention, as a brief form for:

(7) May your mother's vagina be cursed

or

(8) May the soul of your dead ancestors be cursed

Both imperative and optative curses are directly addressed to the target (target = addressee), and their literal meaning is to order or to wish, respectively, a highly destructive action or event against the target, but this in turn aims at communicating: I do not want to have any more social interaction with you!

What events and emotions generally trigger curses? A typical antecedent is that the target performed some misdeed to the detriment of the sender. This serious, unjust damage triggers three mental states in the sender: first, an emotion of anger, typically monitoring the sense of injustice; second, a desire for revenge that is sometimes so overwhelming that no human might carry it out and that could be adequately performed only by some omnipotent entity. This is the point of wishing a very disrupting event on the other. But anger is too short term an emotion to be felt for such a person and such a misdeed; what leaks out from the curse is a more long-term, indestructible feeling, hate, which is something like a long-term anger. And hate triggers, in its turn, a goal of severing any future relationship with the other, who made himself guilty of such a misdeed.

Therefore, in optative curses the literal goal is to ask a third entity to have something bad happen to the target, while in imperative curses the sender asks for a somewhat self-defeating action from the target (and addressee) himself; but in both cases the indirect meaning is to communicate the sender's rejection to the target.

13.3.2 *Imprecation*

An imprecation can be seen as an optative curse or an insult addressed to some inanimate object or, again, to a third entity that one holds responsible for an unlucky event. For example, *Damn!* means: I appeal to a third entity (addressee) to make the target be *damned*.

What events and emotions trigger an imprecation? Typically "*Damn!*" might be uttered, for example, if you stumble in the dark and bang on a chair, or you are going for a picnic and a storm is approaching. In such cases you curse the chair or the weather: that is, by a somewhat animistic attitude, you are making those things responsible for a misdeed that harmed you in some way. This may trigger your anger – an emotion sometimes only felt in connection with the bare frustration of a goal, not necessarily some actual injustice - but not necessarily revenge or hate: (Only if the animistic attitude is very strong could one feel that a social relationship was previously held and, hence, now explicitly reject it.)

In a *curse*, the sender, whether or not making an appeal to a third entity to have something bad happen to the target, is addressing the target to communicate that he does not want to have anything more to do with him. In an *imprecation*, in contrast, the sender does not address a person to tell him he wants something bad for the other; rather, he is making an appeal to a third entity, or at least to the entity he considers responsible for the damage received, and does so in an aggressive way, for instance by using bad words or even by insulting the entity itself.

A sender may perform an imprecation whenever he is surprised or disappointed and needs to express his mental state in an intense way (Poggi 2009), so he may resort to dysphemism: lexical items whose form or meaning is particularly crude or aggressive: interjections like *boia* (executioner), *miseria* (misery), *vacca* (cow), *merda* (shit) that are dysphemistic as to their meaning; or others like *corbezzoli* (good gracious!, literally, arbutus!), which are dysphemistic as to their phonic appearance; or *cazzo* (cock), which is dysphemistic on both the signal and the meaning side.

The main difference between *curse* and *imprecation* is in their communicative status. A *curse* is necessarily an act of communication: the sender has the conscious goal of making the target understand that the sender is feeling an emotion of anger and hate *toward* him, that he wants bad things to happen to the target or the target to do things bad to himself, and that he wants to cut off any social ties with him. An *imprecation*, in contrast, may simply be an expression of an emotion: the sender may only have the goal of giving vent to his anger or disappointment about some unlucky event. So the imprecation does not necessarily have an addressee proper since it is not a communication to others. Sometimes, an imprecation's addressee is a third entity, but often in this case, imprecating is but an indirect meaning for a communicative act with a literal meaning of curse or insult. For example, in Rome, *mannaggia* derives from *male ne abbia* (let him have something bad): clearly a curse, but nowadays only felt as an imprecation. Just as *porco Giove* (you pig Jupiter) has the form of an insult (see below) to a third entity but is used as an imprecation.

13.3.3 *Bad Words*

Unlike curses and imprecations, bad words are not complete communicative acts but only words or phrases – and therefore only fragments of communicative acts – that

1. Mention bodily organs or actions linked to physiological (e.g., sexual or excretory) functions or to other tabooed semantic areas, such as death, disease, or economic resources, in sum various contents protected by norms of privacy;
2. Are mainly phrased in a low sociolinguistic register, that is, vulgar or rude versions of the terms, not, for instance, the scientific nouns or verbs for those

organs and actions (Galli de' Paratesi 1969). (For example, *fuck* is a bad word, *have intercourse* is not.)

Bad words differ from curses and imprecations in that they are not whole communicative acts: a bad word can be uttered in communicative acts with different performatives. I can use *bloody* in an interrogative sentence:

(9) *Where did you put my bloody coat?*

or in an imperative sentence:

(10) *Give me that bloody book!*

or in an informative one:

(11) *I hate this bloody weather.*

If used as an interjection (e.g. *Shit!*), a bad word conveys a whole communicative act of imprecation.

A bad word, if used as a communicative act or within one, has the goal of adding emphasis to the act itself but may also express or communicate anger and possibly a goal of aggression stemming from an unwelcome event, in this case addressing the one held responsible for it. Bad words are a case of “dysphemism”: words loaded with negative evaluation and aggressive import, both for their very meaning and, possibly, because of their harsh sound or the sensation of friction or obtrusion of their production that, if simulated or reproduced by the receiver, thanks to mirror neurons, may give him the same sensation of harsh production as for the sender. Therefore, they often make up part of curses, imprecations, and insults, enhancing their load of aggression, yet they must be carefully distinguished from them.

13.3.4 *Insults*

An insult is a whole communicative act produced by a sender with the deliberate intent of offending a target entity T (a person, a group, even an object, for example, the symbol of an ideology or an institution), and it does so by attributing a very negative property to the target, finally including him in a category that is degrading for him, in such a way as to spoil the target's image and (in the case of a person) the target's self-image, too.

13.3.4.1 **Intent to Offend**

To feel offended means to feel that another person has a worse image of us than the one we want to project. But a person T may feel offended even without a deliberate intention of offending on the part of another person S.

On the bus the young boy *S* gives up his seat to a lady of 60. She is offended because she thinks he has an image of her as an old lady who needs a seat on the bus.

Here *S* did not intend to communicate that he has of *T* an image of an “old” lady, but yet she may feel offended to the extent to which she attributes to him this intention.

In some cases, *S* does have the goal of making *T* understand that *S* has a lower image of *T* than the one *T* wants to project. This is the case with insults, in which the sender has the goal of communicating not only that he includes the target in the degrading category but also that he does so with the goal of offending the target (Castelfranchi 1988).

13.3.4.2 Negative Property and Degrading Category

To attribute a very negative property to the target, the sender attacks the core of his identity, that is, he communicates that he considers the target as belonging to a “degrading category,” one primarily characterized by that negative property and considered on a lower – less noble – level than the one the target (at least based on the sender’s assumption) pretends to belong to, for instance, a category of people of a lower race, a lower social class, or even nonhumans, animals, or objects. The characterization of this degrading category is achieved through the devices of dehumanization, mechanization, and the like (Zamperini and Menegatto 2015), possibly ending with the most severe, on this scale, of negative properties: nonexistence, which is evoked by insults like “You’re a nothing,” or in deliberately and ostentatiously ignoring the target.

13.3.4.3 Spoiling the Target’s Image and Self-Image

Since the goal of an insult is to spoil the target’s image, in this scene we may count two or three actors: (1) the sender, or insulter, the one who performs the insult; (2) the target or insulted, the one who should suffer from the offense of the insult; and, possibly, (3) the audience, one or more persons who witness *S*’s act but are not directly involved in it.

An interesting issue is whether the addressee of an insult must necessarily be the target: we might claim that when the attribution of the negative property or degrading category is only communicated to an audience, this might be gossip, accusation, calumny, or talking badly of someone, but not an insult. We will try to answer this question in the empirical study of Sect. 13.5.

13.3.5 *Bad Words, Imprecations, Curses, and Insults*

An insult differs from an imprecation or a curse in that it is not an optative or an imperative but an informative act (or even only a vocative, as we shall see subsequently) that mentions a negative evaluation of the target and communicates it

explicitly to the target or an audience, with the further goal of offending the target. As in imperative curses, here the relationship is not mediated by some third entity but is only between the sender and the target, who may (or even must) also be the addressee; yet the content of the communicative goal is not an action – as in curses – but a negative evaluation of the target.

As for the emotions expressed or communicated, an insult is a communication of disesteem and lack of respect toward the target; in a sense, as argued by other authors (Vogel 2015) it is connected to disgust; but disgust for objects becomes contempt when addressed to a person. Moreover, while in imprecations the emotion of anger may be simply expressed as a spontaneous way to give vent to the sender's feelings, in an insult the sender deliberately aims at communicating his disgust and contempt to the target. Sender S not only communicates to T that he has a low image of T, but also meta-communicates that he communicates this to T because he has the deliberate intention of offending T. And this is further degrading for T, because if someone defies you without fearing retaliation, he must believe you to be someone with very little power (Castelfranchi 1988).

Therefore, in insulting, Sender S

1. Has the goal of offending T, i.e., letting T know that S has an image of T as belonging to a category that is degrading for T, hence as being entitled to an image lower than the one T pretends to belong to;
2. Has the goal of communicating that S intends to offend T and that he does not fear T's retaliation, which is in itself offensive since it diminishes T's image of power, finally, through this; and
3. Has the goal of attacking T's self-image, thereby causing him to lose face before himself, in addition to before others.

Of course, all this is even more offensive and face-threatening if done publicly, that is, if the attribution of a lower image is publicly displayed in front of an audience, in other words, when not (only) the target but the audience is also addressed.

Table 13.1 summarizes the differences among bad words, imprecations, curses, and insults.

13.4 Verbal and Bodily Direct and Indirect Insults

The communicative act of an insult may be performed with both verbal and bodily signals and can be both direct and indirect, according to whether the insulting content is explicitly stated by the literal meaning of the communicative act – for instance by the insulting meaning of the words used or by a syntactic construction peculiar to insults – or else it must be caught by inference from an apparently noninsulting meaning. Therefore, to distinguish indirect insults implies clearly determining the canonical pragmatic, semantic, and syntactic forms of direct ones. Let us start from the linguistic form of direct insults.

Table 13.1 Bad words, imprecations, curses, and insults

	Bad word	Imprecation	Curse	Insult
Example	<i>Bloody</i>	<i>Damn!</i>	<i>May you sink!</i> <i>Go to hell!</i>	<i>Idiot!</i>
Cause	Frustration	Frustration	Feeling of injustice	Feeling of injustice
Triggering emotion	Anger	Anger	Anger	Anger
Felt emotion	Anger	Anger, disappointment	Revenge, hate	Revenge
Expressed emotion	Anger	Anger, disappointment	Revenge, hate	Disgust, contempt
Expression/communication	Expression, communication	Expression, communication	Communication	Communication + metacommunication
Type of communicative action	Single word	Communicative act	Communicative act	Communicative act
Goal	If expressive, to give vent to anger If communicative, to communicate, aggressive intent	To give vent to anger or disappointment	Wish negative event for Target Revenge over Target Reject relation with Target	Attack target's image Attack target's self-image Revenge against target
Performative		Optative	Imperative or optative	Informative or vocative
Propositional content	Taboo object, action, or event	Negative event for target	Target's action negative for himself or negative event for target	Negative property or degrading category of target
Characters		Sender, third entity	Sender, target, third entity	Sender, addressee, target, audience
Addressee		Addressee = third entity	For imperatives Addressee = target For optatives Addressee = third entity	Addressee = target

13.4.1 Verbal Direct Insults

An insult is an informative act or a vocative by which the sender assigns a very negative property to the target or includes the target in a degrading category. In its most typical cases, the insult may take the following linguistic forms:

1. An informative sentence like “You are X” (*you are idiotic*), where X is an adjective conveying the negative property.
2. An informative sentence like “You are an X,” where X is
 - (a) either a nominalized adjective (*you are an idiot*) conveying the negative property;
 - (b) or a noun (*you are an animal*) conveying the degrading category.
3. An adjective or a noun used as a vocative “You X!”, that is, used to summon the target (*idiot!*).

13.4.1.1 Semantic Aspects

To understand the different import of these forms of insult, we must go back to the meaning and function of nouns. A noun is a word that names a category of entities and contains in its meaning all the distinctive properties that characterize all the entities in that category as being similar to each other and different from entities of other categories. Therefore, when naming someone or something by a noun, one assigns to it all the defining properties of that category. But among these properties one or more may be subject to a negative evaluation, or the very fact of being assigned to that category might even rule out some desired positive evaluation.

An example is the noun “*capra*” (goat), launched as an insult by Vittorio Sgarbi, an Italian art critic and politician, to attack a female opponent during a talk show.¹ The category of *goat* is defined, among others, by the property “NOT INTELLIGENT,” so the very fact of being assigned to that category implies being stigmatized with this property. But more than that, the category of goat is considered (also from being not intelligent) as a lower category than that of person.

All the aforementioned cases assign a negative evaluation to T, but 1 and 2a do this only, and do so directly, while 2b and 3 assign T a negative property indirectly, by inserting him into a “degrading category”: a category of entities seen as definitely inferior to one T pretends to belong to – an animal, an inanimate object, a person of a lower class or race. Therefore, among those cases, insults n.1 and n.2a, where the property is conveyed by an adjective, simply assign to T a negative property, which may be one negative property among others, not necessarily permanent, or characterizing T as such. In cases 2b and 3, however, the mentioned property, being phrased as a noun, becomes not only one among many but the one defining feature

¹All our examples are taken from communicative acts by, toward, or between Italian politicians.

of the target, what characterizes him the most. This is why cases 2b and 3 are more serious insults than 1 and 2a, and 3 is definitely the most serious of all: here the sender summons the other, and requests his attention, by calling him by that noun and making it his name. This makes it the most aggressive – and most prototypical – form of insult.

13.4.1.2 Syntactic Aspects

When an insult is not phrased as a vocative (case 3) but as an informative sentence, the negative property or degrading category must be asserted, not only presupposed. Therefore, it must be mentioned in the main clause, not in a subordinate or relative clause, or in a modifier.

For example,² a sentence like

(12) Renzi tells lies, and Di Maio destroys them

is a direct insult, while

(13) *Di Maio demolisce le balle di Renzi*
(Di Maio destroys Renzi's lies)

is more indirect because Renzi's lying is only implied, not directly stated.

Further, when the negative property is stated as an adjective, it must be used with a predicative function, not simply an attributive one.

In a comment about Prime Minister Matteo Renzi, Beppe Grillo says:

(14) *Tre avvocati milanesi hanno brutte notizie per il condannato Renzi.*
(Three attorneys from Milan have bad news for the condemned Renzi)

Here the negative property is given by an adjective with an attributive function. If Grillo had said:

(15) Renzi, sei un cittadino condannato
(Renzi, you are a condemned citizen)

this would have been a direct insult; but the way it is phrased in (14) is only indirectly insulting because the adjective *condannato* (condemned) is used with an attributive function (as marked in Italian by the adjective preceding the noun), hence it is presupposed, not asserted, information. Thus, this sentence can in fact be considered an insult, but only an indirect one.

²Our verbal examples are mostly taken from the corpus of D'Errico et al. (2014), drawn from the blog and Facebook page of Beppe Grillo, the head of the Italian political movement Movimento 5 Stelle. The examples in italics are real.

13.4.2 *Direct vs. Indirect Verbal Insults*

A verbal insult may be given in either a direct or an indirect way. An indirect insult is a communicative verbal or bodily act in which the negative evaluation of T cannot be understood by applying the bare lexical and syntactic rules of the language but can only be inferred by taking into account, along with the literal meaning, the contextual or cultural knowledge shared by the sender, target, and audience.

We distinguish two types of indirectness:

- (a) *Syntactic indirectness*, where a sentence violates the syntactic constraints mentioned earlier, but contains some insulting meaning that can be gleaned from the presuppositions contained in a subordinate clause, a modifier, or an attributive adjective;
- (b) *Pragmatic indirectness*, where the insulting meaning can be discerned only by applying a more or less elaborate path of inferences to go from the literal meaning to the insulting one.

Indirectness may be a matter of degree: an insult can be more or less indirect depending on how far it is from the aforementioned canonical form of insult and on the number and types of inferences necessary to understand the insulting property or category. Items (13) and (14) were cases of syntactic indirectness, (17) one of pragmatic indirectness.

After the Parliament Chair, Laura Boldrini, applied a rule to cut off too prolonged discussion of a law by the opposition, among Grillo's followers' sharp comments, some are definitely insulting. A case of direct insult is

- (16) *E' fascista*
(She is a fascist)

Then Grillo put a provocative question on his Facebook page, asking his followers what they would have done had they been alone in a car with Boldrini. One of the many harsh answers was:

- (17) *la lascerei al G.R.A. a battere, non sa fare altro.*
(I would leave her on the motorway **to sell herself**, That's all she knows how to do)

This is a case of pragmatic indirectness: the insulting meaning implied is "Boldrini is a prostitute"; but this degrading category is not explicitly mentioned in the sentence; rather, it must be inferred by connecting the meanings of the two sentences.

13.4.3 *Direct and Indirect Bodily Insults*

To insult people or institutions, words are not necessary: we may do so by gaze, facial expression, gesture, or posture, or even without doing anything. *Not to accept an offer* may be considered an insult – not only by a mafia godfather – and be

deliberately used by the (non)sender with the intent of showing disrespect, of offending the addressee's image by diminishing him, in some cases to the point of not acknowledging his presence. This is why *not to say hello* to someone may be a serious insult since it does not credit the other's existence, as if to say: "You're nothing." Nothingness.

Let us now consider some insults by gestures and other bodily actions. To insult by hand movements, one can use both codified gestures (Poggi 2007) – those usually represented in the mind as a gesture-meaning pair – and creative ones – the iconic ones created on the spot by imitating shapes and actions.

(18) A typical insulting Italian codified gesture is the *hand raised with index and little finger extended upward as horns*, which means "You are a cuckold."

(19) Another codified insulting gesture was used by Luigi De Magistris, during the rally for mayor of Naples, in a TV debate against his opponent Giovanni Lettieri: he *waves his right hand, in a fan shape, palm facing left, close to his head from right to left, four times*: an Italian symbolic gesture that means "Are you mad?" (D'Errico et al. 2013).

(20) A less codified gesture, done to insult the European Union, was used in 2014 by Gianluca Buonanno (a deputy from the secessionist Lega Nord party, which is strongly opposed to Europe), who *used the European flag to blow his nose*.

Facial expressions and gazes can be used for insults, too: a disgusted face, *spitting in the target's face* to show contempt, *raising an eyebrow* to display skepticism (Ekman 1979), an *ironic smile, laughter*; finally, a *loose and relaxed posture* may be insulting, especially when talking to a person who wants to be respected and honored.

Among bodily insults, too, some may be indirect.

(21) When interviewing the right-wing minister and university professor Renato Brunetta, the journalist Daria Bignardi misremembers the name of Brunetta's old teacher. To underline Bignardi's error and his own amazement in the face of her ignorance, Brunetta *covers his face with his hands* to convey despair; then he *puts his hands down* on the table and *squints his eyes* as if suffering from a hard blow; finally, *he leans on his left hand, with his elbow on the table*, as one forced to remain where he is listening to a person who has just displayed her profound ignorance. He is communicating: "I am in despair over such unforgivable ignorance," which implies, in turn, "You are very ignorant!". This is why Brunetta's pantomime counts as an indirect insult.

13.4.4 A Serial of Insults

Since insults are a heavy blow to people's image, they may easily trigger heavy reactions: for example, "counterinsults," insults against the insulter. An example comes from a sequence of reciprocal insults launched by Italian political opponents.

(22) In 2013, the left-wing Prime Minister Enrico Letta makes Cécile Kienge, a black doctor from the Democratic Republic of the Congo and an Italian citizen, the Minister for social integration. On 14 July 2013 Roberto Calderoli, a member of the right-wing racist party Lega Nord (North League), fighting for the secession of Padania (North Italy), during a

political discourse to Lega followers, referring to Kienge, says: “When I look at her, I can’t help thinking of an orangutan.”

According to our definition given earlier, this is a pragmatically indirect insult, though a very slightly indirect one: not many inferences are necessary to catch its meaning.

(23) Some days later, during a public talk by Minister Cécile Kienge, someone *threw two bananas* at her.

An indirect bodily insult: a creative gesture again implying inclusion of Kienge in the degrading category of orangutans.

(24) Later again, a picture is published with Calderoli as an orangutan, with the caption: “*Orango padano*” (Orangutan from Padania)

A verbal and visual direct counterinsult.

A way to recover from the continuous escalation of insults and counterinsults is to accept the provocation carried by the insult and to turn a prospective defeat into a gain of power.

(25) On 27 April 2014, a racist fan during a game *hurls a banana* toward Dani Alves, a black Brazilian soccer player. Alves *picks it up, peels it, and eats it*. He is overcoming the offense of being included in the degrading category of apes by taking a proactive attitude, without insulting his insulter. His act indirectly means: “I don’t care about you and your insults. You are not so important to me that I’m offended by a stupid provocation. I am better than you.” From then on, *eating a banana* becomes a symbol of a struggle to combat racism, but a struggle that makes fun of it, and thus diminishes it.

13.5 Aggressive Communication on the Web: Some Real Examples

To test the descriptive adequacy of the foregoing distinctions, we conducted an empirical study on a case of aggressive communication on social media.

During a parliamentary session over pending legislation, members of the Movimento 5 Stelle, a political movement led by Beppe Grillo, tried to launch a filibuster, stretching out the discussion with long talks to block the approval, but President Boldrini applied the “guillotine”, a legal technique to break down the discussion, to have the law approved in time. Boldrini was then strongly attacked by all followers of Beppe Grillo on his Facebook page.

We extracted a corpus of 12,038 words from this interaction (D’Errico et al. 2014). In this corpus, several categories of aggressive communication were singled out.

13.5.1 Direct Insult to Second Person (*Addressee = Target*)

A first category comprises insults in their most canonical form: they are addressed directly to the target and phrased in the form “You are an X” (*boldrini sei una vergogna*

mondiale ! ! != boldrini you are a world disgrace!!!), without the copula (*boldrini schifo != Boldrini suck!*; *peste bubbonica ! != pest!*, Black Death!; *cagna cagna cagna = bitch bitch bitch*), “Name + adjective” (*Boldrini, serva dei potenti e nemica degli italiani = Boldrini, servant of the masters and enemy of Italians*), or, finally, the vocative (*a marcionaaaaaaaaa = oh you big rotten*). This last one is a “creative” insult, in that “*marciona*” (big rotten woman) is a neologism. Other less classical but similar forms are: “insulting adjective + Name” (*cretina boldrini = stupid boldrini*), “insulting verb” (*fai schifo ! != you suck!*), or finally an insulting verb in the form of exclamatory sentence (*Che pena che fai = how I pity you*).

13.5.2 Direct Insult to Third Person (Addressee = Audience)

In some sentences the insulting adjective or noun is assigned to the target as a third person, as if referring to her while addressing the audience: *Boldini VERGOGNOSA* (Boldini shameful), *la boldrini è una nullità* (boldrini is a nothing), *codesta umanoide* (this humanoid), *zarina stalinista di merda* (Stalinist Czarina shit).

In one case, the insult is also argued: *la boldrini è una zozzona quando parla di lampedusa si eccita pensa al c dei suoi amici immigrati africani = boldrini is a dirty woman when she talks about lampedusa (a Mediterranean island where many refugees end up) she gets excited she thinks of the c . . . of her African immigrant friends*).

13.5.3 Indirect Insult: Syntactic Indirectness

As stated earlier, in some cases the insulting word or phrase is present but only presupposed and not asserted, for example, *QUESTI PEZZENTI VOGLIONO LA GUERRA !* (These **beggars** want war!); *li dobbiamo mandare tutti a casa sti porci* (we must send all of them home, these **pigs**). These insults are all in the third person (addressee = audience), which is, in a sense, one more reason to consider them indirect.

13.5.4 Indirect Insult: Pragmatic Indirectness

In other cases, the insulting meaning is not simply presupposed but must be inferred by making reference to rhetorical devices like reticence, insinuation, rhetorical question, or irony. *Vorrei sapere a quale ruolo è adatta la boldrini* (I would like to know what role boldrini is good for) lets you infer a total lack of skills; *mi viene uno strano istinto quando vedo la faccia della boldrini* (I feel a strange instinct as I see boldrini’s face) implies a nonverbal insult: the desire to spit in her face.

While in these cases the addressee is the audience, in what follows the indirect insult is in the second person: *boldrini vieni a pescare con noi ci manca il verme* (boldrini come fishing with us: we are out of worms). Finally, the very fact of *not writing Boldrini’s name with the capital letter* is a way to let you infer disrespect – another kind of indirect insult.

Other cases of aggressive communication in our corpus include direct and indirect criticism.

13.5.5 *Direct Criticism*

Some sentences mention a flaw quite explicitly, for example, *lei fa quello che Berlusconi e Renzo gli dicono di far* (she does what Renzi and Berlusconi tell her to do).

13.5.6 *Indirect Criticism*

A case of syntactic indirectness is: *Questo movimento rappresenta viceversa l'unica speranza degli italiani di poter cambiare uno governo e uno stato che li sta strangolando* (on the contrary, this movement is the only hope of Italians to change a government and a state that is strangling them). That "the state is strangling Italians" is presupposed, so it is a criticism embedded in the main sentence.

Examples of pragmatically indirect criticism are: *Vai a lavorare* (go to work), which implies you generally do not work; the ironic exclamation *quanto è diventata affascinante pure questa signora strappa lacrime* (how charming has become this tearjerker lady); and in the second person (directly addressed to her) *suor boldrini* (sister boldrini): an ironic allusion to her being quite moralistic and struggling for political correctness.

13.5.7 *Aggressive Request and Aggressive Prohibition*

Another type of aggressive communication in our corpus comprises aggressive requests to Boldrini or other politicians, generally in the second person, like *A casa la BOLDRINI !!!!!!!!!!!!!!!* (BOLDRINI home!!!!!!), *boldrini dimissioni* (boldrini resign), *fuori dai coglioni* (begone!).

The aggressive request is an imperative communicative act ordering the other to do something that is bad for him/her and has as its opposite – though rare – case the aggressive prohibition: forbidding the other to do something, as in *non rompere i maroni* (don't bother).

13.5.8 *Aggressive Instigation*

Aggressive instigation is a communicative act by which the sender incites the audience to do something bad to the target. Examples from our corpus include the following: *Spaccateli tutti a sti bastardi !!* (break them, all these bastards!!); *a calci nel culo dobbiamo prenderli* (we must kick them in the ass); *mandatela a zappare in sud africa insieme al suo amico presidente* (send her to hoe in South Africa together with her friend the president); and, more creatively: *rompile gli orecchini* (break her earrings).

13.5.9 *Curse*

Finally, our corpus also contains curses: from fast and classical ones like *Maledetto !!!!!!! !!!!!!! !!!!!!! Giorgio !!!!!!! !!!!!!!* (curse on Giorgio = Giorgio Napolitano, the Italian

president) to more specific and elaborate ones: *spero che un giorno quando creperete tutti letta renzi boldrini monti napolitano ect ect quello che state facendo un giorno ne dovrete rispondere a qualcuno e non ci sara movimento politico o inciuci a cui potrete* (I hope that someday when you all kick the bucket letta renzi boldrini monti napolitano and so on and so forth what you are doing someday you will have to answer for to someone and there will be no political movement or bargain you will be able to [resort to]).

Here the starting verb *spero* (I hope) makes the sentence an expression of desire, that is, an optative communicative act; and the propositional content is some punishment or revenge against the target. Hence, it is a true curse.

13.5.10 *Bad Words*

Our corpus is full of bad words that often also combine with all the foregoing communicative acts, enhancing their aggressive impact.

As can be seen, the sorts of aggressive communication found in our corpus, and their respective linguistic forms, are in large part the same as hypothesized earlier: direct and indirect insults and criticism, and curses. Further acts include aggressive requests and prohibitions, addressed to the target, and aggressive instigation, addressed to some audience against the target. The only type of communicative act that is not present is imprecation: a further demonstration of its being a somehow “egocentric” act, in which an external third entity is called for, different from the audience, and the target of aggression is only indirectly implied.

13.6 Conclusion

This chapter has defined the notion of insult, distinguishing it from other forms of aggressive communication like bad words, curses, and imprecations. While bad words are single words and are aggressive only because they convey taboo contents, the other three are whole communicative acts: a curse, one wishing a bad event or ordering a self-defeating action to a target, and further rejecting any ties with him; an imprecation is a curse or an insult to an object or external entity to which one assigns responsibility for an unlucky event. An insult is a communicative act that assigns such negative properties to a target as to finally include him in a degrading category, with the intent of offending him and spoiling his image and self-image. As tested on a corpus of interactions in TV debates and social media, these forms of aggressive communication often combine but yet can be distinguished from one another; and insults may be performed directly or indirectly, by both verbal and bodily signals. Obviously, in a situation of conflict, insults, especially direct ones, can have deleterious and irremediable effects. Yet, further studies could investigate the role of context and culture in making insults more or less offensive in relation, for example, to their indirectness. Some recent studies have shown the different forms of “rebuffering” in relation to different cultures (Lee et al. 2012).

In sum, studying the linguistic aspects of aggressive communication can help us to better understand insulters' emotions and cognitive processes and their potential differences across contexts.

References

- Castelfranchi C, Poggi I (1990) Blushing as a discourse: was Darwin wrong? In: Crozier R (ed) *Shyness and embarrassment. Perspectives from social psychology*. Cambridge University Press, New York, pp 230–251
- Castelfranchi C (1988) *Che figura. Emozioni e immagine sociale*. Il Mulino, Bologna
- Castelfranchi C (2000) Affective appraisal versus cognitive evaluation in social emotions and interactions. In: Paiva A (ed) *Affective interactions*. Springer, Berlin
- Castelfranchi C (2003) Micro-macro constitution of power. *ProtoSociology* 18–19:208–265
- Conte R, Castelfranchi C (1995) *Cognitive and social action*. University College, London
- D'Errico F, Poggi I, Corriero R (2014) Aggressive language and insults in digital political participation. In: *Proceedings of multiconference on computer science and information systems: Web based communities and social media 2014*. University of Lisboa, Lisboa, pp.105–114
- D'Errico F, Poggi I, Vincze L (2013) Discrediting body. A multimodal strategy to spoil the other's image. In: Poggi I, D'Errico F, Vincze L, Vinciarelli A (eds) *Multimodal communication in political speech. Shaping Minds and Social Action*. Springer, Berlin, pp. 181–206
- D'Errico F, Poggi I, Vincze L (2012) Discrediting signals. A model of social evaluation to study discrediting moves in political debates. *J Multimodal User In*. doi:[10.1007/s12193-012-0098-4](https://doi.org/10.1007/s12193-012-0098-4)
- Ekman P (1979) About brows: emotional and conversational signals. In: von Cranach M, Foppa K, Lepenies W, Ploog D (eds) *Human ethology: claims and limits of a new discipline: contributions to the colloquium*. Cambridge University Press, Cambridge, pp 169–248
- Galli de' Paratesi N (1969) *Le brutte parole. Semantica dell'eufemismo*. Mondadori, Milano
- Lee EAJ, Soto JA, Swim KM, Bernstein J (2012) Bitter reproach or sweet revenge: cultural differences in response to racism. *Personality and Social Psychology Bulletin* 2012, 38(7):920–932
- Miceli M, Castelfranchi C (1998) The role of evaluation in cognition and social interaction. In: Dautenhahn K (ed) *Human cognition and agent technology*. John Benjamins, Amsterdam
- Miceli M, Castelfranchi C (1989) A cognitive approach to values. *J Theory Soc Behav* 19(202):169–193
- Poggi I, D'Errico F (2012a) Social signals: a framework in terms of goals and beliefs. In: Poggi I, D'Errico F, Vinciarelli A (eds) *Social signals: from theory to applications*. Special Issue *Cogn Process*, vol 13(2):427–445
- Poggi I, D'Errico F (2012b) Pride and its expression in political debates. In: Paglieri F, Tummolini L, Falcone R, Miceli M (eds) *The goals of cognition*. *Festschrift for Cristiano Castelfranchi*. London College Publications, London, pp 221–253
- Poggi I, Zuccaro V (2008) *Admiration*. In: *Proceedings of the workshop AFFINE: affective interaction in natural environment, post-conference workshop of ICMI 2008* Chania, Crete, 24 Sept 2008
- Poggi I (2009) The language of interjections. In: Esposito A, Hussai A, Marinaro M, Martone R (eds) *Multimodal signals: cognitive and algorithmic issues*. *Lect Notes Comput Sci Ser*. Springer, Berlin, pp 170–186
- Poggi I (2004) Emotions from mind to mind. In: Paiva A, Aylett R, Marsella S (eds) *Proceedings of the workshop W13: empathic agents*. Third international conference on autonomous agent systems, AAMAS 2004, New York, 19–23 Jul 2004

- Poggi I (2005) Interacting bodies and interacting minds. Proceedings of the 2nd conference of the international society for gesture studies (ISGS) "*Interacting bodies*". École normale supérieure Lettres et Sciences humaines, Lyon-France, 15–18 Jun 2005
- Poggi I (2007) Mind, hands, face, and body. A goal and belief view of multimodal communication. Weidler, Berlin
- Poggi I (2008) Types of emotions and types of goals. In: Proceedings of the AFFINE Workshop, Chania, Greece, 24 Oct 2008
- Scardigno R, Giancaspro ML, Manuti A, Mininni G (2015) The rhetoric of conflict inside and outside the stadium: the case study of an Italian Football Cheer Group. In: D'Errico F, Poggi I, Vinciarelli A, Vincze L (eds) Conflict and multimodal communication. Social research and machine intelligence. Springer, Berlin, pp. 207–221
- Vogel C (2015) Some puzzles of politeness and impoliteness within a formal semantics of offensive language. In: D'Errico F, Poggi I, Vinciarelli A, Vincze L (eds) Conflict and multimodal communication. Social research and machine intelligence. Springer, Berlin, pp. 223–241
- Zamperini A, Menegatto M (2015) Giving voice to silence: a study of state violence in Bolzaneto prison during the Genoa G8 summit. In: D'Errico F, Poggi I, Vinciarelli A, Vincze L (eds) Conflict and multimodal communication. Social research and machine intelligence. Springer, Berlin, pp. 185–205

Part IV
Emotions and Multimodal Communication
in Conflict

Chapter 14

Multimodal Analysis of Low-Stakes Conflicts: A Proposal for a Dynamic Model

Silvia Bonacchi and Mariusz Mela

List of Abbreviations Used in the Transcription (with Cursive Examples)

Simultaneous Utterances (Overlaps)

[] Overlapping utterances
B-Andrea <<all>sag mal was habt ihr von [verantwortung]>
A-Andreas <<f>[sto:pp]>

Latching

= No interval between the end of the prior turn and the start of the next turn
B-Andrea *mir sind die trän_n gekomm_n = ich hab geheult*

Intervals Between and Within Utterances

(.) An estimated micropause of less than 0.2 s long
(-) An estimated short pause of 0.2–0.5 s long

S. Bonacchi (✉) • M. Mela
Department of Applied Linguistics, Institute of Specialised and Intercultural Communication,
Multimodal Communication Culturological Analysis (MCCA), University of Warsaw,
ul. Szturmowa 4, Warsaw 02-678, Poland
e-mail: s.bonacchi@uw.edu.pl; mariusz.mela@gmail.com

© Springer International Publishing Switzerland 2015
F. D'Errico et al. (eds.), *Conflict and Multimodal Communication*,
Computational Social Sciences, DOI 10.1007/978-3-319-14081-0_14

Intonation Contours at Turn Completion

- ? Rising intonation
- ↑ A mid-turn sharp rise in intonation
- ↓ A mid-turn sharp fall in intonation
- <h> High tone of voice

Characteristics of Speech Delivery

- : A colon indicates extension of the preceding sound or syllable, e.g., *tota:l*
- IST Capital letters indicate increased loudness for a focus, e.g., *KINDERzimmer*
- _ Contraction, e.g., *mir sind die trän_n gekomm_n*

The Dynamics of Speech Delivery

- <all> Fast manner of speaking
- <acc> A speaker starts speaking faster
- <f> A loud manner of speaking
- <ff> A very loud manner of speaking

14.1 Theoretical Issues

The aim of this chapter is to propose a dynamic model for the multimodal analysis of conflict situations caused by impoliteness and verbal aggression.

When we talk about conflicts, we have to distinguish between high-stakes conflicts (as underlined by Ellen Giebels in her talk, <http://klewel.com/conferences/sspnet-roma-013/index.php?talkid=20>, see also Giebels et al. 2014), caused by serious divergences, and low-stakes conflicts, which are common in everyday life and are often caused merely by an improper dialogical attitude. Impoliteness and verbal aggression are in effect forms of communicative behavior marked by a struggle for (interactional) power, which leads to a lack of cooperation (according to Grice 1975: 45) and thus to conflict.¹

¹In this sense, we can outline a second-order framework, where politeness is seen as cooperative behavior (Grice), and impoliteness is seen as uncooperative (non-dialogic) behavior, which may lead to conflict and communication problems (communicative standstill, a fight, etc.). See in this sense Lachenicht (1980), Culpeper (1996, 2008, 2011) Culpeper et al. (2003), Spencer-Oatley (2005), Bousfield (2008), for an overview see: Bonacchi (2013): 80ff.

As is well-known, conflict analysis is performed in many disciplines (psychology, linguistics, social sciences, etc.). Each of them has worked out a definition of conflict which should be operationalized in a specific investigation with specific scientific aims. Some disciplines (like psychology) are, among others, interested in finding solutions and reconciliation strategies, while other disciplines (like linguistics) attempt to develop a reliable method for the description and analysis of conflict, and thus to explicate the underlying conflict mechanisms at the level of the dynamics of a communicative interaction. From the pragmalinguistic perspective, the investigation of conflict is focused on the “action-leading” character of speech acts, i.e., on three dimensions related to the performed acts: (a) doing something while saying something (the *locutionary dimension*, what is *said*), (b) doing something in saying something (the *illocutionary dimension*, what is *meant*), (c) doing something by saying something (the *perlocutionary dimension*, which *effects* in the world are caused by linguistic utterances), as Austin² put it. Both impoliteness and verbal aggression end easily in conflict because in both cases they lead to a limitation of freedom of action of one interlocutor to the other interlocutor’s advantage, who in turn claims interactional power (interactional rights) for himself/herself. We will try to show that the central event in conflicts caused by impoliteness and verbal aggression is not only the threat to face (the actual loss of face or the danger of losing face) and the reaction to it, but rather the process of a redefinition of interactional balance between interlocutors.

The notion of face used in our model is a dynamic one. In pragmalinguistic studies on politeness and impoliteness based on E. Goffman’s notion of face (Goffman 1967), later developed by P. Brown and S.C. Levinson in a second-order framework, the face of a person is supposed to be a static element according to the classic definition: “The term *face* may be defined as the positive social value a person effectively claims for himself by the line others assume he has taken during a particular contact. Face is an image of self delineated in terms of approved social attributes [. . .]” (Goffman 1967: 5). The term *face-work* designs the behavior that every participant in social encounters has to assume in order to maintain both his/her face and the face of others (Brown and Levinson 1987: 61ff.), according to the two fundamental human needs: the need to be appreciated (*positive face*) and the need not to be limited (*negative face*) (Brown and Levinson 1987: 101ff. and 129ff.). Politeness is seen as a regulative system of ritual behavior, which has a

²See Austin (1962): 94: “consider from the ground up how many senses there are in which to say something *is* to do something, or *in* saying something we do something, and even *by* saying something we do something.” From the point of view of pragmalinguistics, it is not possible to analyze a conflict exclusively at the locutionary (lexical) level, because every utterance (even an insulting one from the point of view of the lexical meaning) can be used in a supporting way (s. Mateo and Yus 2013: 94). Furthermore, the analysis of the sole locutionary level proves unsatisfactory in the analysis of cold and hidden aggression and tacit conflicts (see “acid speech acts” in Poggi and D’Errico 2013). The phenomenon can be investigated empirically as a perlocutionary effect of communicative behavior (verbal and nonverbal behavior), which is, or is not, intended as conflictive.

prophylactic function of preventing conflicts and ensuring social action. Referring to Emil Durkheim's conception of sacrum (the untouchable) in *The Elementary Forms of Religious Life* (1915), Goffman affirms the thesis that everyday rituals have the function of reintroducing sacrum into everyday life. In Durkheim's theory, the sacrum represents the values of a group which are "embodied" in certain symbols and behaviors (*rituals*), which express the values related to the dignity of a person—his/her right to be respected, i.e., to be listened to and appreciated. Thus it can be concluded that everyday sacrum presupposes a *ritual order*, which gives stability to social encounters and helps to maintain social peace. The politeness system constitutes an instrument for giving stability to the ritual order in a given society, that is, to keep a ritual balance that expresses and maintains the cultural values which are recognized as being essential.

In our model, we propose a dynamic conception of interactional (and ritual) balance (Bonacchi 2013: 97ff.), in which face needs and face expectations of both interlocutors are correlated with other factors. Its dynamics lies in the possibility to properly describe the action as a whole (in its temporal development on a temporal axis) as well as in paying special attention to the interdependence of various factors that we define according to our research needs.

The diagram below (Fig. 14.1) illustrates the core elements of interactional balance: face needs and ritual expectations, the distribution of interactional space, and the allocation of discursive roles and positions in a given situational context and within particular power relations. The dynamics of the notion of interactional balance lies in the consideration that face needs grow throughout an interaction, they are not static, but instead they condition each other and synchronically develop on a temporal axis. The notion of interactional balance allows researchers to capture the way interlocutors create their faces and the expectations they have towards each other in terms of interdependence. They share a *distributed responsibility* (Jacoby and Ochs 1995: 177) for the success of the interaction, a notion according to which it is not possible to identify a clear "fault" for a conflictive situation, but always a correlation of hardly reconcilable communicative and interactional goals.

In this sense, our point of view is a systemic one (Simon 2010): we consider the relationship between interlocutors as well as that between interlocutors and the background (the environment of the interlocutors) as a system, which includes interactants, interactional frame, context, etc. Interactants act simultaneously: they *co-construct* the interaction in a certain given field of reciprocal dependencies.

In a communicative interaction, each interactant has communicative goals (what meanings he/she wants to convey and how) and interactional goals (how he/she defines and constructs his/her relationship to the other in the common communicative space). For our purposes we define conflict³ as a dynamic process

³In scientific literature there are many definitions of conflicts which are compatible with our point of view. Mack and Snyder (1973) define a conflict as a temporal disjunction in the flow of an interaction. They define the following characteristics of conflictive situations: Conflicts involve at least two people (parties) and are a consequence of the position somebody is in and/or a shortage of resources. A conflictive situation is aimed at destroying, harming, frustrating, or controlling the

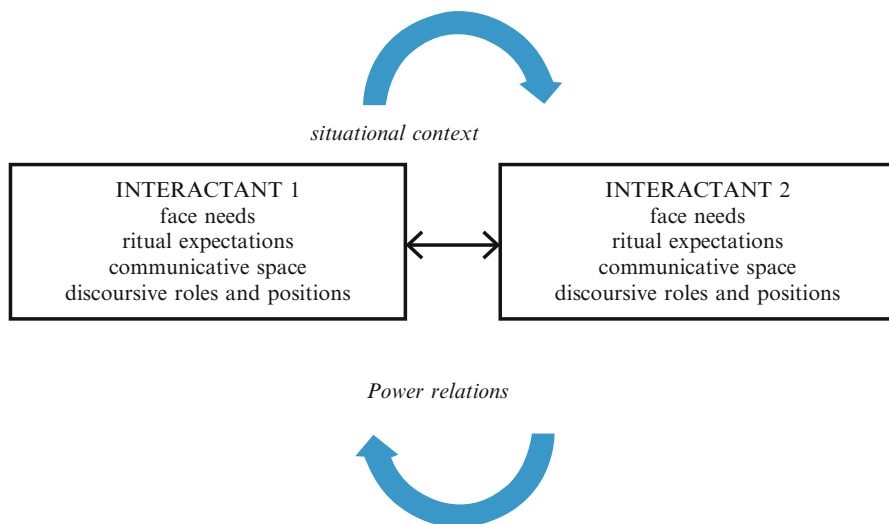


Fig. 14.1 Interactional balance

in which interactants act antagonistically, because each of them attempts to achieve communicative and interactional goals that compete with/are in opposition to the goals of the other.⁴ A goal can be here defined as “a state that regulates the actions of a system” (Poggi 2007: 13). The divergent goals of the interactants cause a change in the state of the system (conflict process). Goals are pursued by making use of external and internal resources and realizing a certain plan, which is a set of actions respectively aimed at a set of hierarchically arranged goals. In a communicative interaction, interactants have *primary* and *secondary goals*, and *instrumental* and *terminal* goals (Poggi 2007: 14). In a conflictive situation a speaker’s instrumental goal to make the other share his/her primary goals and to achieve a given state is intertwined with the terminal goal of both interactants to affirm their faces and their interactional power.⁵

A very important aspect in understanding the dynamics of the balance in an interaction is the notion of “interactional power.” According to Wartenberg’s

other party in some way. In a conflictive situation, a party can achieve its goal only at somebody else’s expense. This is why conflicts constitute a temporal disjunction of interactional flow between interactants. According to Fiehler (1986) a conflict is a serious and unacceptable disappointment of expectations, a violation of interests or a threat to a person’s identity. Shantz (1987) notices that a state of conflict denotes incompatible behaviors or goals. In Galtung’s (1972) opinion a conflict exists in an operating system when within it two or more incompatible target states are sought.

⁴See for example Bousfield (2008): 132, Locher and Bousfield (2008): 8f.

⁵An example: in a conflict about how to educate children, the instrumental goal can be to make the Other share a conduct of behavior (for example to execute punishment for committed violations of rules), the terminal goal can be to affirm one’s authority as a parent.

definition (1990: 85): “A social agent *A* has power over another social agent *B* if and only if *A* strategically constrains *B*’s action environment,” thus in every interaction we have a process of allocating power. Thus the output hypothesis is that verbal aggression and impoliteness can be⁶ expressions of non-dialogic communicative behavior aimed at gaining power, forcing a certain interactional dynamics by rejecting the other and denying him/her conversational rights. Non-dialogic behavior means a refusal to negotiate, and the exercising of power or even violence to affirm one’s goals. Such antidialogic behavior is realized in a multimodal way: through verbal and nonverbal behavior, prosody, facial expression, and gesture.

Given the above considerations, we can think of a conflict as an interactional process characterized by certain stages: a preparatory phase (onset phase), a culminating phase (stroke) which can be further structured in chains, and a retraction/solution phase (offset phase).⁷

ONSET_K₀ → DEVELOPMENT_K₁ → STROKE_K_N ↑
DEVELOPMENT_K → OFFSET (RETRACTION/SOLUTION)_K → L₀

From an initial balance situation which is corrupted by conflict and is characterized by a struggle (tension) we come, in the last of the phases to a new balance, which is a redefinition of communicative balance (it can be negotiation, communicative standstill, the use of physical violence, etc.):

INITIAL BALANCE.....TENSION.....NEW BALANCE

Within these stages, interactants seek to construct their faces according to the possibilities given by the other and by the situation⁸:

FACE A₁/B₁....FACE A₂/B₂....FACE A_N/B_N

This can be clearly seen in the analysis of the empirical material presented and analyzed below in our pilot study.

⁶Of course there are other explanations for the use of impoliteness, for example, asymmetries in communicative competence, but in this chapter we will focus on the use of impoliteness provoked by an attempt to gain power.

⁷For the choice of this terminology see Sager (2005). See also Müller (1998) and Fricke (2007).

⁸See also Simmel (1972). Simmel considers a conflict to be even a chance for the development of social harmony. In his opinion, a conflict involves always a possibility of solving and a mutual will to solve it. In this sense, the stage before managing a conflict is one that actually divides the conflictive parties. As soon as conflict management begins, the first step towards cooperation is made.

14.2 Case Analysis

In order to properly show the feasibility of our model, we will analyze a video of a conflictive communicative interaction where conflict is provoked by a clash of face needs and by the struggle to gain interactional power. The scene is an extract from a reality show and thus a *mediatized interaction* (a television programme). Mediatized interactions present some peculiarities regarding the relationship between interactants and interaction frames. The interactants operate on two intersecting levels: the proper interaction and the interaction with a virtual audience through a medium (in this case television). The interactants behave as if they were on the stage of a theater: they perform for the public in the studio and construct their image with a strong awareness of being watched by the public at home. In this way some aspects of theatricality (Goffman 1959: 29ff.; Schmitt and Deppermann 2009) are more strongly accentuated than in non-mediatized situations. In the interaction presented here, a conflictive process begins when a person acts towards another person in an aggressive way by threatening his/her face, denying him/her interactional rights, and limiting his/her action environment.⁹

The analysis¹⁰ was conducted with the annotation programme ELAN (Eudico Linguistic Annotator), which can be downloaded from the website of The Language Archive of the Max Planck Institute for Psycholinguistics (<http://tla.mpi.nl/tools/tla-tools/elan/download/>). The programme enables the researcher to capture the multimodality of an interaction due to the possibility of using any number and any kind of tiers and thus analyze any aspect of a recording, e.g., verbal language, motion, gestures, facial expressions, and prosody. The transcription is mainly based on the GAT2 system (Gesprächsanalytisches Transkriptionssystem 2, s. Selting et al. 2009) and was carried out by means of the computer programme Folker (<http://agd.ids-mannheim.de/folker.shtml>), developed at the Archive for German Language in Mannheim, Germany. The programme makes it possible to segment and transcribe an audio recording and save it as a file that can be opened in the computer programme ELAN. The presented annotation¹¹ allows a researcher to show how verbal and nonverbal elements interact, and how relevant cues for conflict dynamics are realized in a multimodal way (Fig. 14.2).

⁹In this video we have a conflict with manifest face threats. Other types of conflict (e.g. caused by hidden forms of hostility, such as in an academic discussion) will not be discussed in this chapter.

¹⁰This study was conducted within the research project MCCA (Multimodal Communication: Culturological Analysis, www.mcca.uw.edu.pl) performed at the University of Warsaw, Department of Applied Linguistics thanks to a grant from the NCN (Polish National Research Center, UMO-2012/04/M/HS2/00551).

¹¹The description levels have been defined according to the principles set out in Schneider and Stöckl (2011): 28f and Schmitt (2005).

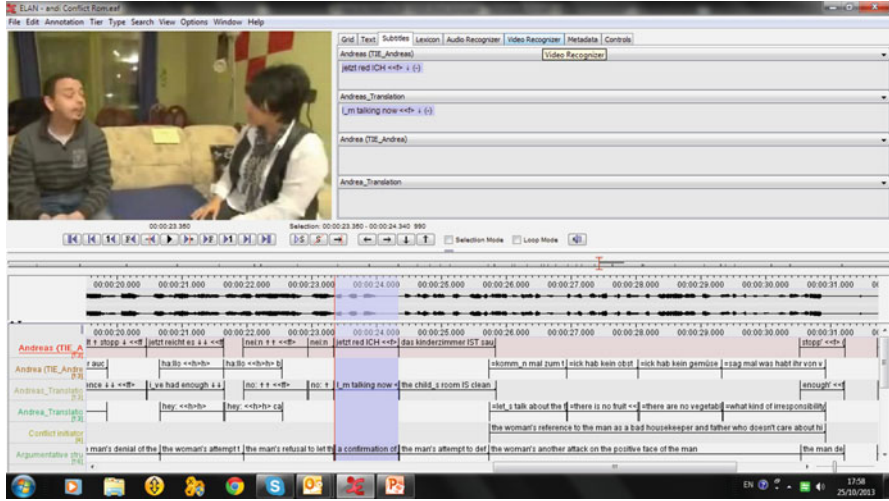


Fig. 14.2 ELAN-Window with the selected tiers. Source: *Der nette Andi beim Frauentausch* (The original video (<http://www.youtube.com/watch?v=IsLrTvc5Hrw>) is 00:02:55:00 long. For our analysis we have chosen the scene from 00:17:80 to 01:05:00 with an internal cut (from 00:43:74 to 00:51:260)). Mediality: The video is a scene from a popular German reality show (*Frauentausch*, “Wife Swap”), where one woman exchanges her household with another woman and lives in the other’s house for 10 days as an experiment. Interaction frame: The selected scene presents the first encounter between the host and the woman who is temporarily replacing his wife. Interactants: The two interactants are a man (Speaker A, Andreas) and a woman (Speaker B, Andrea) (It appears that the aggression was elicited by Andi (Andreas), who seemingly had the task of attacking the woman and inducing verbal violence)

TRANSCRIPTION—(from the original video: 00:17:80 to 01:05:00) with English subtitles

A-Andreas	00:00:00.298–00:00:01.939	und gefällt dir alles? <<<↓> oder gibt_s hier irgendwas >
A-Andreas_Translation		<i>and do you like it at our place or is there anything you don’t like?</i>
B-Andrea	00:00:01.939–00:00:03.589	ick finde dass hier total (-)
B-Andrea_Translation		<i>I find the flat completely ...</i>
B-Andrea	00:00:03.989–00:00:05.178	das is ein dreckstall für mich (-)
B-Andrea_Translation		<i>it looks like a pigsty to me</i>
B-Andrea	00:00:05.685–00:00:06.132	ich wes nich (.)
B-Andrea	00:00:06.318–00:00:07.680	wie könnt ihr hier n kind großzieh_n (.)

<i>B-Andrea_Translation</i>		<i>I cannot imagine bringing up a child in here</i>
B-Andrea	00:00:07.680–00:00:08.977	ich kam in dat kinderzimmer rin (.)
<i>B-Andrea_Translation</i>		<i>I entered the child's room</i>
B-Andrea	00:00:08.977–00:00:10.474	<<acc> dat war für mich der blanke chaos >
<i>B-Andrea_Translation</i>		<i>it was in a terrible mess</i>
B-Andrea	00:00:10.573–00:00:11.528	mir sind die trän_n gekomm_n=
B-Andrea	00:00:11.528–00:00:12.269	=ick hab geheult
<i>B-Andrea_Translation</i>		<i>my eyes started tearing up, I started crying</i>
A-Andreas	00:00:12.269 – 00:00:13.766	so (.) jetzt muss ich dich unterbrechen
<i>A-Andreas_Translation</i>		<i>I'm afraid I have to interrupt you now</i>
A-Andreas	00:00:13.766–00:00:14.763	das KINDERzimmer (.)
<i>A-Andreas_Translation</i>		<i>about the child's room</i>
A-Andreas	00:00:14.763–00:00:16.060	hast du mal auf die gröÙe gesehen?
<i>A-Andreas_Translation</i>		<i>haven't you noticed how big it is?</i>
A-Andreas	00:00:16.360–00:00:18.554	hast die menge an spielzeug äh (.) gesehen alles?
<i>A-Andreas_Translation</i>		<i>or how many umm toys there are?</i>
B-Andrea	00:00:18.554–00:00:18.997	du (.) [aber
B-Andrea	00:00:18.997–00:00:19.317	[könnt mir auch]
B-Andrea	00:00:19.317–00:00:19.851	den
<i>B-Andrea_Translation</i>		<i>but let me</i>
A-Andreas	00:00:18.997–00:00:19.317	[und]
<i>A-Andreas_Translation</i>		<i>and ...</i>
A-Andreas	00:00:19.317–00:00:19.851	<<ff> ha:lt ↑
A-Andreas	00:00:19.851–00:00:20.450	sto:pp ↓ >
<i>A-Andreas_Translation</i>		<i>keep quiet or ...</i>
A-Andreas	00:00:20.450–00:00:21.647	[<<ff> jetzt re:icht es >]

<i>A-Andreas_Translation</i>		<i>I've had enough!</i>
B-Andrea	00:00:20.450–00:00:21.647	[<<h> ha:llo >]
<i>B-Andrea_Translation</i>		<i>hey!</i>
B-Andrea	00:00:21.647–00:00:22.106	<<h> ha:llo >
<i>B-Andrea_Translation</i>		<i>hey!</i>
B-Andrea	00:00:22.106–00:00:22.945	[bleib]
<i>B-Andrea_Translation</i>		<i>calm . . .</i>
A-Andreas	00:00:22.106–00:00:22.996	[<<ff> ne:in ↑ >]
<i>A-Andrea_Translation</i>		<i>no!</i>
A-Andreas	00:00:22.996–00:00:23.386	<<ff> ne:in ↑ >
<i>A-Andreas_Translation</i>		<i>no!</i>
A-Andreas	00:00:23.386–00:00:24.426	<<f> jetzt red ICH ↓ >
<i>A-Andreas_Translation</i>		<i>I'm talking now!</i>
A-Andreas	00:00:24.426–00:00:25.737	das KINDERzimmer IST sauber
<i>A-Andreas_Translation</i>		<i>the child's room is clean!</i>
B-Andrea	00:00:25.737–00:00:26.884	komm_n mal zum thema essen=
<i>B-Andrea_Translation</i>		<i>let's talk about the food</i>
B-Andrea	00:00:26.884–00:00:28.031	=ick hab kein obst gefund_n=
<i>B-Andrea_Translation</i>		<i>there is no fruit</i>
B-Andrea	00:00:28.031–00:00:29.176	=ick hab kein gemüse gefund_n=
<i>B-Andrea_Translation</i>		<i>there are no vegetables</i>
B-Andrea	00:00:29.176–00:00:30.562	<<all>sag mal was habt ihr von verantwortung>
<i>B-Andrea_Translation</i>		<i>what kind of irresponsibility do you</i>
A-Andreas	00:00:30.562–00:00:30.885	<<f> sto:pp >
<i>A-Andreas_Translation</i>		<i>enough!</i>
A-Andreas	00:00:32.005–00:00:33.895	<<ff> sonst kriegst du gleich ein_auf die schnauze >
<i>A-Andreas_Translation</i>		<i>or you'll get a punch in the face!</i>

An analysis of the conversational structure allows one to reconstruct the development of the conflict process and the strategy of redefining the interactional balance. At the beginning of the scene, the man is being friendly; he expresses his care

about the woman's well-being. He has a friendly tone of voice and applies rising intonation ("und gefällt Dir alles?"). He is looking at the woman and is seeking eye contact. His communicative behavior can be described as cooperative. However, his intonation falls gradually in the second part of the question (<<↓>oder gibt_s hier irgendwas >), which can be interpreted as doubt or a sudden loss of self-confidence.



00:00:00.298-00:00:01.939 A-Andreas: und gefällt dir alles?<<↓>oder gibt_s hier irgendwas>(and do you like it at our place or is there anything you don't like?)



00:00:01.939-00:00:05.178 B-Andrea: ick finde dass hier tota:l (—) das is ein dreckstall für mich (—) (*I find the flat completely . . . it looks like a pigsty to me*).

The anacoluthon (disruption of a sentence): "ick finde dass hier tota:l (—) (*I find the flat completely*) can be interpreted as the woman's initial loss for words as she seeks to describe her amazement, followed by a self-correction process in which she chooses highly expressive derogatory vocabulary: "das is ein dreckstall für mich (—)" (*it looks like a pigsty to me*). The woman is looking downwards as she pronounces her verdictive utterances (Austin 1962: 150ff.). There is no eye-contact between the interlocutors. By implying that the owners are untidy people, speaker B attacks the positive face of speaker A, and at the same time she affirms her right to judge him with a verdictive, i.e., she expresses her will to assume interactional authority. Speaker A follows the words of Speaker B with growing nervousness and agitation, which becomes visible in his tense facial expression. In the beginning he seems to be perplexed and confused by the woman's hostile behavior, he doesn't seem to understand what is going on. He only acts nonverbally.

00:05.556: A expresses his incredulity with his gaze



Speaker A's facial muscles are tense, speaker B avoids looking at A in the eyes. Speaker B reinforces the attack with the rhetorical question: "ich wes nich (.) wie könnt ihr hier n kind großzieh_n (-)" (*I cannot imagine bringing up a child here*), which is an aggressive insinuation that Speaker A and his wife are bad parents.



00:00:05.685-00:00:12.269 B-Andrea: ich wes nich (.) wie könnt ihr hier n kind großzieh_n ich kam in dat kinderzimmer rin (.)<<acc>dat war für mich der blanke chaos>mir sind die trän_n gekomm_n = ick hab geheuelt (*I cannot imagine bringing up a child in here. I entered the child's room, it was in a terrible mess, my eyes started tearing up, I started crying.*)

The woman pronounces short staccato sentences which make her story more dramatic and reinforce her credibility as a caring mother and housekeeper and thereby justify her claim for authority. This deliberate construction of her image in front of the interlocutor and in front of the public at home is stressed by the next few sentences which lead to a dramatic climax that stresses her emotional involvement. By describing the child's room as being in extremely bad condition: "<<acc>dat war für mich der blanke chaos>" (*it was in a terrible mess*), the woman portrays herself as a caring sensitive mother: "mir sind die trän_n gekomm_n = ick hab geheuelt" (*my eyes started tearing up, I started crying*).

The climax in the woman's choice of derogatory expressions shows that she is not paying attention to the man's feelings. In this sense, she is not cooperative in that she violates dialogical principles (by ignoring the listener's backchannel signals and acting in a self-referential way). These repeated attacks by B on the positive face of A are stressed by her gestures and facial expressions. Repeated right-to-

left movements with her head (shakes) express her amazement and hopelessness, one short up- and downward movement with her shoulders (a shrug) expresses her disbelief.

When speaker B attacks speaker A and his wife with a serious accusation of being bad parents, A begins to shake his head. As shown in the extract presented below, his head shaking—the man is turning his head left and right along the transverse plane repeatedly in quick succession from extreme kin A to extreme kin B—follows the rhythm of the words uttered by Speaker B, probably to indicate disagreement or denial of what the woman is saying. This can be interpreted as a form of automatic response or even as an accommodation process (an “alignment-related process” according to Karpinski et al. 2014) of head movements with phonetic prominences of Speaker B:



00:10:080 – extreme kin A

00:10:200 – extreme kin B



The man seems unable to answer, he merely listens in silence to what the woman is saying, nonverbal signals give information about his emotional state of helplessness and confusion. An active reaction (proactive behavior) of speaker A, who up to this point seems to be merely a helpless victim of speaker B, comes suddenly with the utterance: “so (.) jetzt muss ich dich unterbrechen” (*I’m afraid I have to interrupt you now*), which can be interpreted as an attempt to force the

woman to interrupt her story in order to prevent her from further attacks on the man's positive face.



00:00:12.269-00:00:18.554 A-Andreas: so (.) jetzt muss ich dich unterbrechen = das KINDERzimmer (–) hast du mal auf die grÖÙe gesehen? hast die menge an spielzeug äh (.) gesehen alles? (*I'm afraid I have to interrupt you now. About the child's room ... haven't you noticed how big it is? Or how many umm toys there are?*)

The man's sudden interruption of the woman's flow of speech is marked by a metacommunicative utterance (“so (.) jetzt muss ich dich unterbrechen”) (*I'm afraid I have to interrupt you now*), forcing a resumption of the topic of the child's room as a left dislocation and an anaphoric reference in the main proposition (“das KINDERzimmer (–) hast du mal auf die grÖÙe gesehen? hast die menge an spielzeug äh (.) gesehen alles?”) (*about the child's room ... haven't you noticed how big it is? Or how many umm toys there are?*). The set of short (rhetorical) questions are probably intended as a counterattack with the aim of making the woman stop talking and at the same time of showing the public at home that speaker B has no right to judge him. Yet the man's tone of voice still reveals an effort to control himself. The man's raised eyebrows and repeated up-and-down movements of the head have an iconographic function of stressing his verbal utterances and at the same time stressing the prominences in his intonation, which are aligned with the beats in the gestures. The falling intonation at the end of the sentence “hast du mal auf die grÖÙe gesehen?” (*about the child's room ... haven't you noticed how big it is?*) indicates that the sentence is intended as a reproach.

The attempt by speaker B to take her turn (“du (.) [aber könnt mir auch den]” (*but let me ...*)) fails because speaker A interrupts her in a very aggressive way with short, loud imperatives: “<<ff>ha:lt↑ sto:pp↓>” (*keep quiet or ...*) and with the expressive utterance: “<<ff>jetzt [re:icht es↓>]” (*I've had enough*), in which speaker A expresses his feelings and his uneasiness about the situation. In the man's decisive up-and-down movement with his head when uttering the sentence “<<ff>jetzt [re:icht es>],” (*I've had enough!*) we notice an alignment of prosodic prominences with gestural prominences.



00:00:19.317-00:00:19.851	A-Andreas: <<ff>ha:lt ↑
00:00:19.851-00:00:20.450	A-Andreas: sto:pp ↓>(keep quiet or ...)
00:00:20.450-00:00:21.647	A-Andreas: [<<ff>jetzt re:icht es >] (I've had enough!)
00:00:20.450-00:00:21.647	B-Andrea: [<<h>ha:llo >] (hey!)
00:00:21.647-00:00:22.106	B-Andrea: <<h>ha:llo>(hey!)
00:00:22.106-00:00:22.945	B-Andrea: [bleib] (calm ...)
00:00:22.106-00:00:22.996	A-Andreas: [<<ff>ne:in ↑ >] (no!)
00:00:22.996-00:00:23.386	A-Andreas: <<ff>ne:in ↑>(no!)
00:00:23.386-00:00:24.426	A-Andreas: <<f>jetzt red ICH ↓>(I'm talking now!)

The woman's further attempts to calm the man down and to claim her turn by repeating with elevated pitch a phatic "[<<h>ha:llo>]" (hey!) twice fail, because the man denies her the right to speak with a repeated "[<<ff>ne:in ↑ >]" (no!). In doing so, the woman uses a high tone of voice to make her words audible. The man's double movement with his head from right to left accompany the word "[<<ff>ne:in ↑ >]" (no!); he is turning his face away from the woman as a proxemic signal of refusal. In his fit of anger, speaker A becomes very aggressive and claims for himself the exclusive right to speak: "<<f>jetzt red ICH ↓>" (I'm talking now!), which is immediately followed by an assertive: "das KINDERzimmer IST sauber" (the child's room is clean!), with a prominence on "IST" (is). The man's denial of the woman's right to speak corresponds to claiming his own right to do so. His eyebrows are raised which corresponds to his raised voice, and he keeps his eyes closed. The man's thumb is pointed in the direction of the child's room as a deictic movement which gives spatial concreteness to the utterance of speaker A.



00:00:29.176-00:00:30.562 B-Andrea: <<all>sag mal was habt ihr von verantwortung>(what kind of irresponsibility do you . . .)

With her few acts of verbal and nonverbal behavior speaker B tries to reaffirm her interactional rights and above all her wish to exercise conversational power. The roles are inverted: she begins to take a more reactive position and she answers reactively. She attempts to impose a second topic in her critique of speaker A: “komm_n mal zum thema essen=” (*let’s talk about the food*)—the topic of food. In this way, she reinforces her attack on him with a stronger face threat and an attempt to allocate interactional power in her favor. The woman stretches out her hand towards the man to draw his attention to a different topic. She attempts to present the man as being a bad housekeeper and father who doesn’t care about his child’s health. This renewed attack on the man’s positive face starts with assertive comments: “=ick hab kein obst gefund_n = ick hab kein gemüse gefund_n=” (*there is no fruit, there are no vegetables*) and follows immediately after that with an aggressive reproach. In the end, she casts doubt on the sense of responsibility of speaker A and his wife as parents: “=<<all>sag mal was habt ihr von verantwortung>” (*what kind of irresponsibility do you . . .*). While suggesting irresponsible behavior on the man’s part, speaker B shakes her head to signalize hopelessness. At this moment the man denies the woman the right to speak with a sudden, loud imperative: “<<f>sto:pp>” (*enough!*) and in order to prevent her from further reproaches, he stretches his arms apart in a position of radial abduction, with his voice raised. This leads to a sudden interruption of eye-contact and the man threatens her with the use of physical violence: “<<ff>sonst kriegst du ein_auf die schnauze>” (*or you’ll get a punch in the face!*). The threat is pronounced with falling intonation, which in German is typical of this kind of sentences. He stands up in anger and leaves the room slamming the door behind him.



00:00:30.562-00:00:30.885 A-Andreas: <<f>sto:pp><<ff>sonst kriegst du gleich ein_auf die schnauze>(enough! Or you’ll get a punch in the face!)

14.3 Selected Results

When resuming the crucial phases of the conflict process, we can distinguish the following conflict development phases:

ONSET_K₀ → development_K₁ → development_K_n → STROKE_K ↑ → OFFSET (retraction/solution)_K → 0

FACE A₁/B₁ → FACE A₂/B₂ → FACE A₃/B₃ → FACE A₄/B₄ → FACE A_N/B_N → FACE A_X/B_X

INTERACTIONAL BALANCE1 → TENSION → INTERACTIONAL BALANCE2

We can distinguish various particular verbal and nonverbal cues for every phase of the conflict. In order to analyze the cues in the verbal, vocal, and kinetic displays (cf. Sager 2005: 10ff.) in the single conflict phases and conflict dynamics, we will analyze some statistical data that we were able to extract via ELAN. It is clear that these data can be meaningful if related to a corpus, in which interactions can be compared.¹²

14.3.1 Verbal Display

Speaker A: Andreas				
Annotation	Occurr.	Aver. Dur.	Time ratio	Latency
und gefällt dir alles oder gibt_s hier irgendwas ?	1	1.641	0.04713350183823529	0.298
so (.) jetzt muss ich dich unterbrechen	1	1.497	0.04299747242647059	12.269
das KINDERzimmer	1	0.997	0.028636259191176468	13.766
hast du mal auf die gröÙe gesehen?	1	1.297	0.037252987132352935	14.763

¹²“Occurr.[ence]” indicates the number of occurrences (contiguous annotations containing the same values); “Aver.[age] Dur.[ation]” defines the total duration of the annotations with the same values divided by the number of occurrences; “Time Ratio” defines the total duration of the annotations containing the same value in the observation period; “Latency” defines the time interval between the beginning of the observation period and the first occurrence of an annotation.

hast die menge an spielzeug äh (.) gesehen alles?	1	2.194	0.06301700367647058	16.36
und	1	0.32	0.009191176470588236	18.997
ha:It!	1	0.534	0.015337775735294117	19.317
sto:pp!	1	0.599	0.017204733455882353	19.851
jetzt re:icht es!	1	1.197	0.034380744485294115	20.45
ne:in!	2	1.28	0.03676470588235294	22.106
jetzt red ICH!	1	1.04	0.029871323529411763	23.386
das KINDERzimmer IST sauber	1	1.311	0.03765510110294117	24.426
sto:pp!	1	0.323	0.00927734375	30.562
sonst kriegst du gleich ein_auf die schnauze	1	1.89	0.054285386029411756	32.005

Speaker B: Andrea

Annotation	Occur.	Aver. Dur.	Time ratio	Latency
ick finde dass hier tota:l	1	1.75	0.05026424632352941	1.94
das is ein dreckstall für mich	1	1.41	0.0404986213235294	3.97
ich wes nich (.) wie könnt ihr hier n kind großzieh_n	1	0.447	0.01283892463235294	5.685
ich kam in dat kinderzimmer rin (.)	1	1.297	0.037252987132352935	7.68
dat war für mich der blanke chaos	1	1.497	0.04299747242647059	8.977
mir sind die trän_n gekomm_n = 1		0.955	0.027429917279411763	10.573
=ick hab geheult	1	0.741	0.02128331801470588	11.528
du (.) aber könnt mir auch den	1	0.443	0.012724034926470588	18.554
ha:llo	2	1.197	0.034380744485294115	20.45
bleib	1	0.831	0.02386833639705882	22.099
komm_n mal zum thema essen=	1	1.147	0.032944623161764705	25.737
=ick hab kein obst gefunden_n = 1		1.147	0.032944623161764705	26.884
=ick hab kein gemüse gefunden_n = 1		1.145	0.032887178308823525	28.031
=sag mal was habt ihr von verantwortung?	1	1.386	0.03980928308823529	29.176

We classified the types of sentences and speech acts as follows:

Types of sentences and speech acts	Speaker A: Andreas	Speaker B: Andrea
Interrogative questions	hast die menge an spielzeug äh(.) gesehen alles?	sag mal was habt ihr von verantwortung?
Rhetorical questions	hast du mal auf die größte gesehen? hast die menge an spielzeug äh (.) gesehen alles?	ich wes nich (.) wie könnt ihr hier n kind großzieh_n? sag mal was habt ihr von verantwortung?
Assertives	so (.) jetzt muss ich dich unterbrechen das KINDERzimmer jetzt re:icht es! jetzt red ICH! das KINDERzimmer IST sauber	das is ein dreckstall für mich ich wes nich (.) wie könnt ihr hier n kind großzieh_n ich kam in dat kinderzimmer rin (.) dat war für mich der blanke chaos mir sind die trän_n gekomm_n = = ick hab geheult komm_n mal zum thema essen= =ick hab kein obst gefunden= =ick hab kein gemüse gefunden=
Expressives	–	ick finde dass hier tota:l mir sind die trän_n gekomm_n=
Directives	ha:lt! sto:pp! (2) jetzt re:icht es! jetzt red ICH!	–
Exhortatives	–	komm_n mal zum thema essen=
Phatics	–	ha:llo! (2)
Negations	ne:in! (2)	=ick hab kein obst gefund_n= =ick hab kein gemüse gefund_n=
Threats (commissives)	sto:pp! sonst kriegst du gleich ein_auf die schnauze	–
Anacoluthons and beginning signals	Und	ick finde dass hier tota:l bleib

Here we can see that speaker A uses numerous assertives and directives, negations, and one threat, which show a high struggle potential or desire to defend himself. He does not use phatic expressions. In the case of speaker B, assertives prevails, in which derogatory vocabulary (*Dreckstall/pigsty*, *der blanke Chaos/a terrible mess*) can be interpreted as expressions of her negative attitude to speaker A and his world. Speaker B uses expressive speech acts twice to manifest her emotional involvement in the encounter. Frequent anacoluthons express her inability to speak, her difficulty in finding the right words to express her outrage and indignation with proper expressiveness. They stress the dramatic nature of the scene. She uses phatic expressions which show her willingness to negotiate or her attempt to present herself as a person who is willing to negotiate. Both speakers use negations (*nein/no*, *kein Obst/no fruit*, *kein Gemüse/no vegetables*) to express their clear conflict position to the other (s. Ogden 2006; 1763).

The struggle for interactional power is clear also in the turn-taking management. In this short scene we have two cases of overlap and a proper interruption, which correspond to a denial of the right to speak (00:18:528-00:19:828). The overlaps are set in the stroke phase and in the offset phase. The overlap in the stroke phase (00:19:343-00:22:530) indicates a struggle to take a turn:

B-Andrea	00:00:18.554	du (.) [aber könnt mir auch] den
<i>B-Andrea_Translation</i>		<i>but let me</i>
A-Andreas	00:00:18.997	[und]
<i>A-Andreas_Translation</i>		<i>and</i>
A-Andreas	00:00:19.317<<ff>ha:lt ↑ sto:pp ↓ >	
<i>A-Andreas_Translation</i>		<i>keep quiet or ...</i>
A-Andreas	00:00:20.450	[<<ff>jetzt re:icht es >]
<i>A-Andreas_Translation</i>		<i>I've had enough!</i>
B-Andrea	00:00:20.450	[<<h>ha:llo >]
<i>B-Andrea_Translation</i>		<i>hey!</i>
B-Andrea	00:00:21.647<<h>ha:llo >	
<i>B-Andrea_Translation</i>		<i>hey!</i>
B-Andrea	00:00:22.106	[bleib]
<i>B-Andrea_Translation</i>		<i>calm ...</i>
A-Andreas	00:00:22.106	[<<ff>ne:in ↑ >]
<i>A-Andreas_Translation</i>		<i>no!</i>
A-Andreas	00:00:22.996<<ff>ne:in ↑ >	
<i>A-Andreas_Translation</i>		<i>no!</i>
A-Andreas	00:00:23.386<<f>jetzt red ICH ↓ >	

The overlap in the offset phase (00:30:530-00:31:220) indicates a retraction moment, in which speaker A gives up any effort to find a compromise with speaker B:

B-Andrea 00:00:29.176<<all>sag mal was habt ihr von verantwortung>		
<i>B-Andrea_Translation</i>		<i>what kind of irresponsibility do you</i>
A-Andreas	00:00:30.562<<f>sto:pp >	
<i>A-Andreas_Translation</i>		<i>enough!</i>

An analysis of the transition relevant places (TRPs) can deliver further cues to the development of the conflict process. The first transition relevant place is the moment when the man reacts to the woman's attack. It is first characterized by a wish to negotiate and to convince the other that she is wrong:

A-Andreas	00:00:12.269	so (.) jetzt muss ich dich unterbrechen=
<i>A-Andreas_Translation</i>		<i>I'm afraid I have to interrupt you now</i>
A-Andreas	00:00:13.766=das KINDERzimmer (-)	
<i>A-Andreas_Translation</i>		<i>about the child's room</i>
A-Andreas	00:00:14.763	hast du mal auf die gröÙe gesehen?
<i>A-Andreas_Translation</i>		<i>haven't you noticed how big it is?</i>

Another TRP is seen when the man needs to fight for the right to speak due to the woman interrupting him:

B-Andrea	00:00:18.554	du (.) [aber könnt mir auch] den
<i>B-Andrea_Translation</i>		<i>but let me</i>
A-Andreas	00:00:18.997	[und]
<i>A-Andreas_Translation</i>		<i>and</i>
A-Andreas	00:00:19.317	<<ff>ha:lt ↑ sto:pp ↓ >
<i>A-Andreas_Translation</i>		<i>keep quiet or ...</i>

As the third relevant TRP one can consider the moment of the woman's change of topic (00:00:25.738). The final TRP is when the man refuses to listen to the woman and leaves the room (00:00:30.526).

In these segments not only verbal and paraverbal elements, but also gestures and facial expressions, are turn-taking-constituting elements, indicating a “multimodal overlap” in the sense of “simultan realisierter kinesischer Beiträge” (“simultaneous realization of kinesic turns,” see Schmitt 2005: 45f.)

14.3.2 *Vocal Display*

The voice has an important function in the dynamics of the conflict. In the onset phase of the conflict up to the stroke (to 19.851) the interlocutors fight for their interactional power first of all at the level of the turn-taking dynamics. We notice some cases of latching, that is to say the speakers do not open a TRP (transition relevant place). To affirm their interactional rights, the interactants have a progressively more excited tone of voice and a higher pitch.¹³ Their emotional involvement is also shown through contractions, anacoluthons, and the stretched (marked) pronunciation of key words. Higher intensity (loudness) and intonation are used to mark the thematic development and information structure (for example for asserting or changing the topic: the child’s room, food, etc.).

Pauses (here annotated as short (.) and middle (–) estimated pauses) also play a crucial role. They are not just interruptions in the verbal flow (“turn vacancies,” Schmitt 2005: 24), but are often a signal of a switch in modality (i.e., what is not said is expressed through facial movements and gestures).

Annotation	Occ.	Av. Dur.	Time ratio	Latency
Intonation suddenly falls down in the second part of the question	1	1.7	0.04882812499999999	0.29
An introduction of a sentence and a sudden pause	1	1.7	0.048828 12499999999	1.99
Falling intonation	1	1.42	0.04078584558823529	3.96
The woman’s voice expresses hopelessness by falling, intonation and interrupted sentence being applied	1	2.0	0.05744485294117647	5.68
Emotional, agitated tone of voice	1	4.56	0.13097426470588233	7.68

¹³The poor quality of the recording did not allow an exact determination of the pitch contour. However, it was possible to notice perceptively an elevated pitch in the stroke phase of the conflict.

Falling intonation	1	1.51	0.04337086397058823	12.24
Rising intonation at the end of the questions	1	3.79	0.10885799632352941	14.76
Raised voice, closed eyes	1	1.45	0.041647518382352935	18.99
High tone applied to make the words hearable	1	1.65	0.04739200367647058	20.44
High tone of voice applied to keep the right to talk	1	1.27	0.03647748161764706	22.11
A strong prominence of the word "I"	1	0.97	0.027860753676470586	23.38
A strong prominence on the word "is"	1	1.365	0.039206112132352935	24.37
Fast, agitated way of speaking	1	4.827	0.1386431525735294	25.735
Falling intonation like in threats	1	0.32	0.009191176470588236	30.562
A threat pronounced by using falling intonation	1	1.89	0.054285386029411756	32.0

14.3.3 Kinetic Display

Gestures

Annotation	Occ.	Av. Dur.	Time ratio	Latency
Repeated right-to-left movements with the head	1	1.7	0.04882812499999999	1.99
The woman's repeated movements with the head one short upward and one downward movement with the shoulders	1	1.41	0.0404986213235294	3.97
The hearer's shaking the head in disagreement	1	1.985	0.05701401654411765	5.695
The woman's repeated nodding with the head; the man's right-left movements with the head	1	4.58	0.1315487132352941	7.68
Repeated up-and-down nodding with the head	1	6.26	0.17980238970588233	12.29

A decisive up-to-down movement with the head	1	2.66	0.0764016544117647	18.995
The man's two movements with the head from right to left and turning the face away from the woman	1	1.287	0.03696576286764705	22.095
One decisive up-to-down movement with the head	1	0.96	0.0275735294117647	23.382
A thumb pointed in the direction of the child's room, a short up-to-down nod with the head	1	1.31	0.03762637867647059	24.42
The woman stretches out her hand towards the man to pay his attention to a different topic	1	4.8	0.1378676470588235	25.73
The man stretches his arms apart	1	0.31	0.008903952205882353	30.573
The man nervously stands up and leaves the room	1	1.89	0.054285386029411756	32.005

Facial expressions

Annotation	Occurr.	Aver.Dur.	Time ratio	Latency
The mimics of the woman expresses amazement	1	1.7	0.04882812499999999	1.99
The woman is looking downwards	1	1.41	0.0404986213235294	3.97
The hearer's tense serious expression	1	2.0	0.05744485294117647	5.685
Visible growing nervousness on the man's face	1	4.57	0.1312614 8897058823	7.685
The man's raised eyebrows and nodding with the head whenever giving arguments	1	4.79	0.13758042279411764	13.77
Raised eyebrows	2	1.97	0.11316636029411764	18.99
Closed eyes	1	0.97	0.027860753676470586	23.385
Sudden interruption of eye-contact	1	0.32	0.009191176470588236	30.565

In the presented material, the main kinds of gestures are deictic gestures as well as gestures with beat and proxemic functions. The first kind of gesture can initially be seen when the man points in the direction of the child's room and thus wants to draw the woman's attention to it, or when the woman wants to draw the man's attention to the topic of food. Thus deictic gestures can be seen as speech-accompanying gestures as they convey meaning which is the subject of the discourse. They underline what has been said (illustrators) and divide bits of information. Deictic gestures can be observed in the following sections of the recording:

A-Andreas	00:00:13.766	das KINDERzimmer (—)
<i>A-Andreas_Translation</i>		<i>about the child's room</i>
A-Andreas	00:00:14.763	hast du mal auf die große gesehen?
<i>A-Andreas_Translation</i>		<i>haven't you noticed how big it is?</i>
A-Andreas	00:00:16.360	hast die menge an spielzeug äh (.) gesehen alles?
<i>A-Andreas_Translation</i>		<i>or how many umm toys there are?</i>
B-Andrea	00:00:18.554	du (.) [aber könnt mir auch] den
<i>B-Andrea_Translation</i>		<i>but let me</i>
A-Andreas	00:00:24.426	das KINDERzimmer IST sauber
<i>A-Andreas_Translation</i>		<i>the child's room is clean!</i>

Beats are gestures typical of excited speech. They are produced with beat-like movements with the hand, or by nodding or shaking the head. This can be seen in the following parts of the film:

A-Andreas	00:00:19.317	<<ff>ha:lt ↑ sto:pp ↓ >
<i>A-Andreas_Translation</i>		<i>keep quiet or . . .</i>
A-Andreas	00:00:20.450	[<<ff>jetzt re:icht es >]
<i>A-Andreas_Translation</i>		<i>I've had enough!</i>
B-Andrea	00:00:20.450	[<<h>ha:llo >]
<i>B-Andrea_Translation</i>		<i>hey!</i>
B-Andrea	00:00:21.647	<<h>ha:llo >
<i>B-Andrea_Translation</i>		<i>hey!</i>
B-Andrea	00:00:22.106	[bleib]
<i>B-Andrea_Translation</i>		<i>calm . . .</i>
A-Andreas	00:00:22.106	[<<ff>ne:in ↑ >]
<i>A-Andreas_Translation</i>		<i>no!</i>
A-Andreas	00:00:22.996	<<ff>ne:in ↑ >
<i>A-Andreas_Translation</i>		<i>no!</i>
A-Andreas	00:00:23.386	<<f>jetzt red ICH ↓ >
<i>A-Andreas_Translation</i>		<i>I'm talking now!</i>

The third group of gestures are proxemic gestures. They serve to allocate communicative space, often to the disadvantage of one of the communicative partners. This can be clearly seen when the woman, on being attacked by the man, feels forced to back down, because the man's gestures have a significant amplitude. Thus the man's proxemic gestures have the function of attacking, whereas the woman uses proxemic gestures to defend herself. In the presented scene, backchannel signals are mainly nonverbal and are, for example, the man's shaking of his head as a sign of disagreement, and the woman's movement of the head as a sign of resignation and disapproval (see Kendon 2002).

14.4 Final Considerations

In this chapter we have tried to show how low-stakes conflict processes caused by verbal aggression can be the object of linguistic analysis. In this kind of conflict process, the primary element is the struggle for interactional power, which is pursued through the strategic use of voice, lexical choice, gestures, facial expression, and proxemics. Through a combined method of conversational and multimodal analysis it is possible to describe the conflict dynamics (escalation) and the cues of every conflict stage in a multidimensional way. In the presented conflict situation we see how accommodation processes influence the lexical choices made, loudness, and movements of the head and body.

In the interaction analyzed above, the claim for interactional power can be retraced through the participants' attempts to gain control of the informative structure and the topics. Several verbal strategies (introduction of topics, change of topics, rhetorical questions, insinuations, hidden and open attacks, the focus on affective states) are used to establish the dominance of one or other interlocutor. The offset phase of the conflict coincides in the analyzed scene with a retraction: one interlocutor waives a further exchange and leaves the place of interaction. In this renunciation of communication we can see an example of the so-called avoidance strategy, which is proof of giving up any attempt to mediate.

References

- Austin J (1962) *How to do things with words*. OUP, Oxford
- Bonacchi S (2013) *(Un)Höflichkeit. Eine kulturologische Analyse Deutsch-Italienisch-Polnisch*. Peter Lang, Frankfurt
- Bousfield D (2008) *Impoliteness interaction*. Benjamins, Amsterdam
- Brown P, Levinson SC (1987) *Politeness. Some universals in language usage*. CUP, Cambridge
- Culpeper J (1996) Towards an anatomy of impoliteness. *J Pragmatics* 25(3):349–367
- Culpeper J (2008) Reflections on impoliteness, relational work and power. In: Bousfield D, Locher MA (eds) *Impoliteness in language*. de Gruyter, Berlin, pp 17–44
- Culpeper J (2011) *Impoliteness. Using language to cause offence*. CUP, Cambridge

- Culpeper J, Derek B, Anne W (2003) Impoliteness revisited: with special reference to dynamic and prosodic aspects. *J Pragmatics* 35(10–11):1545–1579
- Durkheim É (1915) *The elementary forms of religious life*. Allen & Unwin Ltd, London
- Fiehler R (1986) Zur Konstitution und Prozessierung von Emotionen in der Interaktion. In: Kallmeyer W (ed) *Kommunikationstypologie. Handlungsmuster, Textsorten, Situationstypen*. Jahrbuch 1985 des Instituts für deutsche Sprache. Schwann, Düsseldorf, pp 280–325
- Fricke E (2007) *Origo, Geste und Raum*. de Gruyter, Berlin
- Galtung J (1972) Institutionalisierte Konfliktlösung. Ein theoretisches Paradigma. In: Bühl WL (ed) *Konflikt und Konfliktstrategie. Ansätze zu einer soziologischen Konflikttheorie*. Nymphenburger Verlagsanstalt, München, pp 113–176
- Giebels E, Ufkes EG, van Erp KJ (2014) Understanding high-stakes conflicts. In: Ashkanasy N, Ayoko OB, Jehn KA (eds) *The handbook of research in conflict management*. Edward Elgar Publishing, Northampton, MA, pp 66–78
- Goffman E (1959) *The presentation of self in everyday life*. University of Edinburgh SSRC, Edinburgh
- Goffman E (1967) *Interaction ritual: essays on face-to-face behaviour*. Penguin Books, London
- Grice HP (1975) Logic and conversation. In: Cole P, Morgan JL (eds) *Syntax and semantics 3: speech acts*. Academic Press, New York, pp 41–58
- Havryliv O (2009) *Verbale aggression: Formen und Funktionen am Beispiel des Wienerischen*. Peter Lang, Frankfurt
- Jacoby S, Ochs E (1995) Co-construction. An introduction. *Special Issue Res Lang Soc Interact* 28(3):171–183
- Karpiński M, Klessa K, Czoska K (2014) Local and global convergence in the temporal domain in Polish task-oriented dialogue. In: Campbel N, Gibbon D, Hirst D (eds) *Proceedings of speech prosody 7*. Dublin, pp 743–747 (<http://www.speechprosody2014.org>)
- Kendon A (2002) Some uses of the head shake. *Gesture* 2(2):147–182
- Lachenicht LG (1980) Aggravating language. A study of abusive and insulting language. *Papers Ling Int J Hum Commun* 13(4):607–687
- Locher MA, Bousfield D (2008) Introduction: impoliteness and power in language. In: Bousfield D, Locher MA (eds) *Impoliteness in language*. de Gruyter, Berlin, pp 1–13
- Mack RW, Snyder RC (1973) *The analysis of social conflict – towards an overview and synthesis*. In: Jandt FE (ed) *Conflict resolution through communication*. Harper & Row, New York, pp 25–28
- Mateo J, Yus F (2013) Towards a cross-cultural pragmatic taxonomy of insults. *J Lang Aggression Confl* 1(1):86–113
- Müller C (1998) *Redebegleitende Gesten. Kultur – Theorie – Sprachvergleich*. Arno Spitz, Berlin
- Ogden R (2006) Phonetics and social actions in agreements and disagreements. *J Pragmatics* 38:1752–1775
- Poggi I (2007) *Mind, hands, face and body: a goal and belief view of multimodal communication*. Weidler Buchverlag, Berlin
- Poggi I, D’Errico F (2013) Multimodal acid communication of a politician. *ESSEM@AI*IA 59–71*. <http://www.informatik.uni-trier.de/~ley/pers/hd/p/Poggi:Isabella>. Accessed 5 Dec 2014
- Sager SF (2005) Ein System zur Beschreibung der Gestik. In: Sager SF, Bühlig K (eds) *Nonverbale Kommunikation im Gespräch*. OBST, Oldenbourg, pp 19–47
- Schmitt R (2005) Zur multimodalen Struktur von turn-takin. *Gesprächsforschung—Online-Zeitschrift zur verbalen Interaktion* 6:77–109 (www.gespraechsforschung-ozs.de)
- Schmitt R, Deppermann A (2009) Damit Sie mich verstehen: Genese, Verfahren und recipient design einer narrativen performance. In: Buss M, Habscheid S et al (eds) *Theatralität des sprachlichen Handelns*. Fink, München, pp 79–109
- Schneider JG, Stöckl H (2011) *Medientheorien und Multimodalität*. Halem, Köln

- Selting M, Auer P, Barth-Weingarten D, Bergmann J, Bergmann P, Birkner K, Couper-Kuhlen E, Deppermann A, Gilles P, Günthner S, Hartung M, Kern F, Mertzluft C, Meyer C, Mi M, Oberzaucher F, Peters J, Quasthoff U, Schütte W, Stukenbrock A, Uhmann S (2009) Gesprächsanalytisches Transkriptionssystem 2 (GAT 2). *Gesprächsforschung—Online-Zeitschrift zur verbalen Interaktion* 10(2009):353–402 (www.gespraechsforschung-ozs.de)
- Shantz CU (1987) Conflicts between children. *Child Dev* 58(2):283–305
- Simmel G (1972) Der Streit. In: Bühl WL (ed) *Konflikt und Konfliktstrategie. Ansätze zu einer soziologischen Konflikttheorie*. Nymphenburger Verlagsanstalt, München, pp 65–112
- Simon FB (2010) Einführung in die Systemtheorie des Konfliktes. Carl-Auer, Heidelberg
- Spencer-Oatley H (2005) (Im)Politeness, face and perception of rapport: unpackaging their bases and interrelationships. *J Politeness Res* 1(1):95–119
- Wartenberg TE (1990) *The forms of power. From domination to transformation*. Temple University Press, Philadelphia

Chapter 15

Rhetoric of Truthfulness in the Battle Between Social Attributions and Empathic Emotions

F. Milena Marzano, R. Scardigno, and G. Mininni

15.1 Introduction

Interpersonal and intergroup conflicts are strictly connected to individuals incapacity to recognize themselves in others' belief systems and values. It is clear that conflict loses its potential positive effect in terms of knowledge growth when the comparison between interpretative systems is taken in a relational setting that is not supported by the tendency to highlight the reciprocal emotional connotation with regard to the processes of sense making. On the other hand, conflict resolution may happen if the different parts start to consider their similarities or differences, starting from the human experience of emotional sharing. But to establish an emotional resonance with my enemy/opponent, I must be able to attribute to him/her the capacity to tell the truth. The rhetoric of truthfulness is a fundamental socioepistemic condition that gives individuals and groups the opportunity to resolve tensions related to conflictual situations, especially because it maintains a typical dialogical situation and gives access to reciprocal trust. Trusting others is the meta-relational pattern that, starting from attachment relations, regulates interpersonal and intergroup relations, thereby promoting an emotional tuning specifically known as *empathy*.

F.M. Marzano (✉) • R. Scardigno • G. Mininni
University of Bari, Piazza Umberto I, 1, 70121 Bari, Italy
e-mail: marzanomilena@gmail.com

© Springer International Publishing Switzerland 2015
F. D'Errico et al. (eds.), *Conflict and Multimodal Communication*,
Computational Social Sciences, DOI 10.1007/978-3-319-14081-0_15

15.2 Theoretical Background

15.2.1 Definition, Nature, and Measures of Empathy

Understanding what motivates adults to improve social relations has captivated the interest of a number of scholars in different fields, including philosophers, psychologists, and anthropologists. In this framework the study of *empathy* has recently become relevant in the analysis of positive (interpersonal and intergroup) relations. Originally defined by the German *einfihlung*, this concept was subsequently systematically investigated by Theodor Lipps (1851–1914), a German philosopher and psychologist, who developed the theory of *einfihlung* for psychology “from a psychological, nonmetaphysical perspective” and through “a phenomenological method” (Wispe 1987, p. 39). Lipps believed that people responded to each other through *einfihlung*, which was preceded by *projection* and *imitation*, and that as the imitation of affect increases, *einfihlung* increases. According to Lipps, to transfer our psychic reality to an external object, for example a work of art, the object in itself needs to have a kind of internal predisposition toward being “emphatically” perceived. The art object triggers the emotional concern of the person who gaze at it, creating the empathic relationship when the observer is able to project on it certain contents of his/her personal psychological life. The English word *empathy* was actually coined by Titchener (1909) as a rendering of Lipps’ *einfihlung*, which he defined as a “process of humanizing objects, of reading or feeling ourselves into them” (Titchener 1924, p. 417).

It is certain, then, that empathy has been a topic of intense discussion since the years surrounding the birth of phenomenology. Edith Stein, for example, considered empathy a way to establish contact with others, not only openness towards the “other,” but also a way to experience the “other” inside the “self.”

One of Stein’s greatest merits is to have highlighted the social function of empathy: socialization among individuals depends on empathy, which is the process of comprehending others’ mental states. Stein considered empathy, in fact, as a crucial moment in the way from subjectivity to reciprocity, the experience by which it is possible to carry out intersubjective communication.

Although it has been a concept of interest for a long time, empathy did not become truly popular within psychology until the work of Carl Rogers and Heinz Kohut [for reviews, see Bohart and Greenberg (1997)]. Rogers, in fact, started a new tradition that has given birth to many views and understandings on the concept of empathy. In particular, he considered it as a fundamental element within the therapeutic relationship, useful to enter a person’s world without judging him or her (1975). Also, Kohut (1971, 1977, 1984) focused all his theoretical work and therapeutic practice on the concept of empathy, considering it the instrument by means of which the therapist is able to enter the patient’s psychological reality (Kohut 1959). More precisely, Kohut defined empathy at two different levels,

one more abstract and the other more clinically applied. At the most abstract level, empathy was defined as “vicarious introspection” (1959). In a more applied definition, Kohut (1984) stated that *empathy* “is the capacity to think and feel oneself into the inner life of another person” (p. 82).¹

In sum, in psychology as in aesthetics, empathy has been seen as a way of knowing and understanding another person or an object. However, unlike in aesthetics, empathy as a psychological concept has been approached by researchers with an empirical perspective.

Although since the contributions of Rogers and Kohut the concept of empathy has been appreciated particularly in psychotherapy, in fact (cf. Barrett-Lennard 1981; Bohart and Greenberg 1997; Duan and Hill 1996; Gladstein 1977), it has evolved in many ways during the last century, and a substantial body of research has been generated in recent decades on how empathy relates to child cognitive development (Hoffman 1977), altruism (Batson 1991), attribution (Regan and Totten 1975), social judgment (Krulewitz 1982), intergroup relations (Finlay and Stephan 2000; Stephan and Finlay 1999), and others. Such a broad interest supports the claim that empathy forms the very basis of all human interaction and is “an essential constituent” (Kohut 1959, p. 462) of all psychological phenomena. However, such a diversity of attention may have also contributed to the fact that the study of empathy has been characterized by a large amount of theoretical positions and inconsistent, even confusing, results about its *nature*, a dimension along which empathy has been defined (Bonino et al. 1998; Duan and Hill 1996).

Centuries ago and a century apart, Smith (1759) and Spencer (1870) identified two broad classes of responses: *a cognitive, intellectual reaction* on the one hand (an ability to understand another person’s perspective) and a more *instinctive, emotional reaction* on the other. In various forms, this fundamental distinction has been maintained: psychological research on empathy, in fact, has typically been based on one or the other of these general definitions, and in recent decades, the psychological nature of empathy has become a central topic, including for social psychologists.

In particular, already in the first half of the last century psychoanalysts and social psychologists had highlighted the role of affective dimensions of empathy in interpersonal relations, considering such a construct a process of emotional activation, more or less voluntary, in some cases innate, involved in the sharing of another’s experiences. Theories on the nature of empathy were largely influenced by the affective views of Lipps and Titchener until Kohler (1929), who was one of the first researchers to take on the topic in a more cognitive vein. Rather than continuing to focus on “feeling into” the experiences of another, in fact,

¹Kohut (1984) affirms that empathy is useful not only within the therapeutic context but also in everyday life, especially within family relationships. The author gives the example of a mother who must feel what her child is feeling to understand and reassure him or her. If we persistently deny empathy to a child, his or her psychological development will be endangered.

Kohler held that empathy was more the *understanding* of another's feelings than a *sharing* of them. Following this interpretation, from the 1960s researchers focused their attention on the role of *cognition* in generating empathic reactions. Within this approach, many researchers identified empathy with the ability to understand adequately others' way of evaluating and experiencing situations. According to the cognitive perspective, the act of empathizing was defined as the *cognitive ability* to see things from another's perspective by improving our knowledge about him or her (Barrett-Lennard 1981; Kalliopuska 1986; Katz 1963; Rogers 1986; Woodall and Kogler-Hill 1982), referring to the intellectual understanding of another's experience, and the recognition of emotions experienced by the other (Borke 1971).

In recent decades, most researchers who have studied empathy are in agreement that it is a multifaceted concept consisting of both *cognitive and affective* components (Davis 1983; Eisenberg 1986; Finlay and Stephan 2000; Marzano and Serino 2005; Stephan and Finlay 1999; Stiff et al. 1988) and consider empathy a set of related constructs encompassing both cognitive and affective reactions (Davis 1980, 1983; Deutsch and Madle 1975). Instead of defining empathy "solely as affective responses or cognitive reactions, the multidimensional approach recognizes that affect and cognition are intertwined in empathy" (Morrell 2010, p. 55). The resulting model of empathy seeks to articulate a conception of empathy that speaks across the various disciplines in which it plays a role and embraces a range of components ascribed to empathy.

What emerges, then, is awareness of empathy as a process rather than an emotion in and of itself. There is, Morrell argues, "no 'empathy' that we feel; instead, empathy is a process through which others' emotional states or situations have an affect upon us" (2010: 62). Recognizing empathy as a process rather than a state treats it as a multidimensional process that recognizes the intertwining of cognition and emotion and pays attention to the antecedents, process, and outcomes of empathy through which the transformation of relationships can emerge (Morrell 2010, pp. 55–62; see also Cameron 2011, 2012). These characteristics are picked up and developed further by Lynne Cameron's empathy model in ways that contribute to a dynamic and relational understanding of trust, empathy, and dialog.

As we can imagine, the variety of empathy measures is also related to the variety of ways in which it has been defined. Those measures are influenced by the theoretical assumptions underlying the concept of empathy itself (Bonino et al. 1998).

For example, different tools have been created to measure empathy as a stable human ability or *personality trait* (Davis 1980; Mehrabian and Epstein 1972), a *state* (Dymond 1949; Eisenberg et al. 1987), or a *multicomponent* phenomenon (Cochrane 1974; Davis 1983; Elliott et al. 1982).

In the same vein, different components of empathy have been measured through different tools, such as self-reports, and physiological measures² (e.g., Eisenberg et al. 1987; Krebs 1975).

Daniel Batson, in particular, who described empathy as “feeling a vicarious emotion that is congruent with but not necessarily identical to the emotion of another”,³ analyzes participants’ reactions to an (emotionally involving stimulus situation by means of a 24-item scale, namely, the Emotional Response Questionnaire. These stimulus situations can be represented by written narratives or audiotape recordings, followed by ratings scales. This method has the advantage of being easy to administer and makes it possible to focus on a large range of “vicarious emotions” (Batson et al. 1997b).

Batson, who focused most of his 20-year research program on empathy, altruism, and prosocial behavior, also proposes that “a distinction should be made between different emotional reactions” to seeing another person in need. In different studies, in fact (Batson et al. 1997a, b), factor analyses of participants’ emotional responses revealed a consistent tendency for the different sets of emotions to separate orthogonal factors. According to Batson et al. 2012, these distinct emotional responses can lead to qualitatively different kinds of motivation to help: personal distress produces an egoistic desire to reduce one’s own distress, while empathy is related to an altruistic desire to reduce the distress of a person in need. The assertion that feeling empathic emotions for someone in need evokes altruistic motivation to relieve that need has been called by Batson the empathy-altruism hypothesis (Coke et al. 1978; Batson et al. 1997b). More recently, other components of empathy have been individuated, namely, parallel empathy, which refers to one’s ability to feel the same emotions as another person, and pure empathy, which is a generic feeling of concern for another’s plight (Batson 2012; Serino and Marzano 2007).

²Often, researchers interested in empathy refer to the use of physiological indexes, based on changes in the reactions of the autonomic nervous system, such as perspiration, vasoconstriction, heartbeat, and so on. Those procedures are usually used for the study of *emotions* in general and are useful because they are free from biases related to *social desirability* and self-presentation strategies.

Another interesting field of research that is gaining popularity is the one proposed by Gallese et al. (2004). The authors have discovered a class of neurons, called *mirror neurons*, situated in the F5 area, that are activated when monkeys perform a specific action and when they observe another individual (monkey or human) performing the same action. This system seems to exist also in human beings, and a corpus of study is dedicated to analyzing whether those neurons are somehow implicated in the capacity of comprehending or sharing another person’s emotional states.

Even though the study of mirror neurons has achieved (in a short time) very interesting results, those studies are still far from providing an exhaustive explanation of how empathy develops. In fact, this perspective does not explain how those neurons act in relation to so-called social emotions, such as, for example, shame or jealousy. Moreover, they do not explain how these involuntary and automatic levels interact with more complex processes, such as the cognitive attribution that occurs in role playing. In this sense, alternative approaches need to be developed, especially those considering the integrative use of those measures and self-reports or self-ratings.

³In another paper, Batson defined empathy as “an *other* oriented emotional response elicited by and congruent with the perceived welfare of someone else” (Batson et al. 2002, p. 486).

15.3 Empathy and Conflict Management

Although opinions are still diverse about the nature of empathy, a substantial body of research has been generated in recent decades on how empathy relates to things such as altruistic motivation, moral development, similarity (Batson et al. 1996; Hoffman 2000; Houston 1990; Hume 1957; Kohut 1984; Krebs 1975; Kubo and Muto 1984), and interpersonal relationships, and during the same period, empathic emotions have also been considered an essential component of motivation for engaging in prosocial behaviors (Einlof 2008; Stephan and Finlay 1999) and have also received much theoretical and empirical attention because of their contributions to reducing social aggression, together with their ability to promote conflict resolution, social inclusion, and solidarity (Genç and Kalafat 2008). Moreover, some researchers have recently found that empathy is also significantly related to conflict management (LeBlanc et al. 2012; Wied et al. 2007; Halpern and Weinstein 2004). This trend is related to the necessity of connecting empathy to political theory and international relations, including deliberative democracy and communicative action (Crawford 2010; Dews 1992; Morrell 2010), psychological approaches to foreign policy analysis (Jervis 1970, 1976; Jervis et al. 1985; White 1991), and political judgment (Solomon 1988; Arendt 1965, 2006).

15.3.1 *Study: Focusing on Social Attributions and Empathic Emotions*

Although the literature contains important contributions to our understanding of the relationship between empathy and social relations, and researchers have no doubts that the feeling of empathic emotions for another person may compel people to work toward resolving conflicts or enhancing better intergroup relations (Batson et al. 1997b; Davis et al. 1996; Head 2012), what remains a matter of some controversy, however, is the reason why such relations exist.

By definition, such a process as empathizing requires that “inferences” and “attributions” be activated in different forms, in relation to the *other*’s condition, to his/her mood in a certain situation, and so forth (Bem 1972; Deschamps 1977, 1982; Hewstone 1989; Jones and Davis 1965). Weiner, for example, showed that people’s affective responses to a suffering person are shaped by the attributions they make regarding the *cause* of that suffering. In this connection, research on attribution and empathy suggests that it is easier to feel empathy for someone who is not responsible for his or her own condition, someone who is a totally innocent victim (Weiner 1980, 1995). In the same vein, according to Ryan (1976), the perception of *victim’s responsibility* is also an important component of the stereotype of most stigmatized groups. By emphasizing victims’ responsibility, a detached, depersonalized, and negative representation of them can be more easily justified. Batson, instead, shows

that the perception of victims' responsibility yields different effects, depending on the different *groups* considered.

In sum, the idea that people's affective responses to a person are shaped by the *attributions* they make regarding the people's plight found great empirical support (Batson et al. 1997; Weiner 1980, 1995). However, the direction in which this relationship develops has yet to be clearly defined: in line with Cameron's theory (2012), empathy, for example, may or may not represent a response to a *story*, and empathic reactions may depend on the kind of inferences this story can elicit.

In this study, participants are required to express their evaluations on a series of different inferences related to a short text describing the difficult experience of a person who entered a new sociocultural context. Such a situation may activate some conflictual dynamics between the "dialogical self" (Hermans 2001) of the story's protagonist, who must integrate her memories of the past with the difficulties of her present. According to us, type and level of empathy that may be activated by a specific story may be predicted by the inferences and judgments related to the truthfulness of the story, the social desirability of an empathetic answer, and the induced emotional impact.

These three "ways to identify" have different valences and justifications. The more general but, at the same time, more strict connection is the one between empathy and truthfulness since this dimension is related to the relationship between "heads and texts" (Mininni 2010). A person's truthfulness depends on the sense making produced by the "texts" with which a person may be identified. Truthfulness is a "socioepistemic rhetoric" that is extremely relevant because it may be seen as a precursor of trust considered to be the starting point of all interpersonal and intergroup relationships. The construct of "socioepistemic rhetoric" (Berlin 1993) individuates a series of expressive routines that point out belief systems and values that can legitimate individuals' or groups' choices. Truthfulness is a sort of cultural meta-rhetoric in the sense that it defines the specific conditions that make texts coherent, valid, and reliable, for what concerns "what it says" and "how it says it."

15.4 Aims and Method

The present study provides answers to the questions regarding *whether* and *how* empathic emotions are affected by *social attribution processes*. In particular, the assumption that different types of emotions trigger different reactions for the first time, the roles of truthfulness, social desirability, and emotional impact on empathic reactions were simultaneously analyzed.

15.5 Method

Participants in the study included 66 female students of an introductory psychology course at the University of Bari receiving credit toward a course requirement. Participation was by individual appointment. On arrival, participants were escorted into a small research room and asked to carefully read and fill out a questionnaire. In a *between groups* design, participants dealt with the story of a person in need.

*The story.*⁴ The experimenter left the room while participants read a fictitious interview (all participants read the same exact story). In the text, the interviewee described her life since she moved to the new city, her sorrow due to the distance from her family, her own feelings and memories, and the difficulty she had finding a job in Italy and trying to integrate into Italian society, and so on.

Target's membership manipulation. The target's empathy membership (Italian vs. African) was manipulated. The stimulus person was presented as being either a *southern Italian* girl (Annamaria) or an *African* girl (Chandah) interviewed in an Italian city.

Using a randomized-block procedure, participants were equally distributed between the two different experimental conditions.

Self-reported emotional reaction to target's situation (empathy of the reader). Participants completed the Batson Empathy Scales, composed of 24 items (example of items: "After reading this story, how much do you feel . . . grieved, sorrowful, compassionate, etc."). For each item, participants rated the degree to which (1 = *not at all*, 7 = *extremely*) they experienced that emotion after reading the story.

Perceived social desirability of victim. The questionnaire included five items used to measure the social desirability of different emotional reactions to the story. Participants were asked to indicate "how appropriate" it was "to feel concerned about the girl's situation," "to help her," "to be touched by her story," "to show her sympathy," and "to remain detached." Social desirability items were assessed by a seven-point scale (1 = *not at all*, 7 = *extremely*)

Truthfulness of story. Once these answers were provided, participants also had to rate the credibility of the story they had read (1 = *not at all*, 7 = *extremely*).

Emotional impact of story. Finally, participants indicated how "alarming," "interesting," "worrying," "touching," "irritating," "depressing," "annoying," "involving," or "extreme" the story they had just read was (1 = *not at all*, 7 = *extremely*).

⁴Within this study we aimed at controlling the effects related to the interpretation of the stimulus story; this, in fact, became an object of different pretested versions. The final version aimed at presenting a situation of difficulty but not one that was overly dramatic, with no elements of extreme hardship, i.e. those that underlined the dramatic character of the event and triggered discomfort and uneasiness, were eliminated. Moreover, since we wanted to stress the difference between ingroup and outgroup, we "created" a target as far away as possible from the participants' real world, for example, an African girl.

Debriefing. After participants completed the experimental tasks, the experimenter returned and conducted a full debriefing. Then participants were thanked and dismissed.

15.6 Data Analyses and Results

15.6.1 Batson's Empathy Scales

15.6.1.1 Structure of Empathy

Following Batson's procedure, we created distinct indexes for the different components of empathy. To this end, a principal components factor analysis was performed on participants' ratings on Batson's scales. A *t*-test comparison between the two experimental groups (ingroup vs. outgroup) was performed on these two indexes. In particular, the two experimental groups differ in terms of *parallel empathy*, so that this component was more stressed by respondents in the outgroup (versus ingroup) condition ($M_{\text{outgroup}} = 3.54$; $M_{\text{ingroup}} = 2.76$, $t_{64} = -2.478$, $p < 0.05$) rather than the ingroup. Along the same line, *pure empathy* is also stressed toward outgroup members more than toward ingroup members ($M_{\text{outgroup}} = 4.1$, $M_{\text{ingroup}} = 3.2$, $t_{64} = -2.76$, $p < 0.005$) (Fig. 15.1).

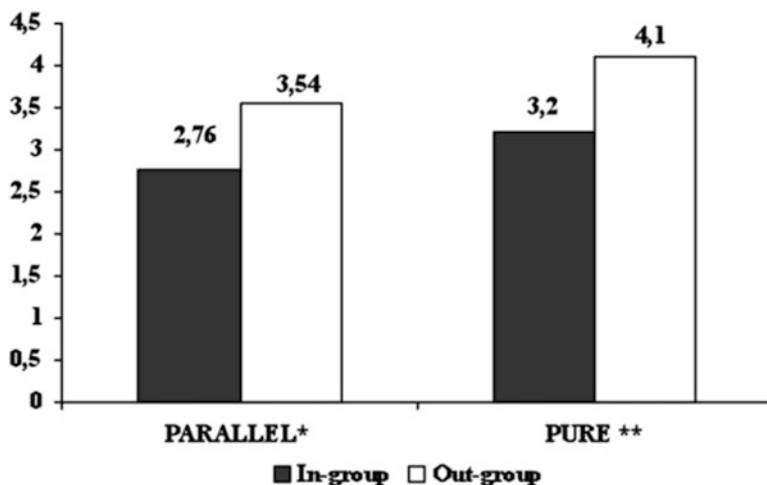


Fig. 15.1 Effects of target's membership (ingroup vs. outgroup): mean scores of parallel and pure empathy indexes (* $p < 0.05$; ** $p < 0.005$)

15.6.1.2 Self-Reported Emotions

Mean comparisons between the two experimental groups (ingroup and outgroup) were performed on each of the 24 items of Batson's scales. In particular, this analysis revealed that participants reported to be more *grieved* ($M_{\text{outgroup}} = 5.18$, $M_{\text{ingroup}} = 4.21$, $t_{64} = -2.32$, $p < 0.05$), *compassionate* ($M_{\text{outgroup}} = 5.33$, $M_{\text{ingroup}} = 4.15$, $t_{64} = -2.967$, $p < 0.005$), *upset* ($M_{\text{outgroup}} = 3.51$, $M_{\text{ingroup}} = 2.42$, $t_{64} = -2.57$, $p < 0.05$), *tender* ($M_{\text{outgroup}} = 3.84$, $M_{\text{ingroup}} = 2.57$, $t_{64} = -2.786$, $p = 0.007$), *feeling low* ($M_{\text{outgroup}} = 5.12$, $M_{\text{ingroup}} = 3.75$, $t_{64} = -3.004$, $p < 0.005$), *sorrowful* ($M_{\text{outgroup}} = 3.27$, $M_{\text{ingroup}} = 2.21$, $t_{64} = -2.425$, $p < 0.05$), *kind* ($M_{\text{outgroup}} = 3.24$, $M_{\text{ingroup}} = 2.28$, $t_{64} = -2.264$, $p < 0.05$), *sad* ($M_{\text{outgroup}} = 3.6$, $M_{\text{ingroup}} = 2.5$, $t_{64} = -2.397$, $p < 0.05$), *touched* ($M_{\text{outgroup}} = 5.21$, $M_{\text{ingroup}} = 4.21$, $t_{64} = -2.402$, $p < 0.05$), and *uneasy* ($M_{\text{outgroup}} = 2.66$, $M_{\text{ingroup}} = 1.65$, $t_{63} = -2.787$, $p < 0.005$) toward outgroup members than toward ingroup members (Fig. 15.2). As emerges from these analyses on Batson's scale of single items, the emotional response is constantly higher toward the outgroup target (even with respect to some emotions that could be described as *parallel emotions*).

How to explain this unexpected result? One possible answer is provided by participants' self-reported evaluations concerning the story and social desirability of different reactions to a person in need, such as those included in this study.

15.6.2 Truthfulness

Although in general our participants considered the stimulus story to be "fairly credible" ($M = 5.47$, $ds = 1.69$, range = 1–7), those in the outgroup condition considered the story as significantly more credible than participants in the ingroup condition ($M_{\text{outgroup}} = 6.03$, $M_{\text{ingroup}} = 4.93$, $t_{63} = -2.721$, $p < 0.01$) (Fig. 15.3).

15.6.3 Emotional Impact of Story

A *t*-test for independent samples was performed on the nine measures of emotional impact. Participants in the outgroup condition considered the story more alarming ($M_{\text{outgroup}} = 4.93$, $M_{\text{ingroup}} = 3.72$, $t_{64} = -2.949$, $p < 0.005$), interesting ($M_{\text{outgroup}} = 5.21$, $M_{\text{ingroup}} = 3.84$, $t_{64} = -3.442$, $p = 0.001$), worrying ($M_{\text{outgroup}} = 4.62$, $M_{\text{ingroup}} = 3.48$, $t_{63} = -2.519$, $p < 0.05$), touching ($M_{\text{outgroup}} = 6.24$, $M_{\text{ingroup}} = 4.39$, $t_{64} = -2.055$, $p < 0.05$), and involving ($M_{\text{outgroup}} = 5.18$, $M_{\text{ingroup}} = 4.18$, $t_{64} = -2.371$, $p < 0.05$, Fig. 15.4) than participants in the ingroup condition.

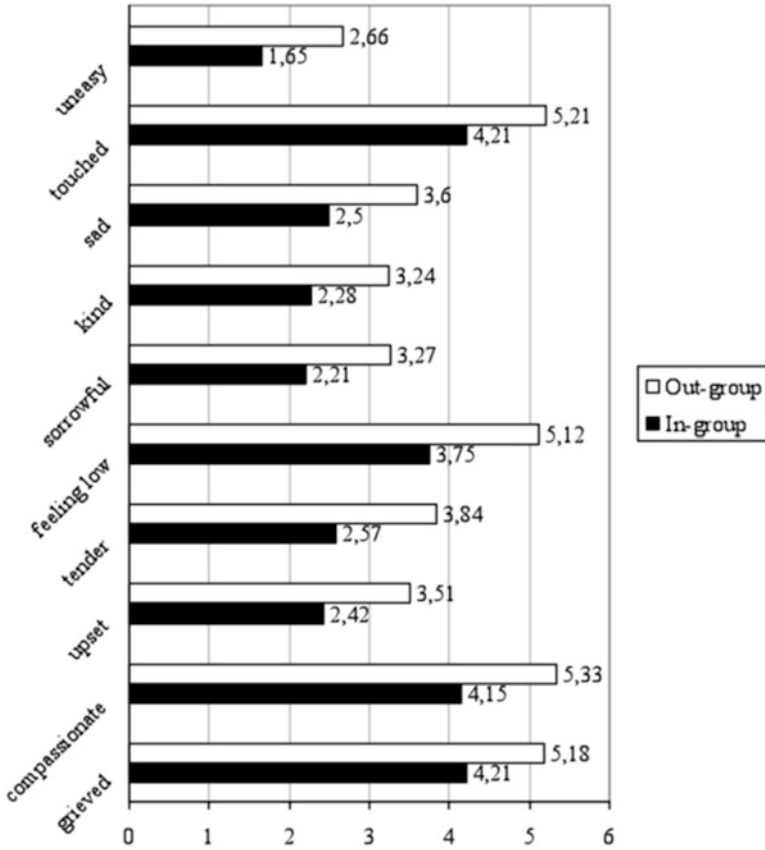


Fig. 15.2 Means of self-reported empathy (from 1 = not at all to 7 = extremely) as a function of the target’s membership

15.7 Social Desirability

The *t*-test on the measures of social desirability of the empathic reaction pointed out a significant effect of a target’s membership (Fig. 15.5): according to our participants, it is significantly more appropriate “to feel concerned with” the girl’s situation in the case of outgroup members than in the case of ingroup members ($M_{outgroup} = 4.69, M_{ingroup} = 3.66, t_{64} = -2.889, p < 0.005$).

On the other hand, it is more appropriate “to remain detached” from the Italian girl (ingroup member) than from the African girl (outgroup member) ($M_{outgroup} = 2.51, M_{ingroup} = 3.42, t_{64} = 2.363, p < 0.05$).

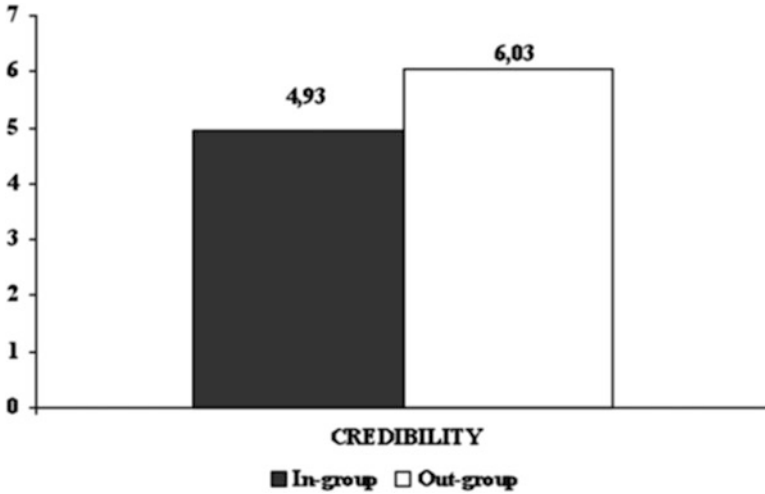


Fig. 15.3 Effects of target’s membership on truthfulness of stimulus story (from 1 = not at all to 7 = extremely): mean scores of each target’s membership condition (ingroup vs. outgroup)

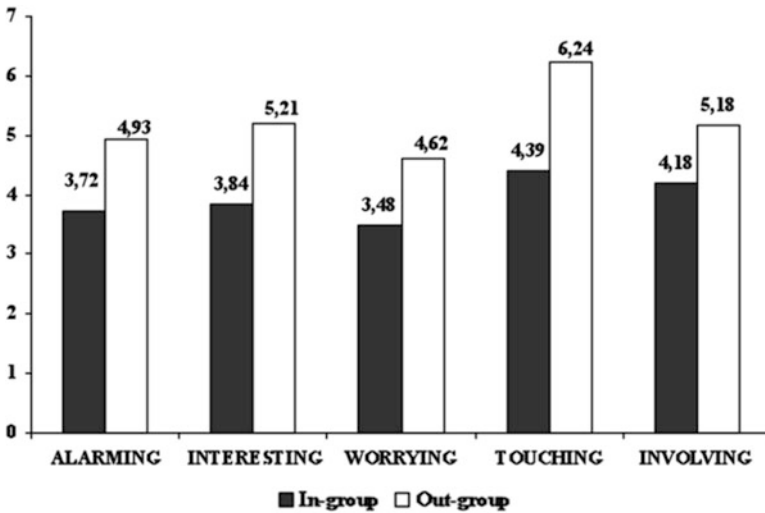


Fig. 15.4 Effects of target’s membership on perceived impact of stimulus story (from 1 = not at all to 7 = extremely): mean scores of each target’s membership condition (ingroup vs. outgroup)

15.7.1 Structural Analyses: Testing a Model

According to our findings, then, the *emotional impact* and perceived *truthfulness* of the story may play an important role in emphasizing the empathic reaction toward the outgroup victim. As in the two previous studies, these reactions seem to refer

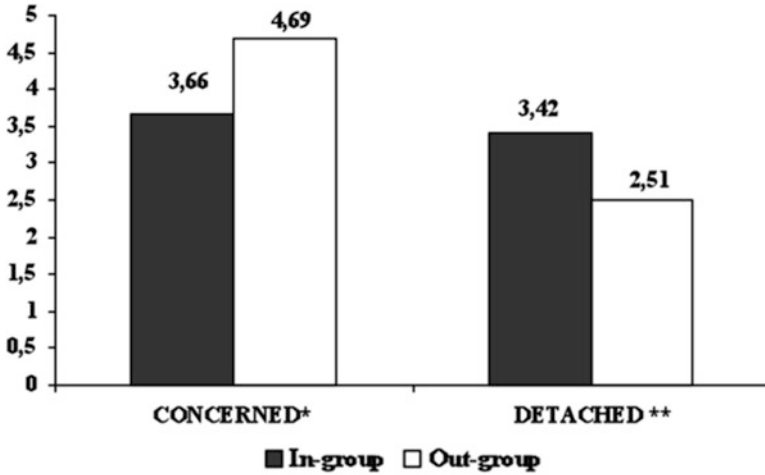


Fig. 15.5 Effects of target's membership on social desirability (from 1 = *not at all* to 7 = *extremely*): mean scores of each target's membership (ingroup vs. outgroup, * $p < 0.005$; ** $p < 0.05$)

to some emotions as being consistent with those felt by the victim herself (parallel empathy), but, unexpectedly, they even seem to refer to a general sympathy and compassion, in terms of what we have called pure empathy.

In this vein, to verify whether attribution processes (such as those involved in the perceived truthfulness and emotional impact of the story) and social desirability simultaneously affect both directly and indirectly empathic reactions toward a person in need, SEM was employed as a confirmatory approach useful for testing theoretical hypotheses concerning both direct and indirect structural relationships between the variables in analyses (Bollen 1989). In other words, the SEM approach compares a theoretically established model with empirical evidence to assess the degree to which data do not fit a theory, instead of simply deriving a model based upon empirical evidence.

In the present study, a model is tested where perceived *truthfulness* is hypothesized to serve as an exogenous variable on the *emotional impact* of the story, which in turn causes *social desirability*. To complete the causal chain leading to the prediction of empathy, social desirability is expected to be the direct predictor of the two indexes of pure empathy and parallel empathy. Finally, in the present model of prediction of empathy, listening to a story evaluated as credible and severe can elicit *comprehension*, *compassion*, and *sympathy* toward a person in need. As a consequence, the emotional impact ascribed to the story is expected to have a direct effect on the pure empathy dimension.

The SEM model was applied on the indexes of truthfulness (one item; Sect. 15.6.2), emotional impact (average value from nine items; see Sect. 15.6.2, Cronbach's alpha = 0.82), social desirability (five items, Cronbach's alpha = 0.75),

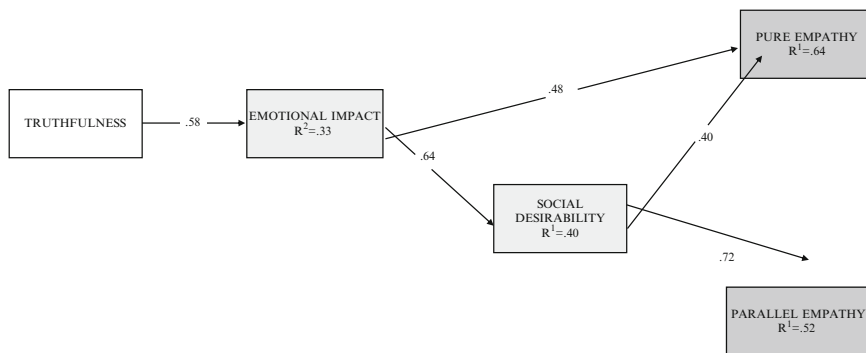


Fig. 15.6 Path diagram on effects of truthfulness, emotional impact of story, and social desirability on pure and parallel empathy in ingroup condition. Note: All paths were significant for $p < 0.05$. For the endogenous variables, the estimated R -squared is given in the squares

pure empathy (Cronbach's $\alpha = 0.87$), and parallel empathy (Cronbach's $\alpha = 0.92$). The model was first separately run on two groups of respondents corresponding to the two categories of the target's group membership (i.e., ingroup and outgroup) in order to assess the model fit for each condition. Subsequently, a multisample analysis in which the two models were directly compared to each other was performed on the data. The SEM approach (Bollen 1989) was applied using the software Lisrel (Joreskog and Sorbom 1996), and the hypothesized pattern of relationships among variables was estimated through a path analysis model. In such a model, explanatory and dependent variables are all observed, and constructs are considered to be directly measured, not assessed through a specific measurement model. As a consequence, no measurement errors are estimated by the procedure because all theoretical constructs are supposed to be fully captured by the corresponding observed indicators. Relationships among variables in the model might be directional or nondirectional. Directional relationships are estimated through regression coefficients. Nondirectional relationships correspond to covariances among variables. For each endogenous variable, the procedure gives an estimate of the proportion of variance unexplained by structural relationships (Bollen 1989; Joreskog and Sorbom 1996). Finally, all computed coefficients are associated with a significance test, which allows the researcher to draw inferences about the generalization of results to the population.

Figure 15.6 summarizes the *path analysis* model applied to the *ingroup condition*. The nonsignificant χ^2 value indicates that the proposed model does not differ significantly from the observed data. This means that the model has a good *fit* to the data in analysis [$\chi^2(5, N = 33) = 1.72, p = 0.89, RMSEA = 0.00, GFI = 0.97, AGFI = 0.93$]. As expected, truthfulness affects emotional impact, which then directly affects pure empathy and, through social desirability, the parallel empathy dimension. Overall, as displayed in the diagram, the R -squared for the endogenous variables in the model appears to be satisfactory.

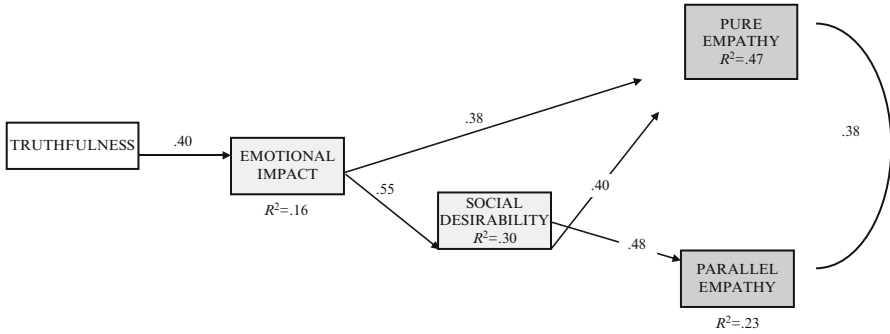


Fig. 15.7 Path diagram on effects of truthfulness, emotional impact of story, and social desirability on pure and parallel empathy in outgroup condition. Note: All paths were significant for $p < 0.05$. For the endogenous variables, the estimated R -squared is given in the squares

The model that results from the analysis on the *outgroup condition* also shows a good fit to the data [$\chi^2 (4, N = 33) = 5.13, p = 0.27, RMSEA = 0.09, GFI = 0.94, AGFI = 0.77, Fig. 14$]. Moreover, a new relationship was found to be significant, accounting for a stronger interconnection between pure and parallel empathy in the outgroup condition. Again, as displayed in the diagram, the R -squared for the endogenous variables in the model appears to be satisfactory, although less than in the previous case (Fig. 15.7).

Finally, to verify whether the two estimated models are consistent between each other, a multisample procedure was employed. The general fit indexes for the comparison appeared to be satisfactory: $\chi^2 (15, N = 66) = 9.97, p = 0.82, RSMEA = 0.00, GFI = 0.92$, percentage contribution χ^2 for the *ingroup* condition = 28.9 %, percentage contribution χ^2 for the *outgroup* condition = 71.1 %). The structural relationships among the variables in the analyses are identical in the two conditions of the victim’s categorization (as an ingroup member or as an outgroup member), except for a stronger interconnection emerging for the two dimensions of empathy in the outgroup condition. This means that the structural pattern of relationships among variables is equal across the two conditions of the present design, although with a significant difference in the outcome reaction of empathy. The present findings certainly deserve further verification; however, they illustrate the crucial role of attributions in empathizing.

15.8 Conclusions

Taken together with the results of the ANOVAs described in the previous sections, the present findings can also contribute to accounting for the effect of a victim’s categorization (as an ingroup or as an outgroup member) on the empathic response.

As emerged from the structural analyses, one possible explanation for these findings, then, deals with the *cognitive antecedents* of empathic emotions (in terms of the *meanings* attributed to the story and its social context) and with the *normative character* of the empathic process (in terms of its social desirability).

Perceived truthfulness, in fact, elicits more or less of the story's emotional impact. In turn, this may have different effects, according to which component of empathy is taken into account. In this sense, attribution processes underlying empathy are clearly evident.

The differences between the various components of empathy are highlighted by the impact of social desirability. This variable, in fact, has effects on both pure and parallel empathy, but while the former is also activated independently of an explicit reflection on the norms, the latter, by contrast, is elicited in a more difficult way. In this sense, pure empathy is more generic and less costly and, for this reason, more immediate and spontaneous, while parallel empathy (which requires a higher identification with the target) is activated only after a process of validation of the credibility of the victim's condition and an evaluation of the social norms to which the reaction must be uniformed.

These results are even more intriguing if we consider that the model's structure is confirmed independently of the target's group membership. Yet, in the outgroup condition a strict correlation between pure and parallel empathy is also emphasized, suggesting that this specific relationship must be more deeply investigated.

Attribution processes, in sum, are called into play in two different ways:

- The actor/observer difference is elicited by asking the participants to discriminate between reactive (attributed to an external observer) and parallel emotions (attributed to the target of the story);
- The effects in terms of self-reported empathy appear to be significantly moderated by an activity of inference that allows our participants to evaluate the *truthfulness* and the *emotional impact* of the story. As a matter of fact, depending on who the actor is (ingroup/outgroup member), the story takes on a different meaning and makes a more or less marked emotional impact on our participants.

In sum, the empathic response appears as the output of manifold and complex phenomena: it requires innate capacities and social competences; it is the result of a complex *interlacing* of cognitive and affective factors, but also the outcome of a battle between social constructs modulated by values and beliefs shared in a certain culture and collectivity. This study has shown that empathy may be activated also in conflict situations and that it may be a way toward a solution since subjectivities may make reference to a rhetoric of truthfulness. The dialogicity of empathy is supported by a minimal dialectic that emerges in the mind's capacity to recognize as plausible others' enunciative positions (Mininni 2013).

Empathy, in fact, may represent an important resource in conflict management, especially for its polyphonic nature, which may be able to reactivate the search for alternative senses that are made inaccessible by the conflict itself.

Being lived as a counterpoint of voices – my experience is mixed with the other's experience – empathy may activate cognitive energies that might contrast

with those tendencies toward indifference and reciprocal skepticism that triggered the relational framing of the conflict. In this sense, the root activated by empathy may support those attempts at conflict management in which persons may still be involved following the principle of responsibility (Zamperini 2007), that is, when the emotional basis of the contraposition has not been transformed into dehumanization.

Being supported by an “embodied simulation” (Gallese 2005), empathy may introduce the dynamics of creative thought into conflict management, although it cannot be considered a magic wand since its cognitive and emotional impact is regulated by uncertainty. There are, in fact, no “exact” conditions that can trigger a positive influence of empathy in conflict management. Empathy is a state of understanding nuances that persons in conflict may use only if they recognize the importance of “approximation” (Pagliarano 2013). When permeable to the magnetic field of empathy, conflict is sometimes able to invoke resources useful for the rejection of extreme positions and may bring about a recognition that differences are not as great as originally perceived.

Although it could be considered an ephemeral and evanescent reality at a socio-epistemic level, empathy aims at establishing the metarelation of trust (Hunt 2010; Mininni 2013) also in those who, being in a situation of conflict, find difficult to consider it reasonable.

References

- About FE, Levy SR (eds) (1999) Reducing racial prejudice, discrimination, and stereotyping: translating research into programs. *J Soc Issues* 55(4)
- Aderman D, Brehm SS, Katz L (1974) Empathic observation of an innocent victim: the just world revisited. *J Pers Soc Psychol* 29:342–347
- Albiero P, Ingoglia S, Lo Coco A (2006) Contributo all’adattamento italiano dell’Interpersonal Reactivity Index. *Testing Psicometria Metodologia* 13(2):107–125
- Allport GW (1954) *The nature of prejudice*. Addison-Wesley, Reading, MA
- Allport GW (1961) *Pattern and growth in personality*. Holt, Rinehart & Winston, New York
- Aquino K, Gover SL, Goldman B, Folger R (2003) When push doesn’t come to shove: interpersonal forgiveness in workplace relationships. *J Manag Inq* 12(3):209–216
- Armezzani M (2008) Gli psicologi e l’empatia. Atteggiamenti e stili di ricerca di fronte alla sfida della fenomenologia e delle neuroscienze. In: Armezzani M, Bosio GF, Cerri C, Lenoci M (eds) *Intenzionalità ed empatia. Fenomenologia, psicologia, neuroscienze*. OCD, Roma, pp 65–91
- Aron A, Aron EN, Tudor M, Nelson G (1991) Close relationships as including the other in the self. *J Pers Soc Psychol* 60:241–253
- Barnett MA (1987) Empathy and related responses in children. In: Eisenberg N, Strayer J (eds) *Empathy and its development*. Cambridge University Press, Cambridge, pp 146–162
- Barnett MA, Tetreault PA, Esper JA, Bristow AR (1986) Similarity and empathy: the experience of rape. *J Soc Psychol* 126:47–49
- Barnett MA, Tetreault PA, Masbad I (1987) Empathy with a rape victim: the role of similarity of experience. *Violence Vict* 2:255–262
- Barrett-Lennard GT (1981) The empathy cycle: refinement of a nuclear concept. *J Couns Psychol* 28(9):1–100

- Batson C, Sager K, Garst E, Kang M, Rubchinsky K, Dawson K (1997) Is empathy-induced helping due to self-other merging? *J Pers Soc Psychol* 73:495–509
- Batson CD (1991) The altruism question: toward a social-psychological answer. Lawrence Erlbaum Associates, Inc., Hillsdale, NJ
- Batson CD (1998) Altruism and prosocial behavior. In: Gilbert DT, Fiske ST, Lindzey G (eds) *The handbook of social psychology*, vol 2, 4th edn. McGraw-Hill, New York, pp 282–315
- Batson CD, Turk CL, Shaw LL, Klein TR (1995) Information function of empathic emotion: learning that we value the other's welfare. *J Pers Soc Psychol* 68:300–313
- Batson CD, Sympson SC, Hindman JL, Decruz P, Todd RM, Weeks JL et al (1996) "I've been there, too": effect on empathy of prior experience with a need. *Pers Soc Psychol Bull* 22:474–482
- Batson CD, Polycarpou MP, Harmon-Jones E, Imhoff HJ, Mitchener EC, Bednar LL, Klein TR, Highberger L (1997a) Empathy and attitudes: can feelings for a member of a stigmatized group improve feelings towards the group? *J Pers Soc Psychol* 72:105–118
- Batson CD, Sager K, Garst E, Kang M, Rubchinsky K, Dawson K (1997b) Is empathy-induced helping due to self-other merging? *J Pers Soc Psychol* 73:495–509
- Batson CD, Eidelman SH, Higley SL, Russell SA (2001) "And who is my neighbour?" II: quest religion as a source of universal compassion. *J Sci Study Relig* 40:39–50
- Batson CD, Chang J, Orr R, Rowland J (2002) Empathy, attitudes and action: can feeling for a member of a stigmatized group motivate one to help the group? *Pers Soc Psychol Bull* 28:1656–1666
- Batson CD (2012) The empathy-altruism hypothesis: issues and implications. In: Decety J (ed) *Empathy: from bench to bedside*. MIT Press, Cambridge, pp 41–54
- Bigler RS (1999) The use of multicultural curricula and materials to counter racism in children. *J Soc Issues* 4:687–705
- Bohart AC, Greenberg LS (1997) *Empathy reconsidered: new directions in psychotherapy*. United Book Press, Baltimore
- Bollen KA (1989) *Structural equations with latent variables*. Wiley, New York
- Bonino S, Lo Coco A, Tani F (1998) *Empatia. I processi di condivisione delle emozioni*. Giunti, Firenze
- Bower GH (1981) Mood and memory. *Am Psychol* 36:129–148
- Bower GH (1983) Affect and cognition. *Phil Trans Roy Soc Lond B Biol Sci* 302:387–402
- Brewer MB (1988) A dual process model of impression formation. In: Srull TK, Wyer RS Jr (eds) *Advances in social cognition*, vol 1. Erlbaum, Hillsdale, NJ, pp 1–36
- Brewer MB, Brown RJ (1998) Intergroup relations. In: Gilbert DT, Fiske ST (eds) *The handbook of social psychology*, vol 2, 4. McGraw-Hill, Boston, MA, pp 554–594
- Brewer MB, Gaertner SL (2001) Towards reduction of prejudice: intergroup contact and social categorization. In: Brown R, Gaertner S (eds) *Blackwell handbook of social psychology: intergroup processes*, vol 4. Blackwell Press, Oxford, pp 451–472
- Brewer MB, Gardner WL (1996) Who is this "we"? Levels of collective identity and self representations. *J Pers Soc Psychol* 71:83–93
- Brewer MB, Miller N (1984) Beyond the contact hypothesis: theoretical perspectives on desegregation. In: Miller N, Brewer MB (eds) *Groups in contact: the psychology of desegregation*. Academic, New York, pp 281–302
- Brown R (1995) *Prejudice: its social psychology*. Blackwell, Oxford, England
- Brown R, Vivian J, Hewstone M (1999) Changing attitudes through intergroup contact: the effects of group membership salience. *Eur J Soc Psychol* 29:741–764
- Bryant BK (1982) An index of empathy for children and adolescent. *Child Dev* 53:413–425
- Calegari P (1994) *Il muro del pregiudizio. Letture in tema di ecologia della mente*. Liguori, Napoli
- Cameron L (2012) Multi-level dynamic model of empathy. *Transforming conflict: trust, empathy, and dialogue*
- Campbell RJ, Kagan N, Krathwohl D (1971) The development and validation of a scale to measure affective sensitivity (empathy). *J Couns Psychol* 18:407–412
- Cantor N, Mischel W (1979) A prototypes in person perception. In: Berkowitz L (ed) *Advances in experimental social psychology*. Academic, New York

- Castelli L, Vanzetto K, Sherman SJ, Arcuri L (2001) The explicit and implicit perception of ingroup members who use stereotypes: blatant rejection but subtle conformity. *J Exp Soc Psychol* 37:419–426
- Castelli L, Arcuri L, Zogmaister C (2003) Perceiving ingroup members who use stereotypes: implicit conformity and similarity. *Eur J Soc Psychol* 33:163–175
- Cialdini RB, Brown SL, Lewis BP, Luce C, Neuberg SL (1997) Reinterpreting the empathy–altruism relationship: when one into one equals oneness. *J Pers Soc Psychol* 73:481–494
- Claire T, Fiske ST (1998) A systemic view of behavioral confirmation: counterpoint to the individualist view. In: Sedikides C, Schopler J, Insko CA (eds) *Intergroup cognition and intergroup behavior*. Erlbaum, Mahwah, NJ, pp 205–231
- Cochrane CT (1974) Development of a measure of empathic communication. *Psychother Theor Res Pract* 11:41–47
- Cohen J (1960) A coefficient of agreement for nominal scales. *Educ Psychol Meas* 20:37–46
- Cooper L, Sawaf M (1997) Emotional intelligence. *J Hum Behav* 31(2):13–17
- Crosby F, Bromley S, Saxe L (1980) Recent unobtrusive studies of Black and White discrimination and prejudice: a literature review. *Psychol Bull* 87:546–563
- Davis MH (1980) A multidimensional approach to individual differences in empathy. *Catalogue of Selected Documents in Psychology* 10:85. (Ms. No. 2124)
- Davis MH (1983) Measuring individual differences in empathy: evidence for a multidimensional approach. *J Pers Soc Psychol* 44:113–126
- Davis MH (1996) *Empathy: a social psychological approach*. Westview Press, Boulder
- Davis MH, Hull JG, Young RD, Warren GG (1987) Emotional reactions to dramatic film stimuli: the influence of cognitive and emotional empathy. *J Pers Soc Psychol* 52:126–133
- Davis MH, Conklin L, Smith A, Luce C (1996) Effects of perspective taking on the cognitive representation of persons: a merging of self and other. *J Pers Soc Psychol* 70:713–726
- DePaulo BM, Coleman LM (1986) Talking to children, foreigners, and retarded adults. *J Pers Soc Psychol* 51:945–959
- Deschamps JC (1977) Effect of crossing category memberships on quantitative judgement. *Eur J Soc Psychol* 7:122–126
- Deschamps JC (1982) Social identity and relations of power between groups. In: Tajfel H (ed) *Social identity and intergroup relations*. Cambridge University Press, Cambridge, pp 85–98
- Dijker AJ (1987) Emotional reactions to ethnic minorities. *Eur J Soc Psychol* 17:305–325
- Dillard JP, Hunter JE (1989) On the use and interpretation of the emotional empathy scale, the self-consciousness scales, and the self-monitoring scale. *Comm Res* 16:104–129
- Dovidio JF, Brigham JC, Johnson BT, Gaertner SL (1996) Stereotyping, prejudice and discrimination: another look. In: Macrae CN, Stangor C, Hewstone M (eds) *Stereotypes and stereotyping*. Guilford Press, New York, pp 276–319
- Dovidio JF, Kawakami K, Beach KR (2001) Implicit and explicit attitudes: examination of the relationship between measures of intergroup bias. In: Brown R, Gaertner S (eds) *Blackwell handbook of social psychology: intergroup processes*, vol 4. Blackwell Press, Oxford, pp 175–197
- Dovidio JF, ten Vergert M, Stewart TL, Gaertner SL, Johnson JD, Esses VM, Riek BM, Pearson AR (2004) Perspective and prejudice: antecedents and mediating mechanisms. *Pers Soc Psychol Bull* 30(12):1537–1549
- Duan C (1988) Effects of reasons for and consequences of a transgression, and perceived mood on moral judgment. Unpublished thesis, University of Maryland at College Park
- Duan C, Hill CE (1996) The current state of empathy research. *J Couns Psychol* 43:261–274
- Dymond RF (1949) A scale for the measurement of empathic ability. *J Consult Clin Psychol* 14:343–350
- Dymond RF (1950) Personality and empathy. *J Consult Clin Psychol* 14:343–350
- Eisenberg N (1986) *Altruistic emotion, cognition, and behavior*. Lawrence Erlbaum Associates, Hillsdale, NJ
- Eisenberg N, Strayer J (eds) (1987) *Empathy and its development*. Cambridge University Press, Cambridge

- Eisenberg N, Fabes R, Bustamante D, Mathy R (1987) Physiological indices of empathy. In: Eisenberg N, Strayer J (eds) *Empathy and its development*. Cambridge University Press, New York, pp 380–385
- Eisenberg N, Fabes RA, Murphy B, Karbon M, Smith M, Maszk P (1996) The relations of children's dispositional empathy-related responding to their emotionality, regulation, and social functioning. *Dev Psychol* 32:195–209
- Elliott R, Filipovich H, Harrigan L, Gaynor J, Reinmschuessel C, Zapadka JK (1982) Measuring response empathy: the development of a multicomponent rating scale. *J Couns Psychol* 29:379–387
- Feshbach ND (1975) Empathy in children: some theoretical and empirical considerations. *J Counsel Psychol* 5:25–30
- Feshbach ND (1991) Empathy and its correlates: cross cultural data from Italy. Eleventh biennial meeting of international society for the study of behavioural development, Minneapolis
- Feshbach ND (1997) Empathy: the formative years—implications for clinical practice. In: Bohart AC, Greenberg LS (eds) *Empathy reconsidered: new directions in psychotherapy*. United Book Press, Baltimore, pp 33–59
- Feshbach ND, Feshbach S (1969) The relationship between empathy and aggression in two age groups. *Dev Psychol* 1:102–107
- Fincham FD, Paleari FG, Regalia C (2002) Forgiveness in marriage: the role of relationship quality, attributions, and empathy. *Pers Relat* 9:27–37
- Finlay KA, Stephan WG (2000) Improving intergroup relations: the effects of empathy on racial attitudes. *J Appl Soc Psychol* 30:1720–1737
- Fiske ST (1998) Prejudice, stereotyping, and discrimination. In: Gilbert DT, Fiske ST, Lindzey G (eds) *The handbook of social psychology*, 4th edn. McGraw-Hill, New York, pp 357–411
- Fiske ST (2004) *Social beings: a core motives approach to social psychology*. Wiley, Hoboken, NJ
- Franco F, Maass A (1996) Implicit vs. explicit strategies of outgroup discrimination: the role of intentional control in biased language use and reward allocation. *J Lang Soc Psychol* 15:335–359
- Gaertner SL, Dovidio JF, Rust MC, Nier J, Banker B, Ward CM et al (1999) Reducing intergroup bias: elements of intergroup cooperation. *J Pers Soc Psychol* 76:388–402
- Gallese V (2005) Embodied simulation: from neurons to phenomenal experience. *Phenomenol Cogn Sci* 4:23–48
- Gallese V, Keysers C, Rizzolatti G (2004) A unifying view of the basis of social cognition. *Trends Cognit Sci* 8(9):396–403
- Gardner RC (1973) Ethnic stereotypes: the traditional approach, a new look. *Can Psychol* 14:133–148
- George JM (1996) Group affective tone. In: West M (ed) *Handbook of work group psychology*. Wiley, New York, pp 77–93
- George JM (2000) Emotions and leadership: the role of emotional intelligence. *Hum Relat* 53:1027–1055
- Giles H (ed) (1977) *Language, ethnicity, and intergroup relations*. Academic, London
- Gilligan C (1993) *In a different voice*. Harvard University Press, Cambridge, MA
- Gnisci A, Bakeman R (2000) *L'osservazione e l'analisi sequenziale dell'interazione*. LED, Milano
- Gordon M (2007) *Roots of empathy: changing the world, child by child*. T. Allen Publishers, Toronto
- Head N (2012) Transforming conflict trust, empathy, and dialogue. *Int J Peace Stud* 17(2):33–55
- Hewstone M (1989) *Causal attribution: from cognitive processes to collective beliefs*. Blackwell, Oxford
- Hewstone M, Brown R (1986a) *Contact and conflict in intergroup encounters*. Blackwell, Oxford
- Hewstone M, Brown R (1986b) Contact is not enough: an intergroup perspective on the 'contact hypothesis'. In: Hewstone M, Brown R (eds) *Contact and conflict in intergroup encounters Social psychology and society*. Basil Blackwell, Oxford, England, pp 1–44

- Hoffman ML (1977) Empathy, its development and prosocial implications. In: Kaesy CB (ed) Nebraska symposium on motivation, vol 25. University of Nebraska Press, Lincoln, NE, pp 169–217
- Hoffman ML (2000) Empathy and moral development: implications for caring and justice. Cambridge University Press, Cambridge
- Hornstein HA (1978) Promotive tension and prosocial behaviour: a Lewinian analysis. In: Wispè L (ed) Altruism, sympathy, and helping: psychological and sociological principles. Academic, New York, pp 177–207
- Houston DA (1990) Empathy and the self: cognitive and emotional influences on the evaluation of negative affect in others. *J Pers Soc Psychol* 59:859–868
- Hubert H (2001) The dialogical self: toward a theory of personal and cultural positioning. *Cult Psychol* 7:243–281
- Hume D (1957) An inquiry concerning the principle of morals. Liberal Arts Press, New York (Original work published 1751)
- Hunt L (2010) La forza dell'empatia. Una storia dei diritti dell'uomo. Laterza, Roma – Bari
- Izard CE (1971) The face of emotion. Appleton, New York
- Jones JM (1996) Prejudice and racism, 2nd edn. McGraw-Hill, New York
- Jones EE, Davis KE (1965) From acts to dispositions: the attribution process in social psychology. In: Berkowitz L (ed) Advances in experimental social psychology, vol 2. Academic, New York, pp 219–266
- Joreskog KG, Sorbom D (1996) LISREL 8: user's reference guide. Scientific Software International, Chicago
- Katz I, Hass RG (1988) Racial ambivalence and American value conflict: correlational and priming studies of dual cognitive structures. *J Pers Soc Psychol* 55:893–905
- Kellett JB, Humphrey RH, Sleeth RG (2002) Empathy and complex task performance: two routes to leadership. *Leader Q* 13(5):523–544
- Kohler W (1929) Gestalt psychology. Liveright, New York
- Kohut H (1959) Introspection, empathy, and psychoanalysis. *J Am Psychoanal Assoc* 7:459–483
- Kohut H (1984) How does analysis cure? University of Chicago Press, Chicago
- Krebs D (1975) Empathy and altruism. *J Pers Soc Psychol* 32:1134–1146
- Krulewitz J (1982) Reactions to rape victims: effects of rape circumstances, victim's emotional response, and sex of helper. *J Couns Psychol* 28:645–654
- Kubo Y, Muto T (1984) The effect of recalling one's own similar experiences on empathic understanding of the feeling of another person. *Jpn J Educ Psychol* 32:296–305
- Lampe M (2001) Mediation as an ethical adjunct of stakeholder theory. *J Bus Ethics* 31(2):165
- Lancaster S, Foddy M (1988) Self extentions: a conceptualization. *J Theor Soc Psychol* 18(1): 77–94
- Lipps T (1903) Einfühlung, inner Nachahmung, und Organumpfindungen (Empathy, inner imitations, and sensations). *Arch Gesamte Psychol* 2:185–204
- McGarty C, Yzerbyt VY, Spears R (2002) Stereotypes as explanations: the formation of meaningful beliefs about social groups. Cambridge University Press, Cambridge
- Mehrabian A, Epstein N (1972) A measure of emotional empathy. *J Pers* 40:525–543
- Menkel-Meadow C (2006) Peace and justice: notes on the evolution and purpose of legal processes. *Georgetown Law J* 94(2):663, 28 pgs
- Mininni G (2010) Teste e testi. Per una psicologia del significare. Aracne, Roma
- Mininni G (2013) Psicologia culturale discorsiva. FrancoAngeli, Milano
- Morrell MM (2010) Empathy and democracy feeling, thinking, and deliberation. Paperback
- Paffenholz T (2004) Designing transformative and intervention processes. Research Center for Constructive Conflict Management 1:1–16
- Pagliarano L (2013) Violenza e bellezza. Il conflitto negli individui e nella società. Guerini Associati, Milano
- Reagan D, Totten J (1975) Empathy and attribution: turning observers into actors. *J Pers Soc Psychol* 32:850–856
- Ryan W (1976) Blaming the Victim. Vintage, Newark, NJ

- Rifkin J (2009) *The empathic civilization: the race to global consciousness in a world in crisis*. J.P. Tarcher/Penguin, New York
- Rogers CR (1975) Empathic: an unappreciated way of being. *Counsel Psychol* 5:2–10
- Sheppard LA (1996) Cognition, affect, and behavior in the prediction of group attitudes. *Pers Soc Psychol Bull* 22:306–316
- Stein E (1989) *On the problem of empathy*. ICS Publications, Washington (Original work published 1917)
- Stephan WG, Finlay K (1999) The role of empathy in improving intergroup relations. *J Soc Issues* 4:729–743
- Stephan WG, Stephan C (2001) *Improving intergroup relations*. Sage, Thousand Oakes, CA
- Stiff JB, Dillard JP, Somera L, Kim H, Sleight C (1988) Empathy, communication, and prosocial behavior. *Comm Monogr* 55:198–213
- Struthers CW, Miller DL, Boudens CJ, Briggs G (2001) The effects of causal attributions on coworker interactions: a social motivation perspective. *Basic Appl Soc Psychol* 23:169–181
- Tajfel H (1982) *Social identity and intergroup relations*. Cambridge University Press, Cambridge, England
- Tajfel H, Turner JC (1979) An integrative theory of intergroup conflict. In: Worchel S, Austine WG (eds) *The social psychology of intergroup relations*. Brooks-Cole, Monterey
- Titchener E (1909) *Experimental psychology of the thought processes*. Macmillan, New York
- Titchener E (1924) *A text book of psychology*. Macmillan, New York
- Toi M, Batson CD (1982) More evidence that empathy is a source of altruistic motivation. *J Pers Soc Psychol* 43:281–292
- Triandis HC (1997) Cross-cultural perspectives on personality. In: Hogan R, Johnson J, Briggs SR (eds) *Handbook of personality psychology*. Academic, San Diego, pp 439–464
- Truax CB (1972) The meaning and reliability of accurate empathy ratings: a rejoinder. *Psychol Bull* 77:397–399
- Turner JC, Hogg MA, Oakes PJ, Reicher SD, Wetherell MS (1987) *Rediscovering the social group: a self-categorization theory*. Blackwell, Oxford/New York
- Vescio TK, Sechrist GB, Paolucci MP (2003) Perspective taking and prejudice reduction: the mediational role of empathy arousal and situational attributions. *Eur J Soc Psychol* 33:455–472
- Weiner B (1980) A cognitive (attribution)-emotion-action model of motivated behavior: an analysis of judgments of help-giving. *J Pers Soc Psychol* 39:186–200
- Weiner B (1995) *Judgment of responsibility: a foundation for a theory of social conduct*. The Guilford Press, New York
- Wenzel M, Mummendey A, Weber U, Waldrus S (2003) The ingroup as Pars pro Toto: projection from the ingroup onto the inclusive category as a precursor of social discrimination. *J Pers Soc Psychol* 29:4
- Wilder DA, Simon AF, Faith M (1996) Enhancing the impact of counterstereotypic information: dispositional attributions for deviance. *J Pers Soc Psychol* 71:276–287
- Wispè L (1987) History of the concept of empathy. In: Eisenberg N, Strayer J (eds) *Empathy and its development*. Cambridge University Press, New York, pp 17–37
- Yzerbyt VY, Dumont M, Gordijn E, Wigboldus D (2002) Intergroup emotions and self-categorization: the impact of perspective taking on reactions to victims of harmful behavior. In: Mackie D, Smith E (eds) *From prejudice to intergroup emotions*. Psychology Press, Philadelphia, pp 67–88
- Zahn-Waxler C, Cole PM, Welsh JD, Fox NA (1995) Psychophysiological correlates of empathy and prosocial behaviors in preschool children with behavior problems. *Dev Psychopathol* 7:27–48
- Zamperini A (2007) *L'indifferenza*. Einaudi, Torino

Chapter 16

Social Behaviour in Police Interviews: Relating Data to Theories

Merijn Bruijnes, Jeroen Linssen, Riëks op den Akker, Mariët Theune,
Sjoerd Wapperom, Chris Broekema, and Dirk Heylen

16.1 Introduction

Automatically generated social behaviour is important for human-like interaction with virtual agents (Vinciarelli et al. 2012). Models of human behaviour can be used to make the behaviour of artificial agents more believable to humans (Steunebrink et al. 2012). Believable agents are being employed in systems for training the skills required for successfully conducting negotiations (Swartout 2010) or job interviews (Vaassen and Wauters 2012). While the interaction with virtual agents may elicit some degree of learning by itself, the most important part of the experience lies in the reflection on what has happened (Koops and Hoevenaar 2013). Along the same lines, *explainable artificial intelligence* (Core et al. 2006) advocates the use of virtual agents that can explain their reasoning. These explanations of actions taken by the virtual agents could be used to improve the user's learning process, especially when the virtual agents clarify their actions in terms of the theories a user has to understand. A simple example of such a clarification might be an agent saying "I started shouting because your competitive stance made me angry."

In this paper, we focus on the interaction between police officers and suspects in police interviews. The long-term goal is to model a virtual suspect that can be used in an application for the training of police students. For this, we are developing a computational model that lets a virtual suspect select the behaviour that is most appropriate. The actions of the user are sensed and interpreted to form meaning, for example "the user is angry". Our computational model then uses these interpretations to form a "mental state" of the virtual suspect, for example

M. Bruijnes (✉) • J. Linssen • R. op den Akker • M. Theune • S. Wapperom
C. Broekema • D. Heylen
Human Media Interaction, University of Twente, PO Box 217,
7500 AE Enschede, The Netherlands
e-mail: m.bruijnes@utwente.nl

“the police officer is angry and that makes me sad”. The mental state (or mood) of the virtual suspect helps select the most human-like action that the virtual suspect has available, for example “I am sad so I will make a sad face”. We do not present a completely specified mental model for a virtual suspect in this paper, but provide the groundwork for such a mental model. This work is a continuation of work by Bruijnes (2013) and op den Akker et al. (2013) who looked into interpersonal attitudes in the same domain. This paper has two main contributions; first, it describes how we analysed interpersonal behaviour by validating ad hoc interpretations of factors resulting from a factor analysis. Second, we show which theories from (social) psychology and their underlying concepts are relevant to capture social interactions during police interviews and how these concepts are interrelated.

16.1.1 Police Interviews

Police interactions are a special type of social encounter, primarily because of the role of authority that the police officer has and the often uncooperative stance that a suspect takes—there may be a conflict between the interaction parties. The police officer receives training to resolve or reduce the conflict—to make an uncooperative suspect more cooperative. In this section, we discuss the training that police officers receive to become skilled at police interviews.

A police interview is often a situation of conflict. Suspects often do not cooperate with the police officer and the police interview in general, but behave in a confronting manner. Suspects may be withdrawn, defiant or even aggressive towards the police officer. The police officer has the difficult task to convince the suspect to cooperate and tell the truth in an interview: resolve the conflict. At the start of a police interview course, Dutch police students receive theoretical training on the use of the theory of interpersonal stance, or as they refer to it “Leary’s Rose” (Leary 1957) (see Sect. 16.3.1). In addition, “negotiation” strategies are taught with which it is possible to try and change the behavior of the suspect. The *Table of 10* by Giebels (2002) describes the strategies a police officer can use when, for example, they want to convince the suspect that cooperation will be of mutual benefit. After learning about the theory, some students get the opportunity to apply what they learned in a role-playing exercise with professional suspect-actors.

Training the proper behaviour for interviews is important for the effectiveness of the interview. For example, Holmberg and Christianson (2002) showed that when suspects perceive the police officer’s behaviour during the interview as dominant they tend to deny criminal accusations. Alternatively, when suspects perceive the interview as humane and respectful they gain the confidence and mental space required to admit criminal behaviour (Holmberg and Christianson 2002). Richardson et al. (2014) investigated the relation between the verbal mimicry (known as Language Style Matching) in police interviews and the confession to criminal behaviour. They showed that interviews that lead to a confession have a

higher rate of the suspect matching the verbal language style of the interviewer than interviews that did not lead to a confession. Further, they suggest that language style matching and mimicry can be employed strategically (Richardson et al. 2014; Rogan 2011). The notion that the behaviour of the interviewer is of influence on the outcome of an interview is critical for training and maintaining the skill to conduct a police interview.

16.1.2 Data-Driven vs. Theory-Driven

We work towards a computational model of virtual suspects and their conversational behaviour in police interviews. To build such a model we can follow different paths. One way is to start with a literature study and see what conceptual frameworks in behavioural and cognitive psychology and socio-linguistics—that focus on police interview practice—have to offer. The question is whether these theoretical frames provide insights that help us in building an operational model for suspect behaviour. Another approach is to start with the application of annotation schemes for specific dimensions of conversational behaviour to perform content analysis of the conversational data. The question then concerns the statistical correlations between aspects of behaviour in this type of data. For example, what is the relation between interpersonal stances that suspect and police take in an interview and the way they manage turn-taking (op den Akker et al. 2013)? The question is whether the concepts (labels) are clear enough and applicable to the data so that the inter-rater agreement is acceptable.

In the current paper we take a more holistic approach. The question is what concepts people use to describe what is going on in a police interview when they experience/observe it. How do they describe the interview and the behaviour of the interlocutors? Does the data, consisting of terms used to describe what is going on in a police interview, reveal interesting patterns? Such patterns might resemble the patterns that theories in behavioural and cognitive psychology and socio-linguistics describe, linking them to the observed practice of police interviewing.

16.1.3 Paper Outline

We look at human behaviour in a corpus of police interviews and try to establish which psycho-social theories might explain what happens in these interviews. In Sect. 16.2, we give a more detailed overview of our approach. A factor analysis of the occurrence of terms describing interactions in our corpus yielded the basis for a selection of theories and models from (social) psychology, which are discussed in Sect. 16.3. In Sect. 16.4, we address how the concepts underlying the theories matched the factors used for the factor analysis and how the concepts may be interrelated. To illustrate how these concepts can be used to understand the

behaviour of police officers and suspects, we describe several fragments from our corpus in terms of these concepts (Sect. 16.5). We conclude with our thoughts on the creation of a computational model for our virtual agents based on the combination of models and theories (Sect. 16.6).

16.2 Corpus Analysis

In this section, we outline how we analysed the behaviours of police officers and interviewees in a corpus of police interviews. We look at the behaviour of suspects and police officers because we feel modelling the interaction between both parties is necessary to create a believable virtual suspect. In Fig. 16.1, we show the steps we took in our analysis.

We started with a corpus of police interviews (step 1), the *Dutch Police Interview Training Corpus* (DPIT Corpus), see Sect. 16.2.1. From this corpus, six observers independently selected fragments that they thought were “interesting” in some way (2). For example, these were fragments in which a change in mood or atmosphere took place or fragments in which behaviour could be observed that

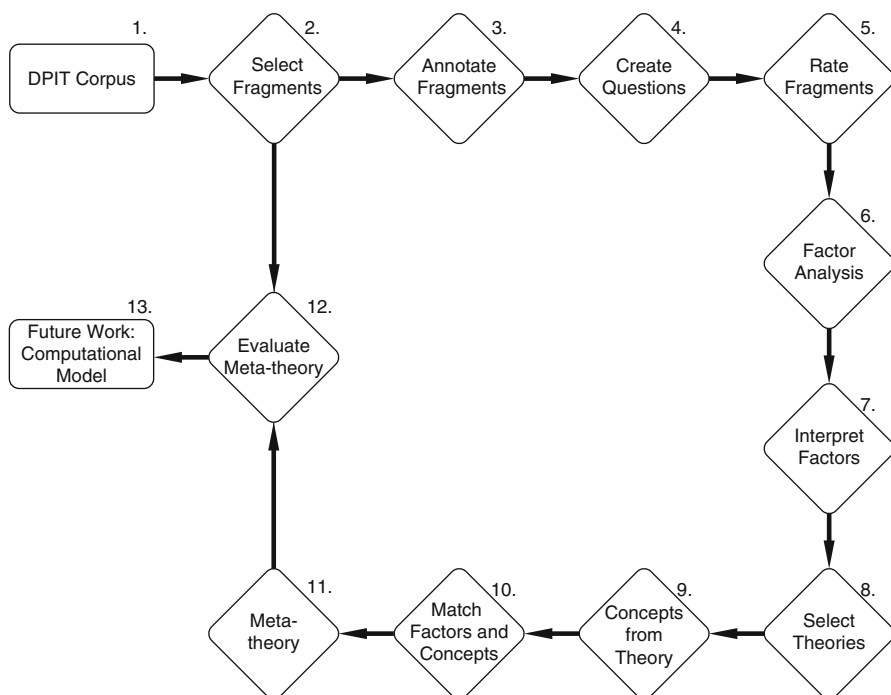


Fig. 16.1 Diagram showing the steps taken towards a computational model of police interviews

indicated how the police officers or suspects felt about the interaction. Next, a subgroup of the observers noted as many different *terms* as possible to describe what was going on in the fragments (3)—these could, for example, be adjectives describing the mood or nouns indicating behavioural traits. Based on these terms, we created questions with variations of the format “*To what extent is [term] the case?*” (4), see Sect. 16.2.2. The original six observers then rated a random subset of the same fragments from the corpus on a five-point Likert scale for every question, once for the police interviewer and once for the suspect. For example, a question from the point of view of the suspect was “To what extent is the suspect dominant?” (5). We performed a factor analysis to find a clustering of correlated questions (6) which we discuss in Sect. 16.2.3. Next, a subgroup of the original six observers reached consensus on the interpretation of the factors (7) and we selected well-known psychological and sociological theories that addressed these interpretations (8), see Sect. 16.3. In Sect. 16.4, we discuss how the concepts that these theories employ were matched to the factors by selecting the concepts that fit each factor (9 and 10). This also revealed the relations between the different concepts that were selected (11), as for some factors, concepts from different theories were applicable. This relation between theories allowed us to integrate the different theories into one “meta-theory” that provides the terms and concepts to describe the interactions in a police interview. We checked whether this “meta-theory” can describe what is happening in a police interview in Sect. 16.5 (12). Our final step (13) is to create a computational model from our “meta-theory” (Bruijnes 2013), but this remains a future endeavour (see Sect. 16.6).

16.2.1 Corpus Description

The *DPIT Corpus* is a corpus that consists of police interviews conducted by trainees of the Dutch National Police, recorded in 2012 and 2013. The police officers in the corpus are novice to moderately proficient police interviewers. The suspects they interview in this corpus are professionally trained actors. Due to privacy requirements, the video and audio data of the corpus is not publicly available.

The corpus consists of 32 interviews from 6 scenarios (cases) with a total length of approximately 13 h. The interviews vary in length from about 9 min to almost an hour. Some scenarios were enacted several times (with different students and sometimes also with different actors), while other scenarios were cut into separate interviews. In the latter, the suspect was interviewed multiple times, for example to give the police officers time to check facts. In these scenarios the same actor is interviewed by different police officers. The scenarios are specified as follows:

Bruintjes Ms Bruintjes is suspected of having bought a stolen smartphone from her cousin. She comes across as being not very bright but knew the phone was stolen.

Huls Mr Huls is suspected of the theft of a small amount of cash from a petrol station. He is a professionally trained actors for the police, has financial problems and has difficulty feeding his family.

Motor Actors from this scenario performed (with criminal exemption) an actual theft of an outboard motor and they played themselves. They try to appear innocent but are instructed to admit when the evidence gets too strong to reasonably deny the crime. The police students are not aware the suspects are guilty and treat them as any other suspect.

Remerink Ms Remerink is suspected to have stolen money from her (ex) husband's bank account. She is a full-time mother and gave up her career for their kids. Her (ex) husband is wealthy and he left Ms Remerink for another woman.

Van Bron Mr Van Bron is suspected of arson with the intent to kill his neighbour. Van Bron has an anti-social or bipolar disorder and has a criminal record. His girlfriend made a statement implicating Mr Van Bron.

Wassink Mrs Wassink is suspected to have physically attacked her neighbour over an argument about the dog of that neighbour. Wassink is a working-class mother whose world is as big as the neighbourhood she lives in and she is suspicious of people not originating from her neighbourhood (like her neighbour).

The actors are allowed to change the scenarios according to their preferences and to fill in the details as they see fit. This means that instances of interviews from the same scenario may be different, yet the police officers are always training with the same *persona*.

16.2.2 Observations of the Corpus

Six observers¹ with previous experience in interaction analysis independently viewed a selection of interviews from the corpus. They selected fragments that they thought were “interesting in some way” (step 2 in Fig. 16.1), for example, fragments in which a change in mood or atmosphere took place or fragments in which specific behaviour could be observed. The observers noted as many different *terms* as possible to describe what was going on in their fragments. The following is an example excerpt from the Remerink scenario (translated from Dutch):

POLICE: My name is Bill [*Surname*].² Can we address each other by our first name?

SUSPECT: Well I don't think so.

¹The first six authors of this chapter.

²Names are fake and/or anonymised for privacy.

P: Don't think so? Then I will use Ms Remerink. You can still call me Bill, if you have any questions you can do it like that... eh... eh... Ms Remerink.

One of the observers wrote down:

The suspect is *invited to call the police officer Bill*, even though she *insists* that he calls her Ms Remerink. He is trying to *be nice*, but he might give away *power*. Now there is *asymmetry* in the way they *address* each other (officer has to say “u” while she can say “je”³).

This description provides six descriptive terms for the fragment: *tutoyer*,⁴ *insist*, *be nice*, *power*, *asymmetry*, and *address*. A subgroup of the observers watched all the interesting fragments and also reported as many terms as possible to describe these fragments using a “think-aloud” strategy (step 3 in Fig. 16.1). In the example above, they added the terms *status* and *cold* to the terms selected by the original observer.

The subgroup of observers added fewer new terms to the entire collection of terms with every successively interview fragment. The first observed fragment yielded over 50 unique terms while only 3 new terms were added to the existing collection after observation of the annotations of the final interview. From this we conclude that we have obtained a sufficiently complete collection of terms necessary to describe the interviews included in our corpus: a *semantic frame* (Allan 2001). Eventually, the collection converged on a total of 251 unique terms.

16.2.3 Rating and Factoring Fragments

Based on the semantic frame of 251 terms, we created 227 questions with variations of the format “*To what extent is [term] the case?*” (step 4 from Fig. 16.1). We excluded terms that were not suited to create meaningful questions, for example, “fact” was a term that is too general to yield a sensible question or every question would have to be specific to the scenario. Example questions that were included are “To what extent is aggressive behaviour the case?”, “To what extent is the speaker indifferent?”, and “To what extent is there an uncomfortable posture?”

The original six observers rated fragments from the corpus on a five-point Likert scale for every question (step 5 from Fig. 16.1). The observers scored 14 fragments (with a total running time of 19 min) of the corpus on the 227 questions. The rated fragments were randomly selected from the fragments that were selected at

³In Dutch there is a difference between the second person pronouns “u” (formal) and “je” (familiar), both are translated to “you” in English.

⁴The (French) term for “to thee and thou”, to be familiar, based on the description “[...] invited to call the police officer Bill.”

Table 16.1 The ten items loading highest on the first suspect factor

Item	Factor loading
“Building pressure”	0.96
“Interruptions”	0.95
“Aggressive behaviour”	0.94
“Angry behaviour”	0.94
“Steering a conversation”	0.93
“Accusing the other”	0.92
“Attacking behaviour”	0.92
“Cutting the other off”	0.91
“Worked-up behaviour”	0.90
“Raised voice”	0.88

This factor was interpreted as *dominant and opposed*

step 2 (see Fig. 16.1). The fragments were scored by asking the rating questions explained above for both the police officer and the suspect, resulting in every question (corresponding to a term) being scored 28 times.

We performed a factor analysis to find a clustering of correlated questions which indicated which categories of questions—and, by extension, which terms—are related (step 6 from Fig. 16.1). Questions that were scored with no variation—that is, they always received the same score—for either the police officer or the suspect were excluded from analysis. This resulted in nine questions being removed (two were excluded from analysis for the subject, seven for the police officer). The excluded terms were found to be very role-specific; for example, crying is something the police never does. The factor analyses (extraction method: Principal Component Analysis, rotation method: Varimax with Kaiser Normalisation) revealed 13 factors for both the suspect and the police, see Tables 16.2 and 16.3.

Based on the related questions, we determined which terms loaded strongly (having a correlation of more than 0.50) per factor. The observers used these terms to interpret the corresponding factors (step 7 from Fig. 16.1). For example, the first factor (explaining 19.4% of the variance) for the suspect was interpreted as *dominant* and *opposed*. In Table 16.1, we show only the first 10 (of 54) items with factor loadings for the first factor of the suspect.

A subgroup of four of the original six observers interpreted all factors. The consensus on keywords describing the strongly loading factors of the suspects and the police officers is reported in Tables 16.2 and 16.3, respectively. In general, the observers’ interpretations were similar. For example, one of the observers interpreted the first factor of the suspect as “negative, confrontational and dominant”, while another observer interpreted it as describing “dominant behaviour and frustrations”. Discussing the interpretations among the observers generally resulted in agreement on the meaning of the factors. Some factors (suspect factors 9, 11, and 13 and Police factor 8) remain unclear as the observers were unable to reach consensus. We attribute this confusion to the few and diverse items that load on these factors.

Table 16.2 Variance explained by each factor for the suspect, with the interpretation of the factors

Suspect factor	Variance (%)	Cumulative variance (%)	Interpretation
1	19.41	19.41	Dominance/opposed (based on frustration), strategy/face (attack)
2	17.40	36.81	Rapport (building), together
3	15.59	52.39	Submissive/opposed, face
4	6.96	59.35	Together
5	6.65	66.00	Strategy (annoy)
6	6.19	72.19	Information exchange (questions)
7	5.74	77.93	Information exchange (lies)
8	5.27	83.20	Strategy (surround a fact)
9	4.70	87.91	–
10	4.01	91.92	Politeness (face)
11	3.02	94.93	–
12	2.90	97.83	Rapport (present)
13	2.17	100.00	– (one item: thank)

Table 16.3 Variance explained by each factor for the police, with the interpretation of the factors

Police factor	Variance (%)	Cumulative variance (%)	Interpretation
1	14.77	14.77	Rapport (missing rapport, negative emotions), arousal, opposed
2	13.57	28.34	Rapport (present), positivity, together
3	11.16	39.50	Strategy (avoid), information exchange (lies)
4	10.46	49.96	Submissive
5	8.83	58.78	Together
6	8.37	67.15	Arousal, dominance (competitive), strategy (attack)
7	8.01	75.16	Dominance/opposed (based on strategy)
8	4.61	79.77	–
9	4.56	84.33	Dominance/together
10	4.52	88.85	Strategy (confront)
11	4.16	93.01	Strategy (confront)
12	3.68	96.69	Dominance
13	3.31	100.00	Strategy (confront)

16.3 Linking Factors to Theories

In this section, we describe how the interpretation of the factors found in the previous section reflects ideas found in theories from (social) psychology (step 7 from Fig. 16.1). Based on the theories discussed in this section, in the following section we present a meta-theory that describes concepts relevant to the interactions in police interviews. The factors describing interpersonal attitudes are taken together in Sect. 16.3.1 on stance; the factors linked to face and politeness are discussed in Sect. 16.3.2; the factors linked to rapport are captured in Sect. 16.3.3. Additionally, two meta-concepts—*information* and *strategy*—were added to accommodate for the concepts that surfaced in the interpretation of the factor analysis but did not fit easily in a theory. Factors relating to information exchange are discussed in Sect. 16.3.4 and factors linked to strategy are discussed in Sect. 16.3.5. In each of the subsections, we describe the relation between these collections of factors and the theories from (social) psychology, including the concepts underlying those theories (step 8 from Fig. 16.1). We provide examples from the corpus and address work done with virtual agents and the mentioned theories. Also, we give some examples of systems using the concepts.

16.3.1 *Interpersonal Stance*

Several interpreted factors for both the suspect (1, 3 and 4 from Table 16.2) and the police (4, 5, 7 and possibly 6 from Table 16.3) are related to the attitude the suspect and the police officer have toward each other. Taken together, these factors sketch the outlines of *Leary's Rose*, a model for interpersonal behaviour (Leary 1957). Leary's Rose represents such behaviour in categories of interpersonal stance on the dimensions of affect (*x*-axis) and power (*y*-axis), see Fig. 16.2a. That is, the underlying concepts of Leary's Rose are part of these axes: the opposing concepts of dominance and submission constitute power, and the opposing concepts of feeling together (positive affect) or feeling opposed (negative affect) constitute affect.

Theories similar to Leary's Rose are known under names such as the Interpersonal Checklist (LaForge and Suczek 1955) and the interpersonal circumplex (Rouckhout and Schacht 2000), but the differences are often superficial. The model is often pictured as an ordering of the stances on a circle, situated on the two axes, which is called a circumplex. The circumplex can be divided into eight areas: these are interpersonal stances. The circumplex shows that stances that are close together are more related than those that are further apart on the circle, with opposites being negatively related (Fig. 16.2a). Leary suggests that human stances are affected by the interaction with the conversational partner. This means that two conversational partners influence each other with their stance during a dialogue. Leary calls these interactions “interpersonal reflexes” and asserts that acts on the dominance dimension are complementary while acts on the affect dimension are symmetric.

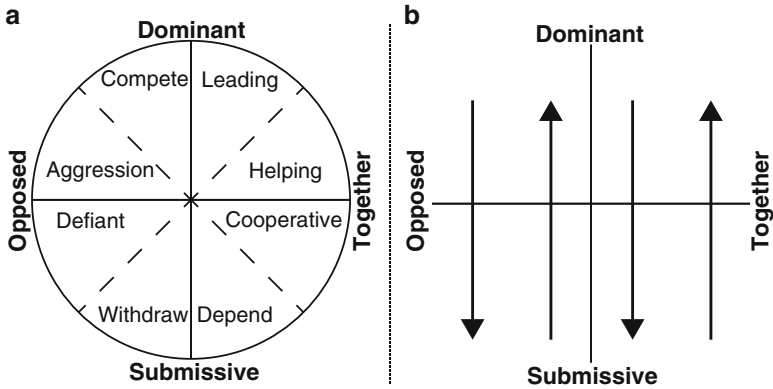


Fig. 16.2 (a) Leary's rose is defined by two axes: a dominance axis (*vertical*), describing the speaker's dominance or submissiveness towards the listener; and an affect axis (*horizontal*), describing the speaker's willingness to cooperate with the listener. The rose can be divided into eight areas that each corresponds to a stance. (b) The *solid arrows* indicate the behaviour-inviting relation between the quadrants according to Leary's (1957) theory. So, dominant-together invites submissive-together behaviour (and vice-versa) and dominant-opposed invites submissive-opposed behaviour (and vice-versa)

This means that a dominant act (for example, power display) elicits submissive acts, whereas an act with positive affect (for example, cooperative behaviour) elicits another positive affect act (see Fig. 16.2b). For example, if someone displays dependent behaviour towards another person (submissive and positive), that other person will feel a tendency to adopt a leading stance (dominant and positive) (Leary 1957).

16.3.1.1 Corpus Examples

In the corpus, we see several examples of different stances. In the Van Bron scenario, the suspect mostly behaves in a detached manner, unwilling to cooperate and expressing this through either competing or defiant behaviour. For example, when Van Bron becomes frustrated about not getting enough time to speak his thoughts and says to one of the officers that they should let him speak, the addressed officer says that he does not need to comply with Van Bron's wishes. As a result, Van Bron becomes somewhat aggressive and acts in a very dominant way, which corresponds to a hostile-dominant stance. On the other hand, the police officer usually displays behaviour with a together stance, for example in the Wassink scenario, in which the police officer does his very best to explain in other words to the interviewee what he was saying just a moment before. In this attempt to help the interviewee, the officer takes a very positive stance towards her by trying to help and cooperate with her.

16.3.1.2 Systems Using This Concept

There have been a few attempts to create virtual agents that act according to the interpersonal circumplex theory. One of these is the serious game *deLearyous*, which focuses on training interpersonal communication skills in a working environment setting, letting users interact with virtual agents through written natural language input (Vaassen and Wauters 2012). However, one of the findings of this project was that determining the stance of dialogue utterances is a very difficult task, even for human annotators. Other work has focused on finding correspondences of non-verbal behaviour with stances (Ravenet et al. 2013). This approach focusses on the generation of upper body movement and of facial animation on a virtual agent, based on human annotation of behaviours.

16.3.2 Face Threats and Politeness

Informed by Goffman's notion of *face* (Goffman 1959)—a person's public self-image—Brown and Levinson (hereafter, B&L) constructed their theory about politeness strategies (Brown and Levinson 1987). Suspect factors 1, 3 and 10 were interpreted as related to face (and politeness), see Table 16.2. The police factors were not interpreted as having a relation with face.⁵

B&L distinguish between negative and positive face, which denote a person's need for freedom (*autonomy*) and a person's need to be approved of and approving of others (*approval*), respectively. Their approach to politeness revolves around the concept of face-threatening acts (FTAs) which are inherent with actions taken by a speaker, as these actions potentially impose on a hearer's face by threatening their needs. B&L view politeness strategies as ways to redress these FTAs in order to minimise their imposition. The four main politeness strategy types follow below, ordered from least to most polite.

Bald on-record Being straight to the point, e.g., “Tell me where you were that night.”

Positive politeness Taking the other's wants into account, e.g., “Would you like to tell me where you were that night?”

Negative politeness Not hindering the other's autonomy, e.g., “If it's not inconvenient to you, could you tell me where you were that night?”

Off record Being indirect or vague about one's own wants, e.g., “I don't seem to have written down where you were that night.”

Conflict situations often arise in the police domain where people may not have the intention to stay polite—on the contrary, they may have the intention to be

⁵Police factor 4 was considered by some interpreters to have a relation with face but this was not unanimous.

impolite. Complementary to B&L's positive and negative politeness strategies, Culpeper et al. (2003) describe impoliteness strategies.

Positive impoliteness Damaging the addressee's positive face wants by excluding him or her, being disinterested, disassociating oneself from the addressee or using taboo words. For example, "Just bloody tell me where you were that night, so I can go home."

Negative impoliteness Damaging the addressee's negative face wants by being condescending, frightening him or her or invading his or her space. For example, "Tell me right now where you were that night, or I'll lock you up till Monday."

16.3.2.1 Corpus Example

On multiple occasions in the van Bron scenario, the suspect demands that the interview takes place according to his wishes. He does this mostly by using sentences that are short and direct, such as "You have to shut your mouth!" when he does not receive ample time to speak and expressing his disinterest by replying to the police with short answers ("It just is."). The first utterance is an example of an attack on the police officer's negative face, as the suspect invades his space and claims room for himself in the conversation. The second may not come across as a direct attack on the police officer's face, but it does impose on his positive face, as it indicates that the suspect does not want to cooperate and does not approve of the police officer. Impoliteness is not limited to solely being used by suspects: police officers use impolite utterances as well. This happens frequently when the police confront a suspect with a lie or an incriminating fact. For example, in the Huls scenario the officer is bald on-record and says "I think you took the money."

Even though police interviews can be uncooperative dialogues, politeness is still abundantly used. For example, in the Huls scenario a police officer explicitly expresses his approval of the suspect's behaviour: "I think it's decent of you that you try to support your family financially." This can be seen as an example of positive politeness, as the police officer takes the suspect's wants (of being approved) into account. In the Motor scenario, an example of negative politeness can be found, as a police officer tries not to impose too much on the suspect's freedom (his autonomy) by saying "I hope you don't mind too much to have this conversation with me."

16.3.2.2 Systems Using This Concept

Based on B&L's definition of politeness, several systems have incorporated virtual agents that can use utterances that vary in politeness. One of the first of these systems was designed by Walker et al. (1997) and involved asking a waiter for drinks with varying degrees of politeness, based on B&L's theory. Work by Gupta et al. (2007) continued this line of research by creating POLLY, a virtual agent that assisted users in learning English as a second language. This agent took into account

how imposing its requests were to the user by redressing these requests according to B&L's theory of politeness.

16.3.3 Rapport

The feeling of rapport can be described as being “in sync” with another person: communication takes place fluently and both interaction partners are roughly on the same level. In our corpus, we see the effects of both the presence and absence of this feeling. Suspect factors 2 and 12 in Table 16.2 and police factors 1 and 2 in Table 16.3 were interpreted as rapport (rapport-like descriptions).

Tickle-Degnen and Rosenthal (1990) conceptualised rapport in order to identify non-verbal correlates. Their description of the nature of rapport focuses on the interaction process as a whole and relies on three components of rapport: *mutual attention*, *positivity* and *coordination*. To develop and maintain rapport, interaction partners need to be mutually attentive so that they can achieve a focused and cohesive interaction. Moreover, their interest in the other party should remain at a high level during the course of an interaction. Figure 16.3 shows a schematic view of relative importance of mutual attention and the other two factors of rapport over time. Tickle-Degnen and Rosenthal mention that being positive towards each other is important during the build-up of rapport, yet becomes less important as time passes during interaction. An example of this is language usage among teens, where insults (a sign of low positivity) are the order of the day (Wang et al. 2012). Lastly, Tickle-Degnen and Rosenthal describe coordination as having a harmonious relationship between partners—this is the key term related to the feeling of being “in sync” and is the factor that becomes more important over time.

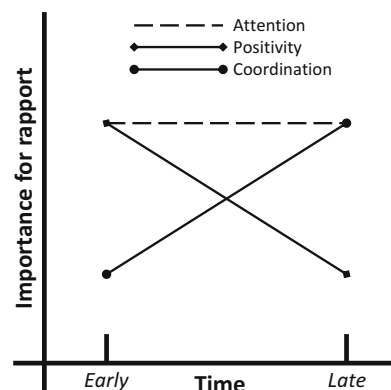


Fig. 16.3 Importance of the three components of rapport over time [from Tickle-Degnen and Rosenthal (1990)]

16.3.3.1 Corpus Examples

In a fragment of the Wassink scenario, the officer and the suspect start to speak more easily and freely to each other after a period of hesitant, slow interaction. The officer starts making gestures and the suspect has her full attention on his comments and responds quickly, without much hesitation. Soon after this, the suspect assumes a more interested body posture and finally both parties start laughing together. What is clear here is that the parties have mutual interest in each other and their coordination increases, resulting in them being “in sync”.

The opposite occurs in a fragment of the Van Bron scenario, in which Van Bron is not listening to what the officers are asking (or does not want to hear what they are saying) and starts making indecent comments about the female officer. In this case, there is little attentiveness of the suspect as well as a lack of intention to be positive towards the officer, resulting in an unpleasant atmosphere in which the police officer is not sure what to say any more.

16.3.3.2 Systems Using This Concept

Huang et al.’s work on the Rapport Agent 2.0—a virtual agent designed to build rapport with users—focuses on backchannelling and turn-taking at the correct moments (Huang et al. 2011). Here, backchannelling and turn-taking are used to inform the user of the attention of the virtual agent. Cassell et al. (2007) address long-term effects of rapport and how these could be modelled by looking at differences in interactions between friends and strangers. In their research, it became apparent that strangers tend to acknowledge each other more—that is, they make sure that the other party understood that they themselves understood what was being said. Friends are much more direct in their interaction, gazing at each other directly and being less explicit about their understanding of each other, which is explained by them having more rapport.

16.3.4 Information Exchange and Framing

Suspect factors 6 (questions) and 7 (lies) and police factor 3 (lies) were interpreted as having to do with information exchange. The discussion between the interpreters revealed more descriptions of information exchange than just “lies” and “questions”, but for these other categories no consensus was reached and they are not included in Tables 16.2 and 16.3. A factor analysis where the questions answered from the points of view of the suspect and the police were taken together revealed more information exchange descriptions during interpretation of these factors, including *give information*, *withhold information*, *lie*, and the notion of *topic or frame*.

Information is exchanged during all conversations between multiple interaction partners. Austin (1975) conceptualised an illocutionary act as the intended

meaning of an utterance, for example a request for information. Based on this notion, Searle (1969) created a classification of five different types of speech acts, namely representatives (informing), directives (requesting), commissives (promising), expressives (expressing a psychological state), and declarations (for official decisions). The main concepts in this theory are those of requesting information (*questioning*) and *giving* (or not giving) *information*.

A special case of giving or withholding information is *lying*: providing information that one knows or believes to be false. Police officers experienced in interviewing have above average lie detecting skills (Mann et al. 2004), mainly because they focus on cues that relate to a suspect's story. In other words, inconsistencies in the information exchange are important during police interviews.

The type of information that is exchanged and how it is interpreted is dependent on context—in other words, the interaction's *frame* determines the type of conversation. The notion of frame was first introduced by Bateson in 1955 as he studied the behaviour of monkeys in different situations (Bateson 1955). Bateson stated that no communication could be interpreted without a meta-message about what was actually going on—that is, what the current frame of the interaction between the monkeys was. During a play frame, all monkeys knew that certain behaviours were accepted (such as biting) which would otherwise be interpreted as a hostile act. Fillmore elaborated on this idea by stating that a frame is “a system of linguistic choices associated with a scene, where a scene is any kind of coherent segment of human actions” (Fillmore 1981). According to Tannen, conversational frames are repositories for social cultural norms of how to conduct different types of conversation, such as storytelling, teasing, and small talk (Tannen 1993). A frame tells us something about what we can and cannot say in that particular frame. The frame that is currently active allows us to decide which assumptions we can make, customs or “social scripts” we have (what we can do), and constraints we have (what we should not do).

16.3.4.1 Corpus Examples

Dutch police interviews start with a social frame during which the police officer tries to get to know the suspect and gathers information about the personal life and emotional situation of the suspect. After getting to know each other they continue with a task frame where they discuss the crime that the suspect has been accused of.

Conversational partners do not always agree on the frame that they are using. During the Wassink scenario the suspect does not agree with the social frame the police officer suggests and she asks: “Why do I have to tell you something about myself?”

In the Huls scenario the suspect eventually admits to the crime of stealing money from the gas station. During this confession the police officer uses an empathy frame (Bickmore 2008) in which he comforts the suspect by telling him that he understands his situation because he too has children. He agrees with the suspect that it is hard to provide for two children without a stable income.

16.3.4.2 Systems Using This Concept

Multiple virtual agent systems have been created that are at least partly based on speech or dialogue act theories. For example, the Mission Rehearsal Exercise and Stability and Support Operations systems and their derivatives feature agent decision making using speech acts (Swartout 2010). The same is true of Kopp's virtual museum guide which distinguishes between the performed behaviours and the communicative function of these behaviours (Kopp et al. 2005). This helps the virtual guide to select responses that vary in their performance, yet have the same communicative function.

Bickmore (2008) developed a health counselling agent that bases its reactions on both interpersonal stances and framing. Bickmore uses four different conversational frames to help the agent decide on how to react: the *task* frame, which is used for information exchange; the *social* frame, which is used for social chat and small talk interactions; the *empathy* frame, which is used for comforting interactions; and the *encourage* frame, which is used for coaching, motivating and cheering up interactions. With this information, combined with interpersonal stance, the agent can decide what behaviour to show in different situations.

16.3.5 Strategy Selection

Suspect factors 1, 5 and 8 and police factors 3, 6, 10, 11 and 13 were interpreted as having to do with strategies in interaction, see Tables 16.2 and 16.3. Specifically, the interpreters used the concepts *confront*, *surround*, *evade*, and *annoy*.

During communication, individuals make use of strategies to achieve their desired goals. These strategies play an important role, especially during non-cooperative communication such as in the police domain, as described in Sect. 16.1.1. Traum et al. (2008) describe a set of negotiation strategies—including finding the issue, attacking to aggressively attain a goal, and advocating or proposing solutions—and assert that the negotiating party must balance three goals to be successful in a difficult negotiation. The negotiator has to find an acceptable solution for the problem, gain and maintain the trust of the other participant(s) and manage the interaction by setting the agenda and controlling the topic.

Campos et al. (2012) explain strategies in terms of conflict. In natural conflict situations the agents will respond using emotional reactions. According to Campos et al., conflict varies around five dimensions: participants, causes, initiating action, responses and outcomes. Thomas (1992) argues that participants can take several approaches to resolve conflicts: *accommodation*, *avoidance*, *competition*, *collaboration* and *compromise*. As mentioned in the introduction, the *Table of 10* by Giebels (2002) describes the strategies a police officer can use when, for example, they want to convince the suspect that cooperation will be of mutual benefit.

16.3.5.1 Corpus Examples

The Remerink scenario in our corpus starts with a frustrated suspect who is apparently angry about something. The police officer uses a negotiation strategy to find out what is bothering her in order to resolve this issue. He asks the woman what is bothering her and eventually she says she is angry about the method with which she was picked up from her house. She is ashamed and angry about the way they came to her house and brought her in with all the neighbours watching.

Later in the Remerink interview, the suspect is accused of taking money from her ex-husband and she becomes emotional and silent every time when the topic of her husband comes around. Due to the fact that the topic is undesirable for her to talk about, she tries to avoid going into it any further.

In the Huls scenario, the police officer is surrounding a specific fact during the conversation, so the topic cannot be avoided. He continues to aggressively ask similar questions to the subject to put pressure on him to tell the truth.

16.3.5.2 Systems Using This Concept

The Mission Rehearsal Exercise and Stability and Support Operations systems and their derivatives feature virtual agents in war scenarios with which users have to negotiate (Swartout 2010). These scenarios deal with dilemmas the user has to solve. For example, a user has to convince a local Afghan doctor to move his clinic to another location, as the user has to conduct military operations in the area of the clinic. One of the ways to convince the doctor is using rational arguments such as offering incentives. Furthermore, this system also takes emotional consequences into account when deciding whether to cooperate with or oppose the user.

16.4 Relations Between Factors, Theories and Concepts

In the previous section, we discussed what theories from (social) psychology match the interpretations of the factors found in the factor analyses (see Sect. 16.2.3) and we explained the concepts from these theories (step 9 from Fig. 16.1). In this section, we discuss how these concepts are related to the factors (step 10 from Fig. 16.1). This gives insight into both the relations between the factors and the concepts, and the relations between the concepts themselves (step 11 from Fig. 16.1). Based on our findings on these relations, we describe how the theories from which these concepts originate are connected.

16.4.1 Concepts in Theories

Psychological and sociological theories use concepts and describe the relations between these concepts. Theories provide us with a way to describe an interaction (in our case, a police interview) and they can be used by a virtual tutoring agent (in our case, a virtual suspect) to predict the effects of its behaviour in an interaction with a human user. For example, the central concepts in the interpersonal stance theory are *dominant*, *submissive*, *together*, and *opposed* and the theory describes how the combination of these concepts creates the notion of “stance” and predicts how people are influenced by the stance of others. A virtual tutoring agent can use this knowledge to create an interesting and useful learning experience. For example, a user might learn by experiencing that if he displays opposed behaviour interaction, the conflict escalates. The virtual suspect can display opposed behaviour in an attempt to get the user to also display opposed behaviour and then let the conversation escalate (Bruijnes et al. 2013). Each of the theories we selected in the previous section has such concepts, see Table 16.4. The concepts from face are positive and negative *autonomy*, and positive and negative *approval*. The concepts for rapport are *coordination*, *attention*, and *positivity*. We added two meta-concepts—*information* and *strategy* to accommodate for the concepts that surfaced in the factor analysis interpretation but did not fit easily in a theory. The information concepts we found are *questioning*, *give information*, *withhold information*, *lie* and *frame* or *topic*. The strategy concepts are *confront*, *surround*, *evade* and *annoy*.

16.4.2 Factors: Theories and Concepts

The interpretation of the factors, see Sect. 16.2.3 and Tables 16.2 and 16.3, and the matching of these factors to theories leads to links between theories and factors. To validate these links, four observers indicated with which concept(s) from the theories (Table 16.4) a factor could be explained. This method provided us with a possibility to validate the intuitive (subjective) interpretation of the factors that is common practice in the field of social science. In other words, we used the initial interpretations to select theories that “cover” the factors and we used the concepts

Table 16.4 Concepts within the theories *stance*, *face*, and *rapport* and the meta-concepts *information* and *strategy*

Stance	Face	Rapport	Information	Strategy
Dominant	Autonomy+	Coordination	Questioning	Confront
Submissive	Approval+	Attention	Give info	Surround
Together	Autonomy–	Positivity	Lie	Evade
Opposed	Approval–		Withhold info	Annoy
			Frame/topic	

from these theories to validate the labels for the factors. This matching of factors to concepts is a data-driven interpretation of the factors (see Sect. 16.1.2) and might bring us closer to a “correct” interpretation of a factor.

In Fig. 16.4, we show the cumulative score the observers gave the concepts for each factor. The colour coding in this figure indicates how much the observers agreed that the concept could explain the factor: the dark-coloured cells indicate unanimous agreement, the light-coloured cells indicate that three out of the four observers agreed. The initial factor analysis interpretation of the factors is indicated with an asterisk.

The fit of the factors and the concepts determines the validity of the interpreted theory for this factor. The observers unanimously matched most factors to the concepts corresponding with the initial factor analysis interpretation of the factor; see Fig. 16.4 in which the asterisks indicate the initial interpretation and the dark cells indicate unanimous matching. The factors where the observers disagreed (not unanimous) with the initial factor interpretation are police factors 1, 9, 11, and 13 and suspect factors 1, 3, and 10 (see Fig. 16.4).

We see several explanations for this disagreement; first, the initial subjective interpretation of the factors might have been wrong. The factors with a higher number had fewer items loading on them (and less explanatory power), which might have made it more difficult to interpret them. Four of these higher factors (suspect factors 9, 11, and 13 and police factor 8) had few and diverse items loading on them, which resulted in disagreements during the initial factor interpretation. It is likely the disagreement persisted in the current analysis for suspect factors 9, 11, 13 and police factor 10. Second, factors could initially have been interpreted as having an “absence of something”. This was the case for police factor 1 which was initially interpreted as “missing rapport”. In the subsequent “mapping concepts to factors” task, the observers did not unanimously match the concepts of rapport to police factor 1. This might be because the instructions were unclear what to do when a concept was explicitly absent: some observers said that this factor contains information about the concept rapport (i.e. rapport is missing) while others said there is no rapport so the concept of rapport is not present. For suspect factor 3 and 10 we can give no alternative explanation and conclude that our initial factor interpretation was incorrect (see Table 16.5).

16.4.3 *Relations Between Theories*

The observers unanimously matched factors to concepts from theories that were not initially included in the factor analysis. In other words, more concepts than initially come to mind might play a role in explaining a factor. For example, suspect factor 4 was interpreted during factor analysis as a *together* stance, but the factor was not only matched to the concept *together* (from theory of stance), but also to *positive approval* (from the theory of face), and *attention* and *positivity* (from rapport) (see Fig. 16.4).

Factor	Stance		Face		Rapport		Information			Strategy		
	Dom Sub	Tog Opp	Aut+ App+ Aut-	App-	Coor Att Pos	Quest G.	Inf Lie W.	Inf Frame	Confr	Sur	Evade	Annoy
P 1	2	4*	2	4	2*	1*	2*	1	1	1	2	4
P 2	4*	4*	4	4	2*	4*	4*	1	2	1	3*	3
P 3	1	3	1	3	1	2	1	4*	4	3	1	3*
P 4	4*	1	1	2	1	2	1	1	2	3	1	3
P 5	2	1	2	4	3	3	4	1	1	1	1	1
P 6	4*	1	4	2	1	1	1	2	1	4*	1	1
P 7	3*	1	1	4	1	1	1	1	1	2	1	1
P 8	1*	1	2	3	1	1	1	1	1	1	1	1
P 9	4*	1	2	2	1	1	2	1	1	1	1	1
P 10	1	2	2	1	1	1	1	1	1	3*	3	1
P 11	1	2	2	1	2	1	1	1	2*	2	1	1
P 12	4*	1	1	3	2	1	1	3	2	2	1	1
P 13	1	3	3	2	1	3	1	2	1*	1	1	1
S 1	4*	4*	* 3*	4*	1	2	1	1	1	3	2	4
S 2	4*	4	3	4	4*	4*	4*	1	1	1	2	4
S 3	4*	1*	1*	2*	3*	1	1	1	2	2	3	3
S 4	1	4*	2	4	3	4	4	2	2	2	3	3
S 5	2	1	1	4	1	1	1	1	2	3	1	1
S 6	2	1	2	4	2	3	3	4*	1	1	2	4*
S 7	1	2	3	1	1	1	1	1	3*	3	4	2
S 8	1	1	1	3	2	1	2	1	2	1	4*	2
S 9	1	2	3	1	2	2	1	1	1	1	1	1
S 10	1	2	3*	1*	2	2	2	1	1	1	1	1
S 11	2	1	2	3	1	2	1	1	1	1	1	1
S 12	1	4	2	3	4*	4*	4*	1	1	1	1	1
S 13	1	1	3	3	1	1	1	1	1	1	1	1

Fig. 16.4 The matching of concepts from the models with the factors derived from the clustering of terms. The numbers indicate how many (out of 4) observers thought the concept from the theory (*column*) fit the factor (*row*). *Asterisks* denote the initial interpretation of the corresponding factor

Table 16.5 The factors from the factor analysis with the concepts that were unanimously matched to the factors

	Factor	Interpretation based on concepts
Police	1	Opposed, negative approval, annoy
	2	Together, positive approval, attention, positivity
	3	Lie, withhold info
	4	Submissive
	5	Together, positive approval, positivity
	6	Dominant, negative autonomy, confront
	7	Opposed, negative approval
	8	–
	9	Dominant
	10	–
	11	–
	12	Dominant
	13	–
Suspect	1	Dominant, opposed, negative approval, annoy
	2	Together, positive approval, coordination, attention, positivity
	3	Submissive
	4	Together, positive approval, attention, positivity
	5	Opposed, negative approval, annoy
	6	Questioning
	7	Evade
	8	Surround
	9	–
	10	–
	11	–
	12	Together, coordination, attention, positivity
	13	–

A dash means no concepts were unanimously matched to the factor

Our methodology makes clear how the theories are related to each other. For each factor, concepts from different theories can be applicable. This *co-occurrence of concepts* suggests that the corresponding theories are related. In Table 16.6 we show in how many factors concepts co-occur. For example, the interpersonal stance *together* co-occurs in more than one factor to: *positive approval*, a concept underlying face (4 co-occurrences), and *coordination* (2), *attention* (4), and *positivity* (5), which underlie rapport (see Table 16.6). This is indicative of a strong link between these concepts. The concept of *dominance* co-occurs with the concepts: *opposed stance*, *negative autonomy*, *negative approval*, *confrontation* and *annoy* (see Table 16.6). To investigate this relation further, we look at Fig. 16.4, which shows that suspect factor 1 was matched with the concepts dominance, opposed, negative approval, and annoy. For the police factors dominance co-occurs with negative autonomy and confront. This is likely due to the different roles the

Table 16.6 The observed relations between the theories based on the co-occurrence of their concepts in the factors

	Stance			Face			Rapport			Info			Strategy	
	Dom.	Tog.	Opp.	App+	App-	App+	App-	Att.	Pos.	Lie	W. info	Confront	Annoy	
Stance	x		1		1							1		1
Dominant														
Together		x		4		2	4	5						
Opposed	1		x		4									3
Face			4	x		1	3	4						
App+														
App-	1		4		x									3
Rapport			2	1		x	2	2						
Coordination														
Attention		4		3		2	x	4						
Positivity			5	4		2	4	x						
Info										x	1			
Lie														
Withhold info										1	x			
Strategy													x	
Confront	1													
Annoy	1		3		3									x

The numbers indicate how often these concepts co-occurred. Only the concepts rated unanimously present in the factors are shown

interactants have. The police officer assumes a dominant stance when he confronts the suspect with an incriminating fact: the act of confronting is dominant. This might be strengthened by the power the police officer has, as he dictates the course of the interview: a concern for the autonomy of the suspect. A dominant suspect might use the strategy annoy to intentionally thwart the progress of an interview and this could negatively impact the approval of the officer. Linssen et al. (2013) proposed that interpersonal stance and politeness (face) are related. They suggest that the dimensions of power and affect used in the model of Leary's Rose can be mapped to the dimensions of face: autonomy and approval. For example, when a person is very dominant, she does not take the other's autonomy into account. A similar relation holds for the dimension of affect, as a person who is opposed to someone else expresses disapproval of that person.

We further investigate the relations between the different concepts in the next section and illustrate them using examples from the corpus. The related theories will be integrated to form the basis for a computational model. As each theory describes relations between the cause and effect of behaviour in an interaction, a virtual tutoring agent (virtual suspect) could use a computational model of these theories to predict the effects of its behaviour in an interaction with a human user (see Sect. 16.6 on future work).

16.5 Illustration of Relations

In the previous section, we showed that certain concepts underlying the theories appear to be related based on the data from our corpus (see Fig. 16.4 and Table 16.6). Here, we illustrate several of these links with example fragments from the corpus that were not used for annotation and the subsequent factor analysis. We illustrate the *co-occurrence of concepts* (see previous section) that shows the relation between the concepts of different theories. Also, we illustrate how our findings might be extended to explain the dynamic aspects in a police interview.

16.5.1 Co-occurrence of Concepts

We found the strongest links between the *together* stance and *positive approval* concepts and between the *opposed* stance and *negative approval* concepts (see Table 16.6). An example from the Brintjes scenario (see transcript below) shows a together stance occurring together with positive approval. In this fragment, the police officers are asking questions about the suspect's leisure time, to which the suspect responds that she spends most of her time at the mall with her girlfriends. The police officers respond to this by indicating that they understand what she means ("Just chilling.") and they all start laughing about this. In this moment, the police officers are very much trying to sympathise with the suspect, thereby

adopting a together stance. They are also expressing approval by saying that they understand the suspect's wish to stay at the mall. In the preceding section, we also showed that there is a strong link between the together stance, concepts underlying rapport (particularly attention and positivity) and positive approval. In the corpus fragment about the suspect staying at the mall, it is clear that both interaction parties are paying a high degree of attention to each other. One of the police officers is asking questions about the suspect's activities which yield immediate responses from the suspect. There is, however, no uncomfortable atmosphere during this part of the conversation, as both the officers and the suspect start laughing about this topic. Thus, the concepts of a together stance, positive approval, and both attention and positivity are displayed in this part of the conversation.

POLICE OFFICER: Those girlfriends, eh, 'cause you said you go shopping with your girlfriends...

SUSPECT: Mm mm. [Confirmatory.]

P: Do you have good friends? Tight friends?

S: [Nods enthusiastically.] Yes.

P: Yes?

S: Yes.

P: So what do you go and do with your friends?

S: Yeah, well, basically, we are often at the mall.

P: At the mall?

S: Yeah, one of those indoor malls.

P: And what do you do there?

S: [Softly:] Kind of hanging around. [Laughs.]

P: [Laughs.] Just chilling.

S: [Laughs.] Yeah!

An example of the strong link between an opposed stance and negative approval can be found in the Wassink scenario (see transcript below). In this excerpt, Mrs Wassink, the suspect, is asked whether she wants to cooperate with the police officer by answering some of his questions, because he wants to form a picture of her situation. Mrs Wassink does not comply and indicates that she does not see the point of doing so.

POLICE OFFICER: I don't know you and you don't know me either.

SUSPECT: No.

P: But maybe it would be convenient if we would first discuss some things about you—about who you actually are. Do you think that's OK?

S: Well... Why?

P: You don't think that's useful?

S: [Shrugs, shakes her head.] No, I don't know why I should tell you who I am.

P: Yeah. [Short pause.] Well, I would like to know.

S: But what for?

- P: Because I would like to know who I'm talking to before I can talk with you—what you just indicated, that you might have physically abused someone.
- S: [Stares blankly.]
- P: Do you understand what I'm saying?
- S: Yes, I do, but that you would like it, well... [Shrugs.] Are you going to tell me something about yourself as well or...? [Shakes head.]
- P: Well, I don't know, would you be interested in what I would have to say?
- S: No, but I also don't understand why you would be interested in me.

In the above fragment, the suspect gives snappy replies to the police officer's questions. This is a first indication that she is not trying to cooperate: she seems to have a very opposed stance to the officer, his approach and his proposals. With her behaviour, the suspect also expresses disapproval (negative approval), repeatedly shaking her head and shrugging, indicating that she does not agree with the police officer or just does not care. Moreover, Mrs Wassink goes a bit beyond simply disagreeing, as she seems to intentionally annoy the police officer. She does this by questioning the police officer's approach (repeatedly asking of what use it is), by expressing that she does not understand what is going on, and by asking a counter-question to the police officer (whether he will say something about himself as well). Here we see that an opposed stance, negative approval, and an annoy strategy occur together.

We found several other relations between concepts from the theories we used that occurred less frequently in our collection of factors. On some occasions, the police officer, but mostly the suspect, used a confront strategy which was accompanied by a dominant stance and negative approval. In these cases, the suspect was trying to lead the conversation by confronting the police officer(s) with his or her own opinions (which were negative in nature most of the time). Another striking co-occurrence of concepts is that of the concepts underlying rapport. As Tickle-Degnen and Rosenthal (1990) assert, coordination, attention and positivity generally occur together to form the feeling of rapport and this is confirmed by our observations.

16.5.2 Concept Dynamics

Our approach to analysing the corpus of police interviews hinges on the annotation of short fragments. However, our annotation did not capture the dynamic aspects of the interviews, for example how and why people change stances or how their feelings of rapport increase or decrease. Here, we illustrate how this may work by describing a change in a situation in the Van Bron scenario (see transcript below) in terms of the concepts from Table 16.6. In this case, the suspect is asking the police officers their name and surname in a dominant way. One of the police officers immediately agrees to give his surname, but the other only gives his first name. To this the suspect replies by making a small gesture with his hand, implying that

he also wants to know the officer's surname. The officer then responds in a laconic way by saying "Oh, you want my *surname*?" which leads to the suspect imitating the officer's response and adding "Wise guy." What we witness here is an exchange in which the police officer has an opposed stance towards the (dominant) suspects and acts disapprovingly of the suspect's behaviour by not granting a full answer. This, in turn, elicits a similarly disapproving response by the suspect. Thus, over time, the suspect changes from having a dominant stance towards one that is more opposed to the officer because of the latter's behaviour.

SUSPECT: [Points at police officer 1.] What was your name again?
 POLICE OFFICER 1: **** [Gives his first name.]
 S: And...? [Makes a gesture with his hand for the police officer to complete his name.]
 P₁: **** [Gives his surname.]
 S: ****? [Repeats the name.]
 P₁: Yeah.
 S: [Chuckles.] And you? [Points at police officer 2.]
 POLICE OFFICER 2: **** [Gives his first name.]
 S: [Makes the 'completion' gesture again.]
 P₂: What is that? [*Mimics the gesture.*]
 S: [To P₁.:] He doesn't understand? **** [P₁'s surname.] And you? [Points at P₂.]
 P₂: Ah, you want to know my last name?
 S: [Stares at P₂.]
 P₂: Yeah, if you could just be clear in your questions...
 S: He understands [Points at P₁.] are you a bit stupid or something like that?
 P₂: Yeah, I'm a bit more stupid than him, OK...
 S: That's clear.
 P₂: [Softly:] All right.
 S: No, it's not all right. What is your surname?
 P₂: **** [Gives his surname.]
 S: Ah... **** [Repeats P₂'s surname.] [Softly:] Wise guy.

16.6 Conclusion and Discussion

In this paper, we presented our methodology for analysing the behaviour of police officers and suspects in a corpus of enacted police interviews. Taking a holistic approach, we described fragments of this corpus in short terms that captured the behaviour of the participants in, and the atmosphere of, the interviews. We used a factor analysis to cluster related terms based on ratings of observers who annotated to what extent these terms were applicable to each fragment. Based on the factors we found, we selected theories from (social) psychology that we intuitively thought could explain these factors. We included theories about interpersonal stance, face

and politeness, and rapport and defined two meta-concepts, namely “information” and “strategy”, to account for the interpretations that remained. To determine whether these theories matched the factors, we investigated whether these factors could be explained by the concepts underlying the theories. We found that our initial factor interpretation and the match between factors and concepts overlapped broadly. We also found that many factors were matched to more concepts than initially were associated with the factors. We used this finding to create a collection of interrelated concepts that gives insight into how the different theories relate to each other. With this collection, we are able to (at least partly) describe the behaviour of both police officers and suspects in an interview setting.

Our combination of holistic and theory-driven methodology does, however, have its limitations. As is the case in most observational studies, our annotations of the police interview corpus were based on our interpretations of the behaviour of the interacting parties and thus subjective. For future work, our methodology may be repeated to include more observers (and more independent observers) which may lead to a broader semantic frame, possibly alleviating problems inherent with interpretation of behaviour.

A second limitation of our approach is that it currently focuses on describing *short* fragments from the corpus. In the previous section however, we illustrated how our findings may be extended to explain changes in the behaviour of interacting parties over longer periods of time. We based these examples on how temporal aspects are explained by the theories from which we drew our concepts. We wish to continue this line of research by investigating how the interplay of these concepts influences the dynamics in police interviews. This may, for instance, be done by locating moments in our corpus in which a person’s behaviour changes. For example, there may be moments when a person changes his or her stance or becomes less polite. Someone may also consciously change his behaviour to evoke desired behaviour of the other party. This may, for example, be the case when a police officer adopts a “together” stance to build rapport with a suspect. Thus, the communicative contexts before and after this type of changes in behaviour should be compared, to discover what may have caused the change in behaviour. This causality is of vital importance for the creation of a virtual suspect agent, as such an agent needs to be capable of taking logical (and explainable) actions. Thus, this calls for an extended empirical study of the corpus. Such a study may also validate the links we found between concepts, as our current work only investigated a number of fragments from our corpus.

Lastly, we also wish to investigate how the methodology we present in this paper translates to other domains. Whether our approach can be used to analyse communicative behaviour in other domains depends on the availability of a corpus and theories on interaction that explain (parts of) the behaviour. A related domain in which we are also involved is that of street interventions by police officers with loitering juveniles. This domain features a different setting and the environment imposes other restrictions on the interaction, such as an easier “way out” for the juveniles because they are not kept in a room like the interviewees. Still, this domain does not differ strongly from the domain discussed in this paper, as they are both

related to police work and display the unique features this work has, such as the status of the police officer. It remains to be investigated whether our approach would allow for feasible analysis of behaviour a completely different domain. However, given a sufficiently rich corpus of such interactions, we expect that our methodology can be used to analyse corpora from other domains as well.

In future work, we will construct a mental model for virtual agents in a police interview setting. As indicated above, we will focus on the dynamics of such interviews, establishing a computational model that enables a virtual agent to perform causal reasoning. This system will go beyond being an “autonomous sensitive artificial listener” as in Schroder et al. (2012). The system will be able to use the “mood” of its mental model to select the most appropriate action it has available. The current work will inform the creation of this model, which will in turn be used for virtual agents in a tutoring application. Thus, having related data to theories, our next step will be to relate theories to practice.

Acknowledgements This publication was supported by the Dutch national program COMMIT.

References

- Allan K (2001) *Natural language semantics*. Blackwell Publishers, Oxford
- Austin JL (1975) *How to do things with words*, vol 1955. Oxford University Press, Oxford
- Bateson G (1955) A theory of play and fantasy. *Psychiatr Res Rep* 2(39):39–51
- Bickmore T (2008) Framing and interpersonal stance in relational agents. In: *Functional markup language workshop at AAMAS 2008*
- Brown P, Levinson SC (1987) *Politeness: some universals in language usage*. Cambridge University Press, Cambridge
- Bruijnes M (2013) Affective conversational models: interpersonal stance in a police interview context. In: *Humaine Association conference on affective computing and intelligent interaction*, pp 624–629
- Bruijnes M, Kolkmeier J, op den Akker R, Linssen JM, Theune M, Heylen DKJ (2013) Keeping up stories: design considerations for a police interview training game. In: *Social believability in games workshop at ACE 2013*
- Campos H, Campos J, Martinho C, Paiva A (2012) Virtual agents in conflict. In: *Intelligent virtual agents*, pp 105–111
- Cassell J, Gill AJ, Tepper PA (2007) Coordination in conversation and rapport. In: *Proceedings of the workshop on embodied language processing*. Association for Computational Linguistics, Prague, pp 41–50
- Core MG, Lane HC, Van Lent M, Gomboc D, Solomon S, Rosenberg M (2006) Building explainable artificial intelligence systems. In: *Proceedings of the national conference on artificial intelligence*, vol 21
- Culpeper J, Bousfield D, Wichmann A (2003) Impoliteness revisited: with special reference to dynamic and prosodic aspects. *J Pragmat* 35(10):1545–1579
- Fillmore C (1981) Pragmatics and the description of discourse. In: Cole P (ed) *Radical pragmatics*. Academic Press, New York, pp 143–166
- Giebels E (2002) Beïnvloeding in gijzelingsonderhandelingen: De tafel van tien. *Nederlands tijdschrift voor de psychologie* 57:145–154
- Goffman E (1959) *The presentation of self in everyday life*. Garden City, New York

- Gupta S, Walker M, Romano D (2007) How rude are you? Evaluating politeness and affect in interaction. In: Proceedings of the international conference on affective computing and intelligent interaction, pp 203–217
- Holmberg U, Christianson SÅ (2002) Murderers' and sexual offenders' experiences of police interviews and their inclination to admit or deny crimes. *Behav Sci Law* 20(1–2):31–45
- Huang L, Morency LP, Gratch J (2011) Virtual rapport 2.0. In: Intelligent virtual agents. Springer, Berlin, pp 68–79
- Koops M, Hoevenaar M (2013) Conceptual change during a serious game: using a Lemniscate Model to compare strategies in a physics game. *Simul Gaming* 44(4):544–561
- Kopp S, Gesellensetter L, Krämer NC, Wachsmuth I (2005) A conversational agent as museum guide-design and evaluation of a real-world application. In: Intelligent virtual agents, Springer, pp 329–343
- LaForge R, Suczek RF (1955) The interpersonal dimension of personality. *J Personal* 24(1):94–112
- Leary T (1957) Interpersonal diagnosis of personality: functional theory and methodology for personality evaluation. Ronald Press, New York
- Linssen JM, Theune M, Heylen DKJ (2013) Taking things at face value: how stance informs politeness of virtual agents. In: Workshop on computers as social actors at IVA 2013
- Mann S, Vrij A, Bull R (2004) Detecting true lies: police officers' ability to detect suspects' lies. *J Appl Psychol* 89(1):137
- op den Akker R, Bruijnes M, Peters R, Krikke T (2013) Interpersonal stance in police interviews: content analysis. *Comput Linguist Netherlands J* 3:193–216
- Ravenet B, Ochs M, Pelachaud C (2013) From a user-created corpus of virtual agent's non-verbal behavior to a computational model of interpersonal attitudes. In: Intelligent virtual agents. Springer, Berlin, pp 263–274
- Richardson BH, Taylor PJ, Snook B, Conchie SM, Bennell C (2014) Language style matching and police interrogation outcomes. *Law Human Behavior* 38(4):357–366
- Rogan RG (2011) Linguistic style matching in crisis negotiations: a comparative analysis of suicidal and surrender outcomes. *J Police Crisis Negot* 11(1):20–39
- Roukhout D, Schacht R (2000) Ontwikkeling van een nederlandstalig interpersoonlijk circumplex. *Diagnostiekwijzer* 4:96–118
- Schroder M, Bevacqua E, Cowie R, Eyben F, Gunes H, Heylen D, Ter Maat M, McKeown G, Pammi S, Pantic M et al (2012) Building autonomous sensitive artificial listeners. *IEEE Trans Affect Comput* 3(2):165–183
- Searle JR (1969) Speech acts: an essay in the philosophy of language, vol 626. Cambridge University Press, Cambridge
- Steunebrink BR, Dastani M, Meyer JJC (2012) A formal model of emotion triggers: an approach for BDI agents. *Synthese* 185(1):83–129
- Swartout W (2010) Lessons learned from virtual humans. *AI Mag* 31(1):9–20
- Tannen D (1993) What's in a frame? Surface evidence for underlying expectations. In: Tannen D (ed) Framing in discourse. Oxford University Press, New York, pp 14–56
- Thomas KW (1992) Conflict and conflict management: reflections and update. *J Organ Behav* 13(3):265–274
- Tickle-Degnen L, Rosenthal R (1990) The nature of rapport and its nonverbal correlates. *Psychol Inquiry* 1(4):285–293
- Traum D, Marsella SC, Gratch J, Lee J, Hartholt A (2008) Multi-party, multi-issue, multi-strategy negotiation for multi-modal virtual agents. In: Intelligent Virtual Agents, Springer Berlin Heidelberg, pp 117–130
- Vaassen F, Wauters J (2012) deLearyous: training interpersonal communication skills using unconstrained text input. In: Proceedings of the 6th European conference on games based learning, pp 505–513

- Vinciarelli A, Pantic M, Heylen DKJ, Pelachaud C, Poggi I, D'Ericco F, Schroeder M (2012) Bridging the gap between social animal and unsocial machine: a survey of social signal processing. *IEEE Trans Affect Comput* 3(1):69–87
- Walker MA, Cahn JE, Whittaker SJ (1997) Improvising linguistic style: social and affective bases for agent personality. In: *Proceedings of the international conference on autonomous agents*, pp 96–105
- Wang WY, Finkelstein S, Ogan A, Black AW, Cassell J (2012) “Love ya, jerkface”: using sparse log-linear models to build positive (and impolite) relationships with teens. In: *Proceedings of the 13th annual meeting of the special interest group on discourse and dialogue*, pp 20–29

Chapter 17

From Personalization to Parrhesia: A Multimodal Analysis of Autobiographical Recalls in Barack Obama’s Political Speech

Giovanna Leone, Francesca Di Murro, and Livia Serlupi Crescenzi

17.1 Beyond Personalization: An Analysis of Barack Obama’s Autobiographical Recalls

Since his surprising victory as a former outsider in the race to become the president of the USA, many scientific works have been devoted to exploring how much of the political force of Barack Obama may be explained by taking into account the features of his unconventional life (Da and McClain 2009; Zogby 2009; Hammack 2010). In this interesting field of study, a specific facet that we will try to understand more in depth here is the way in which Barack Obama himself uses self-reflections on his own autobiographical novelty – which made him a somehow “strange” incumbent senator from Illinois –, as a persuasive tool aimed at enhancing his political rhetorical skills. We will propose the idea that at these moments of his political speech, personalization turns into *parrhesia* (Foucault 2001): a way of speaking the truth about himself that allows this leader to later propose to his audience to use the same frankness to cope with controversial points exposed in subsequent arguments of his political speech, introduced after this first self-exposure.

Within this general framework, a previous paper (Leone 2013) already pointed out that the theory of personalization – a widespread aspect of political leadership characterizing present-day democracies that can be traced back to Thatcher’s and Reagan’s election times (1979, 1980), or even to Trudeau’s election time (1968) (McAllister 2007) – may not fully explain the specific features of autobiographical recalls of Barack Obama, often shared with his audience during his political speeches. In the present work, we pursue this line of thought in more detail by

G. Leone (✉) • F. Di Murro • L. Serlupi Crescenzi
Sapienza Università di Roma, Dept. CORIS, Via Salaria 113, 00198 Rome, Italy
e-mail: Giovanna.leone@uniroma1.it

suggesting that perhaps the main characteristic that makes Obama's choice to include autobiographical recalls in his political speeches so effective is linked to his parrhesia, i.e., his risky decision to speak frankly about difficult or controversial aspects of his own origins (Foucault 2001).

To explore empirically this new line of thought, two aspects must be developed.

First, a theoretical hypothesis must be advanced to better understand why these self-disclosures of the social disadvantages of his origins may bolster Obama's public image, rather than detract from it, as might have been expected at a first sight.

Second, from a methodological point of view, innovative tools must be developed because, according to this new theoretical framework, an analysis of the verbal contents of Obama's statements is insufficient, and his body language must also be taken into account. This methodology aims to grasp the powerful blend that is observable in these communicative acts, conveying at the same time a deep self-awareness of social prejudices targeting him and his family and a strong expression of pride and self-confidence when freely sharing with his audience such difficult and even humiliating contents.

17.2 Memories of Being a Pariah: The Importance of the Parrhesiastic Attitude Characterizing Obama's Recalls

One aspect of the novelty of Barack Obama's autobiographical recalls that makes them somehow a unicum in contemporary political speech is the frank attitude used by him in stressing the difficulties linked to the unconventional circumstances of his birth. To better grasp this aspect, two theories may be invoked. The first theory, proposed by Hannah Arendt (1978), is based on the difference between self-conscious pariah and parvenu. The second theory, proposed by Michel Foucault (Foucault 2001) in the last part of his scientific work, is linked to the analysis of parrhesia, i.e., the communicative choice to speak clearly and fearlessly about difficult topics. Within the theoretical framework proposed in this paper for analyzing how Obama recalls his autobiographical memories during political speech, both these theories are combined to explain why these communicative actions of frankly describing the more embarrassing and sometimes even humiliating aspects of his life, which have the potential to besmirch or threaten his image, are on the contrary one of the keys to his success.

The first theory may help us to better understand the importance of the contents shared in Obama's memories. The second theory may allow us to grasp the similar importance of the multiple modalities used for this social sharing: considering at the same time not only the words but also the way in which they are conveyed by Obama's body language: his face, his gaze, his voice, his gestures, his posture, his emotive reactions to the difficult moments reenacted during his memories.

The theory proposed by Arendt (1978) on the difference between a self-conscious pariah and a parvenu was inspired by the concept of *pariah*. This word refers to the rigid division of classic Indian society into social groups or castes that are expected not to merge or to have reciprocal contacts. This rigid attitude may be explained by the religious belief that being born into a particular caste is due, not to chance, but rather to acts committed by the newborn during previous lives. Therefore, people of low or no caste are judged to be unworthy of social opportunities since they are expected to atone in this current existence for the improper acts committed earlier. Referring to this old social model, the word *pariah* arrived in time to be used as a way to analogically represent any social outcast. Taking inspiration from these classic remarks, Arendt (1978) explored how the word *pariah* could be used to address the social danger linked to a person's stigmatized social origins. To better understand this point, she examined (Arendt 1997) the diaries and letters of Rahel Varnaghen, a German-Jewish writer who hosted one of the most prominent salons in Europe during the late eighteenth and early nineteenth centuries. In Varnaghen's salon, the greatest scientists and artists of the time debated the most pressing topics of their culture, which made Varnaghen a big social success. However, studying Varnaghen's letters and diaries, Arendt identified a growing psychological hardship that drove this successful woman to eventually abandon this way of life and to change dramatically her interactions with her social circle. The crucial point of Varnaghen's inner and social changes, which shocked friends and relatives, was the unhappiness she felt surrounding her ethnic and cultural background, which caused her to hide her shameful birth, take her husband's name, and remain silent on her Jewishness. In the last part of her life, Rahel made the bold decision to proclaim, in a very public manner, her birth into a Jew family, going so far as to study Hebrew so that she could write her last letters in that language. In that decision of Varnaghen, Arendt sees the end of a struggle, both psychological and social, between the attitude of becoming a parvenu, perhaps successful but always threatened with being destroyed if the truth were to become widely known, and the attitude of the *self-conscious pariah*, fearlessly exposing the facts of her birth to social judgment. This second attitude is, in Arendt's opinion, the most efficient one, both psychologically and socially. From the point of view of the individual life, in fact, to be aware of the conditions of one own's birth and to accept them gratefully are the main roots of personal well-being because such an awareness signals that the individual is free from the fear of social refusal and aims to defend the originality of her personal life. In addition, from a social point of view, the choice of attitude of the self-aware pariah is the most efficient one because it signals a lack of selfish goals, always implied when a lie is devised to manipulate others' knowledge about social connections (Arendt 1978).

By this last consideration Arendt implicitly encouraged her Jewish friends, sharing with her the social status of refugee from the Nazis, to openly declare their pariah origins, which caused them to be grouped among the socially despised and outcast. In that attitude, the philosopher recognized the best example of her theoretical point, which seized the natality of each newborn as the key feature that

made social advancement possible, much more so than some obscure psychological quality, such as creativity, originality, or any other charismatic gift endowing special leadership qualities.

The same respect for factuality shown by Arendt's theory on the personal and social values expressed by the self-aware pariah is the root of Foucault's (2001) theory on *parrhesia*, seen as the most advantageous communicative choice when speaking about controversial or difficult aspects of personal and social life. In this work, developed in the last part of his life, Foucault tried to grasp the main characteristics of the decision to speak frankly and fearlessly about difficult topics. What distinguishes *parrhesia* from other social circumstances when it is essential to speak the truth (e.g., when conveying scientific knowledge or technical details) is the fact that in the case of *parrhesia* the one who is speaking accepts the risk of her frank and open speech. This is how the concept was first described in ancient Greek dramas (e.g., in Euripides's plays) or in ancient Greek philosophy (e.g., in the teachings of the Cynic School). Here, good examples of *parrhesia* include the words of Teiresias, an oracle who revealed to Oedipus his tragic destiny, or the episode of Demosthenes replying to his king that the only desire of Demosthenes that the king could really fulfill would be to step aside so that the philosopher could enjoy the sunlight. Later on, this same idea of *parrhesia* entered the first Christians' philosophy, though in slightly changed form because of the Gospels' teachings on truth. On the one hand, Jesus' words "the truth shall set you free" stressed the empowering effect of *parrhesia*, reducing the fear that always accompanies a lack of knowledge and awareness. On the other hand, his decision never to lie about his religious proposal – a decision that was subsequently repeated by many of his disciples – was perhaps the best expression of a generous attitude toward society, full of hope that the future would be more open to change and understanding than their present society would allow. Taking together both the ancient Greek considerations about *parrhesia* and the later importation of this concept into Christian culture, Foucault came to propose the idea that *parrhesia* was perhaps the most empowering communicative act in social discourse. On the one hand, he considered *parrhesia* the best way to take care of oneself, refusing to hide or hinder one own's thoughts or inner states. On the other hand, he stressed that *parrhesia* was the best way to empower the disempowered as well, giving them the tools to better understand the situations they find themselves in. In this last consideration, however, recipients must agree to play the stressful "communicative game" that the *parrhesiastic* speaker is inviting them to play (Foucault 2001).

The following in-depth analysis of four autobiographical recalls, embedded in some famous political speeches of Barack Obama, aims to explore both the verbal contents and the body communication evident in these extracts by framing them according to both of Arendt's (1978) theory on self-aware pariahs and Foucault's (2001) theory on *parrhesia*. To achieve this aim, however, an original methodology was needed that would combine a tool focused on facial expressions [facial action coding system (FACS)] (Ekman and Friesen 1978) with the multimodal analysis of communication proposed by Poggi (2007) that takes into account simultaneously the verbal content and body language used during the recollections of autobiographical

events. What we expect, in fact, is that not only words, but also all bodily communication when conveying them would turn out to be important when a parrhesiastic communication is delivered to an audience. In the case of Barack Obama, we expect more in particular that his body communication, when recalling in a parrhesiastic way his difficult childhood as a pariah, would express at the same time the negative emotions enacted by these memories and the emotions' regulation (Frijda 2013) needed to turn these aspects of his life into a tool to communicate to audiences a trusting attitude toward his political proposal of empowerment and of social change.

17.2.1 *Methodological Choices*

The analysis was conducted on a selection of four public speeches made by Barack Obama in the period from 2004 to 2009. The first speech was made in the United States, when he was still a senator from Illinois. The other three speeches were made abroad during his international trips after becoming president of the United States. In all the speeches there reoccurs, as a constant of his political discourse, the choice to share with his listeners some of his autobiographical memories. These moments of reminiscence were studied by combining a multimodal analysis of communication (Poggi 2007) and an analysis of facial expression and emotions conducted through the use of "facial action coding system" (FACS) (Ekman and Friesen 1978).

The coding of the facial expressions through "facial action coding system" (FACS) begins with the assumption, originally proposed by Darwin [1872], of the regularity of muscular movements in the face as an emotional reaction, which was linked to the idea that the communication of emotions is innate and universal. This analysis permits, then, the detection of the expression of emotions through the recognition of a typical configuration of movements of facial muscles, present in the atlas of possible facial expressions given by FACS.

A multimodal analysis of communication begins instead with the assumption that, just as a set of rules exists that, acting together, create a language, so communication is created by the joint actions of the rules that underlie facial expressions, modulation of the voice, gestures, and body movements.

Thus, this technique of analysis proposes considering jointly *five principal modalities* of transmission of communicative signals:

- *Verbal modality*: based on the analysis of words;
- *Prosodic-intonative modality*: based on the analysis of the voice, with attention to the temporal aspects of the speech, the rhythm, pauses, length of the vowels and accents, intensity, and tone;
- *Gestural modality*: analysis of gestures based on hand, arm, and shoulder movements;

- *Facial modality*: analysis of the gaze, head movements, smiles and laughs, facial expressions, and mouth movements (in our case, the analysis is integrated by “facial action coding system” (FACS) analysis);
- *Corporeal modality*: analysis of the posture, torso and leg movements, orientation in relation to interlocutors, and movement of the body in space.

All the productive modalities are used simultaneously and synchronized with speech, each one offering a precise semantic contribution with different goals. In fact, when senders deliver the verbal contents of their discourse, conveying the information at their disposal to the world, they communicate at the same time through the other modalities the purposes of their communication, the thoughts and feelings they have about their specific message (Poggi 2007). The addressee, furthermore, observing the body of the speaker while they listen to their words, in a few seconds perceptively acquires or infers specific knowledge about the sex, age, ethnic and cultural background, and personality of the sender. This information is often communicated by senders against their own will. This information, deduced by recipients through observation of the speaker’s body, interacts with the effects of the speaker’s strategies of self-presentation (Goffman 1967), that is, with the image that the speaker wishes to give of him- or herself, consciously producing signals or monitoring them while speaking, to induce a type of specific perception of the speaker in recipients.

Describing speech as a sort of “communicative symphony,” Poggi (2007) proposes the score of multimodal communication as the best tool to transcribe, analyze, and classify the joint action of several signals transmitted in the different modalities present in a communicative fragment.

In the score, the five aforementioned modalities are written and analyzed on parallel lines as in a musical score. Thus, for each signal of each modality, analysis takes place on five levels:

- *DS* is the description of the signal; it relates to the physical characteristics of movement, gestures, gaze, posture and vocal elements in terms of length, intensity, fundamental frequency, and pauses for prosodic-intonative signals.
- *Ts* is the type of signal; each signal may be classified, for instance, batonlike gesture, deictic gaze, iconic gesture (Ekman et al. 1984). The gestures can be, for instance, *deictic*, when they point to an object or person with the index finger or open hand; *iconic*, when they draw in the air a form or imitate the typical movements of an object, animal, or person; *symbolic*, when a gesture in a cultural moment has a meaning easily translatable into words or phrases. Lastly, the *batonlike* gesture, where the hands go from up to down to emphasize what is spoken; batonlike gestures are often unconscious.
- *S* is the meaning of each signal, its verbal translation.
- *TS* is the type of meaning; each signal is classified as information about the world (Imo), information about the identity (IIM) of the sender, or the mind (IMM) of the sender.
- *F* is the function, the semantic relation between the signal being analyzed and the concomitant verbal signal or another signal taken as a point of reference and that

may be *repetitive* when the two signals have the same meaning, *additional* when a signal adds a meaning congruent with that of another signal, *contradictory* when the meaning of the signal in question contrasts with that of a concomitant signal, or *independent* when the signal in question is not in relation to another signal produced simultaneously because they are part of two independent plans of action.

In multimodal communication the signals expressed in the various modalities combine and are integrated with coherence. Sometimes, however, the signals manifest discordant, clashing, or contradictory meanings. These are cases of error, ambivalence, and deception in communication. Contradictory communicative behaviors, for instance, are a sign of deception. In the hypothesis where deception is expressed in relation to genuinely felt emotions, the emotions seep out through an imperfect simulation given by micromovements not belonging to the expressive category typical of the emotion that the sender desires to convey or certain time lags in the expression. When instead the content of what is being said is false, the deception is revealed by the filtering through of emotions set off by the very act of deceiving, and there is contradiction between meanings expressed in different modalities. For this reason, the task of the score is to establish whether there is a correspondence between the perceivable behavior of the speaker and his thinking.

Through further analysis carried out using the “facial action coding system” (FACS) it was possible to classify muscle movements of the face. This classification was developed by a Swedish anatomist, Carl-Herman Hjortsjo, and then recodified by Paul Ekman and Wallace V. Friesen in 1978, with updates made by Ekman and Friesen together with Joseph C. Hager in 2002. Using this technique of analysis of facial microexpressions it is possible to identify the inner emotional state of a person, obtaining indications on the subject’s hidden thoughts and feelings.

The technique attributes a combination of corresponding codes to certain facial micromovements [called action units (AUs)] made by the person. The combination of these movements may lead to a further decodification or “translation” of the code into a predominantly emotional and generally unconscious meaning.

17.3 27 July 2004: Barack Obama Speech at the Democratic National Convention

On 27 July 2004 in the campaign of the presidential election that would lead to George W. Bush’s election for a second mandate, the unknown Illinois politician Barack Obama makes a programmatic speech at the Democratic Convention in Boston. It should have been one of the many preparatory speeches in the lead-up to the main speech by John Kerry, but it became the Obama revelation at the national level.

The autobiographical recall happens at the very beginning of the speech (min. 1.34), in what could be seen as the social part of the self-presentation of the speaker to his audience. It starts with the description of his father's life:

"My father was a foreign student, born and raised in a small village in Kenya. He grew up herding goats, went to school in a tin-roof shack."

While the audience, gathered in a big hall and waiting for the more important speakers to come, move around the hall and chat, Obama *turns his gaze* all around, stressing his will to reach all the people with his words. *He smiles* expressing affection and nostalgia for the memory of his father. He describes the poor school he attended using *iconic gestures* that stress how this building was little more than a hut. He speaks rather quickly, but *the tone of his voice* shows his will to describe clearly and sincerely the poor situation of his family. His voice, in fact, has a peak of intensity on the words "herding goats" as if to signal "this was the story of my father; I do not want to hide his humble origins."

The description of the family goes on, turning to his grandfather (min. 1.45):

"His father – my grandfather – was a cook, a domestic servant to the British. But my grandfather had larger dreams for his son."

Here again Obama *slowly turns his gaze* around and then looks firmly at the audience in front of him. The voice is clear, while *batonlike gestures* stress the meaning of the words. When communicating clearly his grandfather's condition as domestic servant, Obama's *facial expression* shows his distress; however, it is quickly brought under control. His entire body seems to communicate his pride for the sufferings linked to the colonial past of his family: a proud attitude expressing his dignity and gratitude for his ancestors, which in this case may be clearly distinguished from an arrogance due to his prominent social position (Poggi and D'Errico 2012). He gazes fixedly into the camera and turns slowly from one side of the audience to another, as if to ask to all his listeners to pay attention to his words. He describes also with iconic gestures the tin roof of the poor school shack attended by his father. All of his body language seems to add to his words, to make as clear as possible the poor condition of his father's life.

Having so clearly stated the social condition of pariah of the family into which he was born, Obama goes on to describe the life projects of his parents (min. 1.55):

"Through hard work and perseverance my father got a scholarship to study in a magical place, America, that shone as a beacon of freedom and opportunity to so many who had come before."

In this passage, Obama stares right in front of him, *looking steadily* into the camera, as if expressing his will to speak not only to the audience gathered in the hall but also to those watching him on television. His batonlike gestures are constant, especially when he describes America as a "magic place." He pauses briefly before these words, as if to emphasize them. After stating clearly his pariah condition as a descendant of colonial servants and declaring his love for America, Obama introduces the most difficult part of his autobiography, i.e., his birth as a son of a racially mixed couple. As was clearly shown by studies on racial prejudice, in fact,

being an offspring of a mixed couple is the worst social position to have since such children may be seen as a living defeat for the myth of racial purity (Volpato et al. 2010). Obama chooses to introduce this topic by talking about the “strange” name he was given, intertwining it with the faith expressed by his parents on American democracy (min. 3.02):

“My parents shared not only an improbable love, they shared an abiding faith in the possibilities of this nation. They would give me an African name, Barack, or ‘blessed’ (3.03) believing that in a tolerant America your name is no barrier to success.” (min. 3.11)
 “They imagined – they imagined me going to the best schools in the land, even though they weren’t rich, because in a generous America you don’t have to be rich to achieve your potential.”

Score 1: Boston (Table 17.1).

While describing his parents’ dream, Obama changes his former attitude of pride, linked to his self-description as a social pariah, into a warmer attitude toward his audience. He turns his body and leans toward the audience, which is now listening to him attentively, as if to indicate a desire to reduce the social distance between them and involve his audience in a confidential communication. When referring to his “strange” name, he draws attention to himself by putting his open hand on his chest. He looks sternly at the audience but also at the camera, as if asking his entire audience, both in front of him and watching him on television, to share with him such an intimate description of his parents’ dreams. After declaring his condition of social pariah and sharing with his audience the story of his parents’ dreams for him, in the last sentence of this fragment, which opened the speech that led, for the first time, to his being considered a plausible contender in the US presidential race, Obama resumes his self-presentation as the offspring of a family socially outcast by linking his autobiographical roots to the larger American story, expressing at the same time his gratitude to his ancestors and to his country (min. 3.47):

“I stand here knowing that my story is part of the larger American story, that I owe a debt to all of those who came before me, and that, in no other country on earth is my story even possible.”


Here again he moves his body toward the audience, as if trying to represent his connection to them, while looking to the camera too, as expressing his will to speak not only to those in the hall but to all Americans watching the political campaign on television. The efficacy of his rhetorical skills is immediately confirmed by the audience, who rise to their feet and applaud him loudly. At this feedback Obama reacts by stopping in embarrassment (min. 4.09), as if signaling that this applause was uninvited by him (Bull 2006), regaining by this humble attitude the role of second-rank politician he was expected to fill at that time, as if to deny any suspicion of personal ambition in this self-presentation to the people voting for his political party. Also, his emotions, clearly shown during the autobiographical recall and expressing his anger when remembering the status of his grandfather as a domestic servant in a British family in a colonial setting (min. 1.45–1.54), are kept under control in order to present an attitude of never going “over the top” of a well-managed political speech. This regulation of his emotions (Frijda 2013) silently

Table 17.1 Score 1. Boston

v.Ds.	My parents shared not only an improbable love, they shared an abiding faith in the possibilities of this nation	They would give me an African name, Barack, or blessed, believing that in a tolerant America your name is not a barrier to success	They imagined me going to the best school in the land, even though they weren't rich	Because in a generous America you don't have to be rich to achieve your potential
p.i. Ds.	Descending	Ascending – peak of intensity – peak of tone – peak of duration at “African” – Barack – tolerant America	Ascending	Descending
S	Tone becomes confidential	Emphasis, intensity	The line of thinking still is not finished	There is nothing else to add
II	It is a revealing, intimate story	These keywords are important	Of this I'm firmly convinced	
TS	I IMM	IMM	IMM	IMM
II	IMM	IMM	IMM	
F	I Additional	Repetitive	Additional	Additional
II	Additional	Additional	Additional	
G Ds	He raises first his right hand with index finger and thumb together and then both hands parallel in front of him	He raises his right hand and beats it several times on his chest at the top of his heart	Raises right arm and then moves it forward with palm facing outward	Right arm is again raised and then moved forward with palm facing outward
Ts	Batonlike and iconic	Iconic and deictic	Iconic	Batonlike and iconic
S	I He underlines the concept	He indicates himself	He underlines their distance from wealth	He illustrates and underlines the uselessness of wealth
II	He seems to indicate the greatness of their faith in the possibilities of that nation	He also indicates the bonds of affection with his African origins		
TS	I IMM	IIM	Imo	Imo
II	Imo	IIM		

F	I	Repetitive	Repetitive	Repetitive	Repetitive
	II	Additional	Additional		
F Ds		His gaze is on the audience straight in front of him	His gaze ranges warmly over the audience	His gaze is again focused in front of him	His gaze is still focused in front of him, then moves right
S	I	I am speaking to you	He tries to establish a connection with the audience	He is speaking to everyone	He is speaking to an ideal audience
	II	I am sure of what I am saying	He seems to say: I am proud of my name		He wants to affirm a universal concept
TS	I	IMM	IMM	IMM	IMM
	II	IMM	IIM		Imo
F	I	Additional	Additional	Additional	Additional
	II	Additional	Additional		Additional
C Ds		His body bends and he leans forward	His upper torso turns to take in the whole auditorium		He shakes his head
	I	He tries to establish closer contact with the audience	He seeks the consensus of the whole auditorium on his assertion and he obtains it		He underlines the nonnecessity of wealth
TS	II	He puts himself in confidential terms as if to reveal something			
	I	IMM	IMM		Imo
F	II	IMM			
	I	Additional	Additional		Repetitive
	II	Additional			

Table 17.2 FACS analysis 1

1		
But the West is not responsible for the destruction of the Zimbabwean economy over the last decade or wars in which children are enlisted as combatants		
<i>Coding AUs</i>	<i>Decoding</i>	
1 + 2	Doubt + certainty + signal illustrator	

conveys to the audience that the risky choice of speaking frankly about his family's poor and humble circumstances is due to the will to communicate openly his trust in democracy and in spite of the personal discomfort that these memories might cause him to this day. His preference for the political outcome of his speech over the more natural inclination to escape from that discomfort in such a situation of self-exposure is therefore striking evidence of his emotional resources and of the energy behind his will to advance the political discourse to face the crucial point of democracy, i.e., the equal opportunities offered to all, regardless of the starting conditions of their lives.

FACS analysis 1 “But the West is not responsible . . .” (Table 17.2).

17.4 5 April 2009 Obama Prague Speech on Nuclear Weapons

This speech was given in Prague on 5 April 2009, on the occasion of the US-EU summit, to an audience of 30,000. Obama declared that the United States would undertake concrete actions for a world without nuclear weapons. In this speech he used autobiographical memories to equate the realization of his dreams with the realization of the people of the Czech Republic. The message of hope includes everyone without distinction.

Here again Obama introduces his own autobiographical memories in the first part of his speech, aimed at fulfilling the social role of self-presentation to his audience. He starts out by stressing his positive attitude toward the people hosting him and his family, remembering the Czech friends he had during his youth in Chicago (min. 2.27):

“I’ve learned over many years to appreciate the good company and the good humor of the Czech people in my hometown of Chicago.”

He looks at his audience, gathered in the main square of the old town of Prague, *smiling* to them, thereby creating a climate of intimacy despite the *distance* between

him and the large crowd listening to his speech. As he pronounces these words, his gaze relaxes into a smile, expressing in an intimate way the pleasure he gets from being there. After a few minutes, however, he changes this good-humored beginning by recalling to his audience the difficult historical circumstances in which he was born, which influenced in different yet equally serious ways both his personal story and the history of the Czech people (min. 4.09):

When I was born, the world was divided, and our nations were faced with very different circumstances.

Both the serious and concerned *gaze* he shows the audience and his *body movements* – *he moves brusquely to the side and forward, while shifting his weight alternately to either leg* – express the emotional distress he is feeling when recollecting such a dramatic past. To better stress the violence of the past historical situation and how helpless it seemed to be – a helplessness sadly witnessed by the wall in Berlin – Obama goes on to frankly describe how “realistic” politicians could have foreseen both the future of a social outcast such as himself at the time and the future of a nation situated behind an Iron Curtain that was segregating into two parts the peoples of Europe (min. 4.20):

Few people would have predicted that someone like me would one day become the president of the United States. (Applause.) 4.34 Few people would have predicted that an American president would one day be permitted to speak to an audience like this in Prague.

Score 2. Prague (Table 17.3).

In this part of his speech, Obama’s facial expression changes, becoming sad, but he is also regulating his emotion, as if accepting the reality of the old historical situation, which wrought havoc on lives with the same violence that tore apart the socially segregated American society and the societies of the Soviet bloc countries, which were forcibly removed from their natural ties and connections with their European neighbors. Also, Obama’s body movements clearly express his distress, evidenced by how he shifts frequently from one leg to the other. However, in this same sentence he expresses not only the sadness of the situation but also the comparison of these apparently “realistically” poor chances assigned both to “someone like me” as a political leader and to the Czech people as an important nation of the European community. By evoking his position as pariah in the American society of the times, Obama points proudly to himself by patting his own chest (Tracy and Robins 2007), as if to express with his entire body the self-confidence that helped him to react defiantly to social prejudice. The audience stresses this proud sentence with a big round of applause, which Obama, now president of the USA and no longer a “strange” incumbent among powerful competitors, accepts and pauses in gratitude, looking straight back at the audience. In the second sentence, he assumes the same parrhesiastic attitude he used regarding his personal unfortunate circumstances to recall to his audience the fact that they were pariahs, too, when Europe was still divided by the Berlin Wall.

Table 17.3 Score 2. Prague

v.Ds.		Few people would have predicted that someone like me would one day become the president of the United States	Few people would have predicted that an American President would one day be permitted to speak to an audience like this in Prague
p.i. Ds.		Ascending, emphatic peak in duration at <i>become</i>	Ascending then descending
S	I	The subject is not yet closed	The subject is closed; emphasis on the word <i>people</i>
	II	The world is changing and my presence here is proof of that. I'm convinced of it.	The world is changing and my presence here is proof of that. I'm convinced of it.
TS	I	IMM	IMM
	II	IMM	IMM
F	I	Additional	Additional
	II	Additional	Additional
G Ds		Moves right hand with fingers closed up and down repeatedly and brings it to his chest at the words <i>like me</i> , then raises his index finger at the word <i>become</i> and then brings his hand back to the lectern	Raises his right index finger and then brings together thumb and index finger and alternates these movements, moving the hand up and down several times
Ts		Batonlike and deictic	Batonlike – assertive, didactic
S	I	I want to make what I'm saying clear and I'm involved in the themes of the speech	I want to make this concept clear
	II	It's important and I'm talking about myself	Most people were wrong
TS	I	IIM	IMM
	II	IIM	Imo
F	I	Additional	Additional
	II	Repetitive	Additional
F Ds		His gaze is serious and grim	His gaze is serious and grim and directed at the entire audience
S	I	I'm aware	I'm here and you're here
	II	The speech concerns me and I'm serious in giving it	I'm speaking to the entire audience to tell a truth
TS	I	IMM	IMM
	II	IMM	IMM
F	I	Additional	Additional
	II	Additional	Repetitive – additional

(continued)

Here again, the complementary analysis of the FACS (AUs 4, 7, 14, 23) shows how Obama clearly expresses his anger when recalling Europe's division and immediately regulates this negative emotion (actively suppressing the 14 AUs).

Table 17.3 (continued)

C Ds		His body moves slightly forward to the left. He is immobile and erect at the words <i>President of the United States</i>	His body is leaning forward and moves from right to left
S	I	I'm like you	I'm speaking to you
	II	But I've come this far	Those people were wrong because I'm here in front of you as President of the USA
TS	I	IIM	IMM
	II	IIM	IMM
F	I	Additional	Additional
	II	Additional	Additional

17.5 7 July 2009. US President Barack Obama's Speech to the New Economic School in Moscow

In a speech to the students of the New Economic School of Moscow on 7 July 2009, Obama claims "*America wants a strong, peaceful and prosperous Russia.*"

The visit of July 2009 had the aim of supporting relations between the US and Russia, also aiming for a drastic reduction in nuclear arsenals.

In this speech, the autobiographical recall is unusually set in the middle of the discourse and not in the first, social, part. In this speech, in fact, the recall of the social situation of a pariah is not used by Obama as a crucial aspect of his self-presentation as the leader of the USA but as a reminder of the political shortcomings that violently divided the people living both in the democratic USA and in the former Soviet Union. Obama is trying by this rhetorical move to stress, for his young and cultivated audience, the enormous historical progress made by both their societies in recent years. To achieve this goal, he links together in two subsequent sentences two descriptions of the preceding historical circumstance that obtained in both the USA and Russia before the birth of his young audience: his situation as the son of a mixed-race couple in a segregated America and the isolated situation of the schools in the former Soviet Union (min. 28.16):

"When I was born, segregation was still the law and lands in part of America and my father's Kenya was still a colony. When you were born, a school like this would have been impossible, and the Internet was known to only a privileged few."

In the first sentence, a multimodal analysis of Obama's communication clearly reveals his psychological difficulty in speaking frankly about the segregated state of the USA, where his family was still bearing the burden of past colonial domination. Among other signals of stress, it is worth noting how Obama repeatedly *blinks* when pronouncing the word "segregation" (referring to America) and "colony" (referring to the country of his father). However, at the end of these sentences he *nods* as if to acknowledge (Poggi et al. 2010) that this parrhesiastic account of his personal and social past is stressful, yet it is the right rhetorical move.

After describing his life in a parrhesiastic way, Obama invites his audience to join him in this risky communicative game. He is aware of the danger of proposing such a realistic description of the cultural isolation characterizing the social past of Soviet schools. However, he points overtly to his young audience, so as to involve them more directly in his speech, saying (min. 28.29):

“When you were born, a school like this would have been impossible, and the Internet was only known to a privileged few.”

Score 3. Moscow (Table 17.4).

He looks at them seriously with a frowning gaze, without hiding his distress and smiling only slightly, at the end of the sentence. With his *facial expressions* of frowning and keeping his *body from moving*, he acts as if he is urging all his listeners to consider very seriously and in depth the enormous historical change in which they are all involved and that destroyed the violent conditions haunting their societies at the time of their birth to open up opportunities that would have been considered pipe dreams only a few years before.

17.6 President Obama’s Speech to the Parliament of Ghana, Saturday, 11 July 2009

This speech was made in the Parliament building in Accra, the capital of Ghana, on 11 July 2009. It was Obama’s first trip as President of the United States to sub-Saharan Africa and was made after the G8 summit in Italy. During the speech, Obama made use of autobiographical themes to assert, among other things, that it was no longer a time to play the role of the victim and that the future of Africa depended on the Africans themselves.

Also in this example, Obama uses his autobiographical memories as an integral part of his self-presentation, just in the social part beginning the speech. But of course this reminder of his pariah origins as the offspring of a colonial family acquires a new meaning when pronounced before an African audience as the first African-American president of the USA.

After thanking his hosts for their warm welcome to him and his family, Obama introduces himself to his audience declaring that (min. 3.53)

“After all I have the blood of Africa within me (muted applause, and the speaker goes on), and my family’s own story encompasses both the tragedies and triumphs of the larger African story.


In this first sentence, Obama maintains a very dignified posture: his chin raised, his head and bust slightly outstretched, he moves his gaze around the entire audience, where a large number of African leaders, sometimes dressed in traditional furs, are sitting. He stops the applause that starts following his declaration that he is partly African, while his gaze remains very serious and he swallows during the beginning of the uninvited applause (Bull 2006) as if stressing that he is stating a fact

Table 17.4 Score 3, Moscow

v.Ds.	When you were born	Pause	A school like this would not have been impossible	And the Internet was known to only a privileged few
p.i. Ds.	Ascending, peak of intensity at <i>you</i>		Ascending. Peak of intonation at <i>impossible</i>	Peak of duration at <i>only</i> and <i>known to</i> – ascending
S	The subject is not over; he emphasizes <i>you</i>	Reflection	The subject is not over	The subject is not over
	When you were born too things were different	I'm about to say something very important	This seems like a time impossible to conceive of	It's a glaring example of how much things can change
TS	I IMM	IMM	IMM	IMM
	II Imo	IMM	Imo	Imo
F	I Additional	Additional	Additional	Additional
	II Repetitive	Additional	Additional	Additional
G Ds	He raises his left hand half-closed with index finger turned toward the audience	His hand returns to the lectern	His hands are raised circumscribing a space with the palms one in front of the other, then they are lowered imperceptibly and the palms are turned up, fingers spread. They both jerk at the word <i>impossible</i>	He raises his right hand with palm open and fingers outward and upward, moving it in the air
Ts	Deictic		Iconic – batonlike	Batonlike – iconic
S	I I'm speaking about <i>you</i>		I'm drawing the school in the air and it was impossible to conceive of it once	I'm representing the limited number of people who could use the Internet in the past
	II After speaking about myself I'm speaking about you		I'm certain that the information I'm giving you is true	It's a certainty
TS	I Imo		Imo	Imo
	II IMM		IMM	IMM
F	I Repetitive		Repetitive – additional	Additional
	II Additional		Repetitive	Repetitive

(continued)

Table 17.4 (continued)

F Ds	His gaze is grim, his eyes are narrowed and turned to the audience	His gaze is serious and grim	His grim expression relaxes and seems to leave space for resignation	Expression more relaxed, very slight smile
S	I'm speaking to <i>you</i>	I'm about to say something very serious	This was how things were	Even the Internet was the privilege of a few; to you it's an everyday tool
II	I'm giving you an important piece of information	You must listen to me carefully	We can't change the past	I'm speaking of a recent time and yet it was so different
TS	I	IMM	Imo	Imo
	II	IMM	IMM	IMM
F	I	Additional	Additional	Additional
	II	Additional	Additional	Additional
C Ds	Body immobile, directed forward	His body remains in the same position as before	His body stays in the same position as before	After the word <i>Internet</i> his body moves right and he shifts his weight to his left foot
S	I'm resolute and sure of what I'm asserting	I'm serious and sure of what I'm about to say	I'm serious, sure, and resolute with regard to what I'm saying	It's something that concerns all of you
II				Let's all be aware of this
TS	I	IMM	IMM	Imo
	II			IMM
F	I	Repetitive	Repetitive	Repetitive
	II			Additional
				

and not fishing for consent. He goes on without allowing the applause to continue, stating frankly the humiliating station of his grandfather, a domestic servant of no importance (“boy”) in a British family (min. 4.10):

“My grandfather was a cook for the British in Kenya, and though he was a respected elder in his village, his employers called him “boy” for much of his life.”

He maintains the same attitude of dignified pride (Poggi and D’Errico 2012) – chin raised, gaze moving around slowly, chest puffed out slightly toward his audience all the while – as if to suggest that he is keenly aware of this humiliation of his grandfather, but that he also accepts it as a historical fact that cannot be changed. He wants, however, to denounce all the injustice suffered by his elder and continues (min. 4.22):

“He was on the periphery of Kenya’s liberation struggles, but he was still imprisoned briefly during repressive times.”

He slowly vacillates as he goes on discussing all aspects of his family story, hiding nothing, not even the most distressing aspects. Only at the end of his recollection about his grandfather’s brief imprisonment, though an old person and certainly not a dangerous one, does Obama resume his proud posture (Poggi and D’Errico 2012; Tracy and Robins 2007), raising up his head and chin. After this brief narrative about his grandfather’s imprisonment, he closes his mouth firmly, stops vacillating slightly, and closes his arms in front of him, as if going back to a resting pose after trying to keep under control the distress of recalling the useless and undeserved violence perpetrated against his family. The evident effort demanded of him to communicate these autobiographical events despite showing in public his personal discomfort accounts for his self-determination (Frijda 2013) to present to his audience a parrhesiastic narrative of African history, not a glorified one (Foucault 2001). In fact, the more he recalls his family story, the more he tries to highlight the bonds inextricably intertwining his personal feelings with the larger African story (min. 4.29):

“In his life, colonialism wasn’t simply the creation of unnatural borders or unfair terms of trade; it was something experienced personally, day after day, year after year.”

The way he looks at his audience, sternly and without smiling, the movements of his head, nodding in acknowledgment of the historical truth of what he is saying (Poggi et al. 2010), the way in which his head keeps up at the end of each sentence, stressing his proud frontal pose (Poggi and D’Errico 2012; Tracy and Robins 2007) toward his audience with his chest, which now remains almost still as he recalls not only the humiliation of his family but of all colonized peoples, the frequent batonlike gestures, the rhythm of his voice, which shows a peak of intensity on the word “personally,” the brief silent pause following each important concept presented to the audience – all these signals taken together seem to underscore how focused Obama is on what he is saying and how he is asking his audience to stay focused, too, and simply listen to him without unnecessary feedback (Bull 2006). The audience shows a deep acceptance of his communicative proposal, to frankly review in his official speech the difficulties they share in their common colonial past. Many

leaders nod or look at the floor, expressing the same focused attitude shown by the speaker. After giving a frank, parrhesiastic account of colonial violence, exemplified by his family memories, Obama turns now to recall the historical change that ended this structural social imbalance, to change Africa's destiny starting with the history of the country that is now hosting him as a democratic nation and an important political partner (min. 4:42):

"My father grew up herding goats in a tiny village, an impossible distance away from the American universities where he would come to get an education. He came of age at an extraordinary moment of promise for Africa. The struggles of his own father's generation were giving birth to new nations, beginning right here in Ghana."

During this part of the autobiographical recall, Obama keeps signaling the need for his audience to pay close attention to his words. His voice reaches a peak of intensity on the words "impossible," "new nations," and "here": keywords to stress the pariah status of his father, very far from everything, and the historic achievement of democracy in the very place where he is now giving his speech. He supplements his sentence with many gestures: iconic, representing the small village of his father and the huge distance separating it from the rest of the world; batonlike, stressing the importance of what is said; and deictic, pointing both to himself and to the place where they are now all gathered. He moves his eyes around continuously, looking at his audience, except when he is imagining his father's distant village, when he almost closes his eyes (Vincze and Poggi 2011), as if inviting his audience to imagine how it was to live at that time and in that place. Also, his body vacillates slightly as if joining in his effort to link together autobiographical memories and African history: moreover, he jumps slightly in jubilation when he recalls how the first new nation of the free Africa was precisely in Ghana. In this last part of the sentence, his gestures alternately point to himself and to the place where the speech is being held, connecting his pride for his family's memories with his pride for the democracy of the new nations today living in an independent Africa. Nevertheless, this sharing of positive emotions lasts only a little while: since Obama, after openly describing the difficulties suffered by his family, seriously invites his audience to play the risky but salutary *parrhesiastic* game as well. Therefore, he openly declares that (min. 5:53):

"In many places, the hope of my father's generation gave way to cynicism, even despair. (6.08) It is easy to point fingers, and to pin the blame for these problems on others. Yes, a colonial map that made little sense bred conflict, and the West has often approached Africa as a patron, rather than a partner. But the West is not responsible for the destruction of the Zimbabwean economy over the last decade, or wars in which children are enlisted as combatants.

Score 4. Ghana (Table 17.5).

This is perhaps the most risky and important passage of this difficult speech. In a somehow unexpected way, Obama in fact uses here his commonalities with the audience, owing to their shared suffering in the colonial past, as a way to encourage his listeners to look as soberly at their condition as he did when recalling the humiliations his family was forced to undergo. He looks sternly and sometimes

Table 17.5 Score 4. Ghana

v.Ds.	In many places the hope of my father's generation gave way to cynicism, even despair	It is easy to point fingers and to pin the blame for these problems on others	But the West is not responsible for the destruction of the Zimbabwean economy over the last decade or wars in which children are enlisted as combatants
p.i. Ds.	Ascending then descending. Peak of intensity at <i>cynicism</i> and <i>despair</i>	Ascending and descending. Peak of duration at <i>blame</i>	Ascending and descending
S	I It's a sad story	I'm turning to and stating an important concept	I'm sure of what I'm saying
	II His personal story reflects the story of a people; you should be aware of this	It's necessary to understand who is responsible because it's easy to blame others	The responsibility is not the West's
TS	I Imo	IMM	IMM
	II Imo	Imo	Imo
F	I Additional	Additional	Additional
	II Additional	Additional	Additional
G Ds	His hands are open and turned towards each other then, still open, are turned to Obama himself and moved back and forth, and at the word 'despair' only his left hand moves in the same way	He raises his right hand and moves it up and down in the direction of the audience to the right, then brings his thumb and index finger together, making the same gesture	He moves his left hand upwards slightly with palm open
Ts	Iconic—Batonlike	Symbolic (reprimanding gesture)	Symbolic
S	I I want to show you my father's mood	You must learn to take on responsibilities	I'm saying obvious but very important things
	II It's right that you should be aware of this	I want your full attention because you must know your history	This is your history
TS	I IMM	IMM	IMM
	II IMM	IMM	Imo
F	I Additional	Additional	Additional
	II Additional	Additional	Additional

(continued)

Table 17.5 (continued)

F Ds	His gaze is turned to one side of the audience and then the other, but his expression is calm. He lifts his chin and his expression is scornful	His gaze is turned to the audience, his expression is proud, and he lifts his head and chin at the end of the sentence	His head turns rightward with micromovements that express a “no,” then turns left and again right, with chin rising, then he raises both eyebrows and again lifts his head and chin leftward
S	I He is speaking to everyone present	I’m saying it to everyone	I’m saying something important and this is how things are – you absolutely must be aware of this
II	What I’m recounting is something sad and despicable but we need to accept it	I’m sure of what I’m saying and you must accept it	Contrary to what you might think, this is the truth
TS	I IMM	IMM	IMM
II	IMM	IMM Performative	IMM
F	I Additional	Repetitive	Additional
II	Additional	Additional	Additional
C Ds	His body stays still	He sways his body right and left	His body stays still and his head moves, rising rightward when he raises his eyebrows
S	I He expresses concentration	He’s agitated	It’s something important
II	The concept is important	It’s important to me to convey this concept to you	It’s a fact, whether you accept it or not, but you must take cognizance of it
TS	I IMM	IMM	IMM
II	Imo	IMM	IMM
F	I Additional	Additional	Additional
II	Additional	Additional	Additional
			

reproachfully at the audience of African leaders, overtly claiming that explaining the social problems currently plaguing their society by blaming past perpetrators is a way to escape from their own responsibilities. The audience listens quietly, staring down and sometimes nodding to the most reproachful parts of these severe but frank description of the many problems the African leaders must deal with if they really want to empower themselves after centuries of exploitation. This concept becomes even clearer in the last sentence that ends this first part of his speech to Ghana's leaders (min. 6.40):

In my father's life, it was partly tribalism and patronage in an independent Kenya that for a long stretch derailed his career, and we know that this kind of corruption is a daily fact of life for far too many.

Throughout this difficult parrhesiastic game with his audience, Obama often signals through facial expressions his involvement in his speech. He betrays his anger when remembering his grandfather's humiliations and imprisonment, anger that turns into disdain when he claims that "colonialism was something he (the grandfather) experienced personally, day after day, year after year." A similarly embodied communication is also evident when, expressing his enthusiasm in recalling that his ancestors' struggles gave "birth to new nations, beginning right here in Ghana," he rises up on his foot when pronouncing the words "new nations." But the enthusiasm that could easily be predicted considering this was the first political speech addressed to African leaders by a US president who "after all" had the blood of Africa within him is scarcely shown during this parrhesiastic and sometimes even reproachful discourse. The last part of this introduction, which certainly shocked his audience, ends up in fact by showing mostly negative emotions in Obama's facial expressions: disdain, quickly turning into sadness, when he states that "we also know that much of that promise has yet to be fulfilled"; doubt if saying these words overtly, when asserting that "the West is not responsible for the destruction of the Zimbabwean economy over the last decade, or wars in which children are enlisted as combatants"; sadness, coupled with a proud attitude, when claiming that

"In my father's life, it was partly tribalism and patronage in an independent Kenya that for a long stretch derailed his career, and we know that this kind of corruption is a daily fact of life for far too many."

17.7 Concluding Remarks

In this paper, we presented an in-depth analysis of several examples of autobiographical recalls used by Barack Obama during four relevant political speeches in order to explore the idea that he chose this kind of self-exposure to propose to his audience a difficult yet empowering "*parrhesiastic game*" (Foucault 2001). To frame this work, we relied on two main theories. The first, advanced by Hannah Arendt (Arendt 1978), proposes that an overt declaration of social disadvantages,

claimed by self-aware *pariahs*, connected to their birth enables them to contribute in an original and creative way to a political agenda since they are – more than conventional individuals – forced to recognize the importance of their natality, i.e., of the influence of their ancestors in their current social life. The second theory, proposed by Michel Foucault (2001), focuses on the risky choice of telling the truth that makes parrhesia a very specific kind of “communicative game.” Although this way of speaking may cause aggressive reactions from its recipients, since it reveals a truth that is hard to deal with, it may also be the most effective tool for enabling them to cope with their difficulties, which are presented much more as challenges than as defeats. If we consider not only the words pronounced by Barack Obama during the autobiographical recalls we observed but also the other modalities of his communication (Poggi 2007), as well as his regulation of his emotional expressions (Frijda 2013), made evident by an examination of his face, body, and voice, both Arendt’s and Foucault’s theories seem particularly suited to account for several features of these communicative acts. Much more than the general theory of personalization (McAllister 2007), in fact, these theories may explain why narrating humiliations and shortcomings of his ancestors and showing how at times it may be hard for him to regulate his negative emotional signals when reenacting these difficult memories, may ultimately become an opportunity to turn these events into a magnification of Obama’s public image and account for his self-determination to communicate frankly and fearlessly with his audience. This in turn allows him to go beyond the natural inclination to escape from the discomfort (Frijda 2013) he must publicly experience when playing his parrhesiastic game (Foucault 2001) on the basis of his autobiographical memories during political speeches (Leone 2013).

To conclude this first analysis of parrhesia in the political speeches of Barack Obama, inspired by these two theories, we would like to propose the idea that Obama’s leadership may benefit from his parrhesia. It empowers both him, through his “strange” autobiography, and, at the same time, audiences, who may need to cope with their own difficulties using this same attitude of fearless frankness. Our results suggest, in fact, that these autobiographical memories conveyed parrhesiastic narratives about the social origins of Obama as a pariah. Together with these risky rhetorical moves, emotional expressions (mainly negative) were clearly evident yet well regulated – his emotional expressions served to convey his sincerity, and his emotional regulation demonstrated his self-determination and self-maintenance (Frijda 2013).

Summing up, this first explorative analysis suggests that Obama sometimes assumes a parrhesiastic attitude – making clear to all his listeners his socially disadvantaged origins – in order to persuade his audience to accept a similar parrhesiastic game for communicating about difficult aspects of their political situation. Being an offspring of a despised social group, he may urge his audience to choose self-affirmative and empowering actions rather than indulging in irresponsibility. Communicating his pariah origins in a frank way, he may offer to his audience an occasion to observe how emotions may be regulated, controlling both the inclinations to escape from discomfort and to exaggerate it.

Of course, much work must be done to develop these first intuitions by applying to Obama's political speech both Arendt's (1978) and Foucault's (2001) theories.

The results of this study, together with observations already discussed in a previous paper (Leone 2013), suggest that the methodological choice to explore in depth the multimodal communication of a speaker (Poggi 2007), paying close attention to facial expressions (Ekman et al. 1984) and emotional regulation (Frijda 2013), may contribute useful insights into the issue. However, a sharper distinction between parrhesiastic and personalizing uses of autobiographical memories must be drawn, while continuing to analyze extracts from Obama's speeches with this same mixed methodology. Persuasive effects of these two communicative strategies must be explored as well, using chosen extracts of Obama's autobiographical memories as experimental stimuli.

Finally, we think that audience reactions must be better explored, starting with the particular feedback of applause (Bull 2006).

Taking into account all these limitations of our study, we hope nevertheless that it may advance our understanding of one of the many facets of the political speech of one of the finest political rhetoricians of our times.

References

- Arendt H (1978) *The Jew as Pariah: Jewish identity and politics in the modern age*. Edited and with an introduction by R.H. Felman, Grove Press, New York
- Arendt H (1997) *Rahel Varnhagen: the life of a Jewess*. Edited by L. Weissberg, Johns Hopkins University Press, Baltimore
- Bull P (2006) Invited and uninvited applause in political speech. *Br J Soc Psychol* 45(3):563–578
- Da Costa MK (2009) Interracial intimacies, Barack Obama, and the politics of multiracialism. *The Black Scholar* 39:4–12
- Ekman P, Friesen WV, Hager J (1978) *The Facial Action Coding System (FACS): a technique for the measurement of facial action*. Consulting Psychologists Press, Palo Alto
- Ekman P, Friesen WV, Bear J (1984) The international language of gestures. *Psychol Today* 18(5):64–69
- Foucault M (2001) *Discourse and the truth* (1983). In: Pearson J (ed) *Fearless speech*. Semiotext(e), Los Angeles, CA
- Frijda NH (2013) Emotion regulation and free will. In: Clark A, Kiverstein J, Vierkant T (eds) *Decomposing the will*. Published to Oxford Scholarship Online, Sept 2013
- Goffman E (1967) On face-work. *Interaction Ritual*:5–45
- Hammack PL (2010) The political psychology of personal narrative: the case of Barack Obama. *ASAP* 10:182–206
- Leone G (2013) Sometimes *I*, sometimes *Me*: a study on the use of autobiographical memories in two political speeches by Barack Obama. In: Poggi I, D'Errico F, Vincze L, Vinciarelli A (eds) *Multimodal communication in political speech. Shaping minds and social action*. Springer-Verlag, Berlin
- McAllister I (2007) The personalization of politics. In: Dalton RJ, Klingemann H-D (eds) *Oxford handbook of political behavior*. Oxford University Press, Oxford, pp 571–585
- Poggi I (2007) *Mind, hands, face and body. A goal and belief view of multimodal communication*. Weidler, Berlin

- Poggi I, D'Errico F (2012) Pride and its expression in political debates. In: Paglieri F, Tummolini L, Falcone R, Miceli M (eds) *The goals of cognition: Festschrift for Cristiano Castelfranchi*. London College Publications, London, pp 221–253
- Poggi I, D'Errico F, Vincze L (2010) Types of nods. The polysemy of a social signal. In: Calzolari N et al (eds) *Proceedings of the seventh conference on international language resources and evaluation (LREC'10)*. Malta 11–12 May (ERLA)
- Tracy JL, Robins RW (2007) The psychological structure of pride: a tale of two facets. *J Pers Soc Psychol* 92:506–525
- Vincze L, Poggi I (2011) Close your eyes and communicate. In: Paggio P, Ahlsén E, Allwood J, Jokinen K, Navarretta C (eds), *Proceedings of the 3rd Nordic symposium on multimodal communication*. NEALT proceedings series. 15(2011): 62–71. Electronically published at Tartu University Library (Estonia) <http://hdl.handle.net/10062/22532>
- Volpato C, Durante F, Gabbiadini A, Andrighetto L, Mari S (2010) Picturing the other: targets of delegitimization across time. *IJCV* 4(2):269–287
- Zogby J (2009) Barack Obama: America's first Global President. *Politics* 45

Part V
Technologies for Conflict Detection
and Simulation

Chapter 18

Detecting Speech Interruptions for Automatic Conflict Detection

Marie-José Caraty and Claude Montacié

18.1 Introduction

Research in organization and management has investigated phenomena, such as the causes, effects, and handling of interpersonal or intergroup conflicts (Korabik et al. 1993; Macintosh and Stevens 2008; Thomas et al. 2008). Data on these social and psychological phenomena are collected from people who are involved in the conflict, witnesses of the conflict, or, by extension, looking at a recording of the conflict escalation between the protagonists. A large quantity of audio and/or video metadata can be extracted from these recordings, such as the conversation, face, and gesture interactions. In this chapter, the conversational interactions during political debates have been studied to develop an automatic conflict detector from voice analysis. A reliable detector of conflict would be useful for many applications, such as security in public places, the quality of customer services, and the deployment of intelligent agents. The development of such a system requires modeling of the conversational interactions as well as the search for specific interactions in relation to a given measure of conflict handling (Rahim 1983; Daly et al. 2010).

M.-J. Caraty (✉)

Paris Descartes University, 45 rue des Saints Pères, 75006 Paris, France

STIH, Paris-Sorbonne University, 28 rue Serpente, 75006 Paris, France

e-mail: Marie-Jose.Caraty@ParisDescartes.fr

C. Montacié

STIH, Paris-Sorbonne University, 28 rue Serpente, 75006 Paris, France

18.1.1 Model of Conversational Interaction

Conversation is a social interaction between two or more people, where taking turns to talk is naturally observed. In the pioneering work of Sacks et al. (1974), an organizational model of turn-taking for conversation that is context-free, capable of context sensitivity, and having a cross-cultural validity was investigated. The constraints of their model were set in reference to the high cross-cultural flexibility of conversation accommodation, with a wide range of interaction in which there is a variety of persons and numbers of persons who are taking part. The authors proposed a model that relies on two components that are related to the turn-constructive units (TCUs, the basic units of talk) and the turn allocation at the end of each TCU for the next unit (the next speaker's TCU). TCUs end with points of possible completion (e.g., gap, query) called transition-relevant places (TRPs), in which the turn transition could be relevant but is not necessary. Observed in any conversation, 14 facts were listed. An excerpt of this list is the following: (a) mostly one party talks at a time; (b) the vast majority of turn-taking transitions is composed of transitions that have no/slight gap and no/slight overlap; (c) the turn size varies; (d) overlapping speech is common, but brief; (e) two basic turn-allocation techniques are used: the "current selects next" technique when a current speaker can select a new speaker (e.g., addressing a question) and the "self-select" technique when a speaker can self-select in starting to talk; (f) repair mechanisms exist for addressing turn-taking violation; e.g., when overlapping speech occurs, one (or more) of the speakers will stop prematurely. A set of rules was edited for addressing turn transitions from TRP in such a way as to minimize the gap or overlap in the transitions. The turn transfer is defined according to the construction of the TCU, regardless of whether the "current speaker selects next" technique is used as well as the eventual application of "self-selection." The rules are based on the purpose of no-gap-no-overlap transitions, for which ability is required in anticipating the precise moment at which a TCU is going to come to a completion point (i.e., a TRP). In related work (De Ruiter et al. 2006), the lexical and syntactic content of TCU was shown to be necessary for this anticipation, while the intonation contour was neither necessary nor sufficient for this projection. According to the turn-taking rule-set applied to a multiparty conversation, overlap is expected in the neighboring transition-relevant places: when a possible completion of the current TCU is wrongly projected by a party or when parties are competing in a self-selection mode for a next turn. In a work that is related to turn-taking organization and that is beyond the ordinary conversation and is mostly unconstrained in terms of a role, a wide range of publications have studied the turn-taking practices and characteristics within various contexts of multiparty interactions. Distinctive features of turn-taking were found in institutional interactions in which a turn-taking organization is more constrained and specialized according to the roles that are assigned to the group members (e.g., interviewer vs. interviewee, chair vs. participant). Studies on turn-taking management were investigated in institutional settings such as in a classroom (Mac Houl 1978; Mehan 1985; Lerner 1995),

in courts (Atkinson and Drew 1979), in political interviews (Beattie 1982), in press conferences (Schegloff 1987), in mediation (Garcia 1991), in professional meetings (Boden 1994), in talk shows in which interpersonal conflicts are expressed (Brinson and Winn 1997), in auctions (Heath and Luff 2007), in political debates (Valente and Vinciarelli 2010), and in political meetings that involve large groups of people in which everyone can contribute ideas, opinions, and proposals and in which opposition is also expressed (Mondada 2013). The role of the chair has been analyzed in various studies (Boden 1994; Svennevig 2008; Mondada 2012). Prediction of the speaker order in turn-taking was investigated in news, talk shows, and meetings (Barzilay et al. 2000; Vinciarelli 2009).

18.1.2 Guidelines and Overview

In related work on conflict detection in conversational interactions (Valente and Vinciarelli 2010; Pesarin et al. 2012), turn-taking patterns and overlaps between speakers are shown to be informative with respect to classification into the presence or absence of conflict. The total amount of overlap and the minimum pitch during overlap were found to be the features that correlated the most with conflict (Kim et al. 2012c). A widely adopted classification of interruptions/overlaps is collaborative or competitive in reference to the “cooperative-competitive” dimension of the conflict-handling style. While communication strategies are naturally collaborative, this preponderance is not the case for conflict dialogues, in which competitive strategies are the norm. The detection of competitive interruption is a difficult problem in relation to the search of the TRPs. Spectral content and intonation contour are not sufficient to locate these places. Furthermore, the perception of the conflict can be different in the case of the constrained organization of turn-taking, such as institutional interactions (interview, debate, meeting). Competitive strategies such as those of the moderator or the chairman appear to be natural in this context and are not perceived as conflicting. Our experiments relate to the classification of audio clips into two classes of conflict level (low and high) during the Interspeech 2013 Conflict Challenge. The clips, which were extracted from political debates, have been annotated into conflict levels, using crowdsourcing to model the perception of the people. For our design of the conflict detector, we categorized the overlapping speech into low- and high-level conflict overlap. We made the assumption that these categories can be detected from acoustic cues. We focus our study on a multi-resolution framework for the detection of the overlaps and a multi-expert architecture to include knowledge about overlap in the automatic conflict detector.

This chapter is organized as follows: Section 18.2 presents the speech material that we used for the experiments on conflict detection; it describes and analyzes the statistical characteristics of the corpus while focusing on interruptions and the moderator’s role. Section 18.3 describes the Conflict Challenge and the various audio feature sets that were used for our investigations. In Sect. 18.4, the

multi-resolution framework of the overlap detectors is outlined, the relation of the types of overlap with the conflict level is introduced and assessed on the Development set, and the results are discussed according to the official measure of the challenge in terms of the UAR. Section 18.5 describes the multi-expert architecture of the conflict detector. Various audio features that are related to the overlap detectors are presented. The results on the conflict detector task on the Test set are discussed. Section 18.6 presents the study's conclusions.

18.2 Speech Material

The SSPNet corpus (Kim et al. 2012a) is an international reference for social signal databases. In the context of political debates, this corpus allows investigations on conflict to occur during interactions between group members. SSPNet was used for our study in analyzing various turn-taking characteristics and testing models for conflict level detection.

18.2.1 SSPNet Corpus

The “SSPNet Conflict Corpus” is a collection of 45 political debates in the French language that were televised in Switzerland. It represents approximately 12 h of speech signals; 1,430 audio clips of 30 s duration were extracted from the corpus. A total of 157 individuals were speaking in the collection of debates (23 females and 134 males). In the various multiparty discussions of the debates, the roles of the group members were distinguished: a member of the group held the role of moderator, and the other members were participants who were taking part in the debate. Four moderators (1 female, 3 males) and 153 participants (22 females, 131 males) were counted in the database. The SSPNET corpus was distributed for the Interspeech 2013 ComParE Challenge. Data were split into the Train, Development, and Test sets: 793 clips were in the Train set, 240 clips were in the Development set, and 397 were in the Test set. Metadata are available for the Train and Development sets.

The clips were annotated in terms of the conflict score in the range -10 to $+10$ by crowdsourcing, to model the perceptions of the data consumers at a nonverbal level; metadata were taken to be low-level conflict (LLC) when the score was lower than 0; otherwise, it was taken to be high-level conflict (HLC). Figure 18.1 shows the distribution of the clips of the Train set as a function of the conflict score range (CSR). The clips are split into the two classes of level conflict (LLC and HLC); the dashed line shows the boundary between the LLC and HLC clips. LLC clips are predominantly represented in the database (63 % for LLC vs. 37 % for HLC).

Segmentation metadata are available for each clip, indicating the diarization (“who spoke when”). From these metadata, we can compute the following statistics:

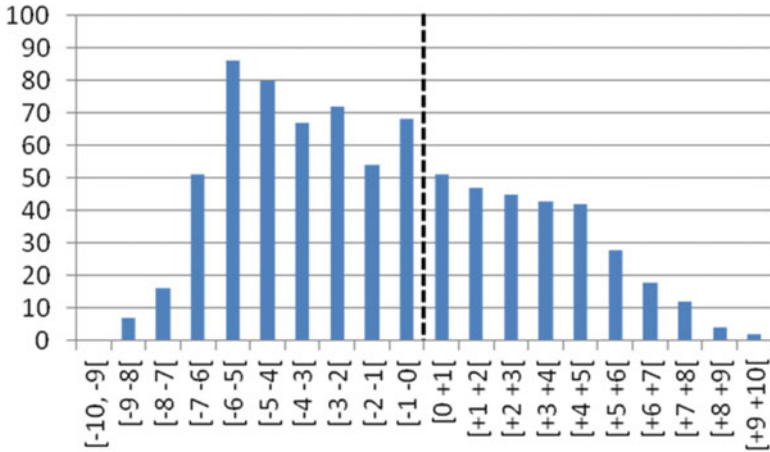


Fig. 18.1 Clip occurrence on the Train set as a function of the CSR

(a) the overlap segment duration, (b) the clip overlap duration as the summation of each overlap segment duration of the clip, (c) the mean overlap duration of a clip as the ratio of the clip overlap duration to the number of overlaps occurring in the clip, and (d) the percentage of overlap duration of the clip as the ratio of the clip overlap duration to the clip duration.

18.2.2 SSPNet Train Set Statistics

We analyzed the statistics of the SSPNet database Train set in focusing on the main characteristics of overlap segments; some statistics of the moderator were also investigated. The Train set includes 793 clips and has a total duration of 23,774 s (two clips' duration is inferior to 30 s), with 82 speakers (one moderator and 81 participants).

We analyzed the 4,143 segments of 23,774 s duration that were obtained by the clip diarization given in the SSPNet database. These segments were split according to the number of speakers that occurred in the segment: (1) 34 segments of a total duration of 89.9 s, which correspond to gaps in which nobody is speaking, (2) 2,638 segments of a total duration of 20,083.5 s, in which a lonely subject is speaking, and (3) 1,471 segments of a total duration of 3,600.6 s, in which two subjects are speaking. No segment was identified that had three or more speakers.

Figure 18.2 shows the histogram for each CSR of the average of the number of interruptions (i.e., the segments of overlapping speech) of the CSR clips. The horizontal dashed line represents the average of the number of interruptions of the Train set clips. Except for the CSR $[-1, 0[$, all of the CSRs of LLC have a mean number of interruptions that are below the average value ($1.85 = 1,471/793$). The HLC clips have more interruptions than the LLC clips.

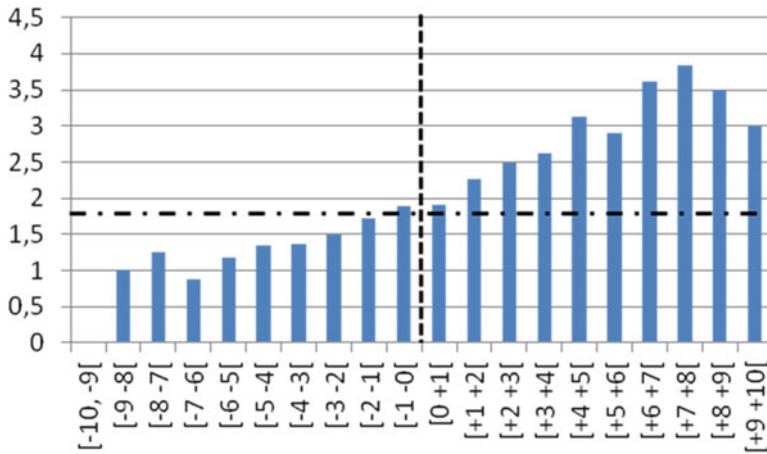


Fig. 18.2 Average of the number of interruptions as a function of the CSR

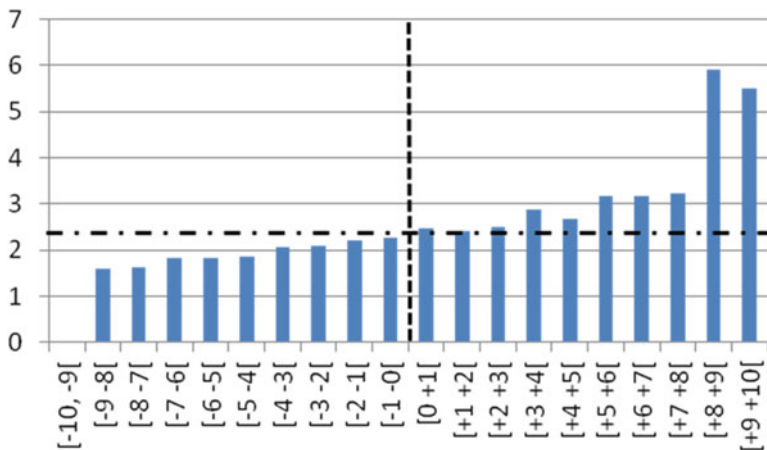


Fig. 18.3 Overlap mean duration (in s) as a function of the CSR

Figure 18.3 shows the histogram of the overlap mean duration for each CSR. The horizontal dashed line represents the average of the overlap duration in the Train set (2.45 s = 3,600.6/1,471). HLC clips have a mean duration of overlap that is higher than the LLC clips.

Figure 18.4 shows the histogram for each CSR of the percentage of overlap duration. The horizontal dashed line represents the mean percentage of the overlap duration of the Train set clips (15.1 % = 3,600.6/23,774). The conflict level is shown to be highly correlated to the percentage of overlap duration as in related work (Kim et al. 2012b).

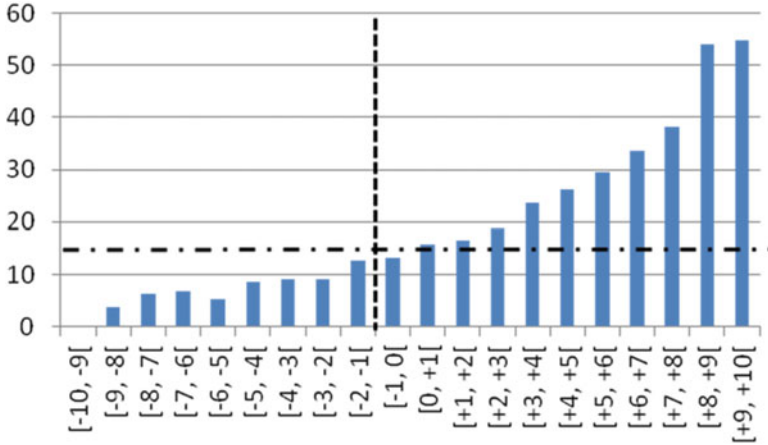


Fig. 18.4 Percentage of overlap duration as a function of the CSR

Table 18.1 Statistics on the speech duration of the moderator

Moderator—spk-050	Train set	Non-Ov	Ov	LLC-Ov	HLC-Ov
Total speech duration (in s)	27,284.7	20,083.5	7,201.2	2,619.0	4,582.2
Speech duration of the moderator (in s)	5,149.7	3,183.7	1,966	1,029.8	936.2
Percentage of the moderator speech duration	18.9	15.9	27.3	39.3	20.4

In the multiparty discussions of the debates, a member of the group held the role of moderator among the participant members who were taking part in the debate. We analyzed various statistics that were related to the moderator from the Train set.

In Table 18.1, the statistics on the speech duration of the moderator are given. The total speech duration (27,284.7 s) that we accounted for is different from the total segment duration (23,774 s) of the Train set; it was estimated as the duration of a segment in which a lonely subject is speaking plus twice the segment duration in which two subjects are speaking. The speech duration of the moderator was computed for the various classes of speech: Nonoverlap (Non-Ov) and overlap (Ov) were split into the two conflict level classes, low-level conflict overlap (LLC-Ov) and high-level conflict overlap (HLC-Ov). The speech duration of the moderator is given (in s). The percentage of the moderator speech duration was computed as the ratio between the speech duration of the moderator and the total speech duration. From the previous statistics, we note in Table 18.1 that the moderator speaks more during LLC-Ovs than during HLC-Ovs (39.3 % vs. 20.4 %) and Non-Ovs (39.3 % vs. 15.9 %).

In Table 18.2, the modes of the interruptions that were related to the moderator were analyzed according to the following occurrences: “the moderator interrupted a participant” or “the moderator was interrupted by a participant.” Clips were extracted from the video into 30-s duration segments. The mode of interruption that was related to the moderator was defined from an overlap in which the moderator

Table 18.2 Statistics on interruption mode occurrences of the moderator

Moderator—spk-050	Ov	LLC-Ov	HLC-Ov
# of interruptions	1,353	604	749
# of interruptions by the moderator and occurrence percentage	645 (47.7 %)	357 (59.1 %)	288 (38.4 %)
# of interruptions of the moderator and occurrence percentage	198 (14.6 %)	127 (21.0 %)	71 (9.4 %)

was speaking, by examining the previous segment: if in this segment the moderator was speaking, then the moderator was interrupted by a participant; otherwise, the moderator interrupted a participant. Taking off the first segment of each clip, an interruption occurs at the beginning of each overlap segment; the total number of interruptions in the Train set is 1,353 split into 604 interruptions in LLC-Ovs and 749 interruptions in HLC-Ovs. The number of interruptions by the moderator (respectively, the interruptions of the moderator) was computed for the overlaps and their two categories (LLC and HLC) as well as its percentage of occurrence. We note that the moderator interrupted the participants more often than the moderator was interrupted by the participants (47.7 % vs. 14.6 %). Moreover, the moderator interrupted the participants more in the LLC-Ovs than in the HLC-Ovs (59.1 % vs. 38.4 %).

18.3 Conflict Challenge

The Conflict Challenge was one of the shared tasks that was organized during the Interspeech 2013 Computational Paralinguistics Challenge (Schuller et al. 2013), which took place from January 15 to May 24, 2013. The task consisted of an automatic analysis of the group discussions, to retrieve the conflicts. The goal of this competition was to bridge the gap between research in automatic conflict detection and the low compatibility of the results. The task data were split into the Train, Development, and Test sets. The speaker dependence between these sets was reduced to a minimum that was needed in the real-life settings. As usual, the criterion to guide the detection strategy is the maximization of the UAR on the Development set. This set is also used to tune the parameters of the learning algorithms. Metadata are available only for the Train and Development sets. The participants did not have access to the labels of the Test set. However, each participant could upload the instance predictions up to five times, to receive the confusion matrix and the results from the Test set. The official measure of the competition is the UAR. An official system of conflict detection was also provided with the following characteristics: the WEKA data mining tool kit was used as a framework for the classification task (Hall et al. 2009), and the support vector machine (SVM) classifier with linear kernel and sequential minimal optimization (SMO) was used for learning; the official set of features (6,373 features), which

is referred to as the IS-2013 set, was a representation of the utterances, and the complexity parameter of the SVM classifier was optimized by using UAR on the Development set.

18.3.1 Audio Feature Sets

In this section, we describe the audio feature sets that we used for analyzing speech segments. This speech representation (Vogt and André 2005; Schuller et al. 2008) is a new paradigm for speech analysis. It contrasts with the standard paradigm for speech analysis (the sequence of observation vectors): regardless of its duration, a speech utterance is represented by a large set of features, which is termed an audio feature set. The feature set is based on several low-level descriptors (LLDs) that are computed from short overlapping windows of the audio signal. These LLDs comprise the loudness, the harmonics-to-noise ratio, the zero-crossing rate, the spectral and prosodic coefficients, the formant positions and bandwidths, the duration of voiced/unvoiced speech segments, and features derived from the long-term average spectrum such as band energies, roll-off, and centroid as well as voice quality features such as jitter and shimmer. Various global statistical functions (functionals) are computed on these LLDs to obtain feature vectors of equal size for each speech utterance. The sequence of LLDs that are associated with speech utterances can have different lengths, depending on the duration; the use of functionals allows us to obtain one feature vector per speech utterance, with a constant number of elements. It avoids the use of the expensive procedures of time warping between sequences of different lengths such as dynamic programming algorithms. Some functionals aim at estimating the spatial variability (e.g., mean, standard deviation, quartiles 1–3), and others aim at the temporal variability (e.g., peaks, linear regression slope). The four audio feature sets that we used for our experiments include the set of features that was provided by the organizers of the Interspeech 2010 (IS-2010) Paralinguistic Challenge (Schuller et al. 2010), the set of features for the Interspeech 2011 (IS-2011) Speaker State Challenge (Schuller et al. 2011), the set of features for the Interspeech 2012 (IS-2012) Speaker Trait Challenge (Schuller et al. 2012), and the set of features for the Interspeech 2013 (IS-2013) Conflict Sub-Challenge (Schuller et al. 2013). All of the features were extracted using the open source openSMILE feature extraction tools (Eyben et al. 2010). The IS-2010 feature set consists of 1,582 audio features, which were computed from 38 LLDs and 21 functionals. The spectral features include loudness, mel-frequency cepstral coefficients, mel-frequency band energy, and line spectral pair frequencies. The prosodic and voice quality features comprise the pitch frequency and envelope, jitter, and shimmer. Functionals such as the mean, standard deviation, kurtosis, skewness, minimum and maximum value, relative position, linear regression coefficients, and quartile and percentile coefficients were applied on the LLDs. The IS-2011 feature set consists of 4,368 audio features, which were computed from 59 LLDs and 39 functionals. Additional LLDs, such as the auditory spectrum-derived loudness

Table 18.3 Official feature sets of Interspeech Challenges

Feature set	IS-2010	IS-2011	IS-2012	IS-2013
# of LLDs	38	59	64	59
# of functional	21	39	40	48
# of features	1,582	4,368	6,124	6,373

measure, RASTA-style filtered auditory spectra, and statistical spectral descriptors (such as flux, entropy, variance) have been introduced. Additional functionals, such as quadratic regression and linear predictive coefficients and peak distances, allowed a better estimation of the temporal variability. The IS-2012 feature set consists of 6,124 audio features, which were computed from 64 LLDs and 40 functionals. Few LLDs have been added, including the logarithmic harmonics-to-noise ratio, spectral harmonicity, and psychoacoustic spectral sharpness. Functionals that are related to the local extrema, such as the statistics of inter-maxima distances, have been introduced. Useless functionals have been removed to limit the number of the audio features. The IS-2013 feature set consists of 6,373 audio features, computed from 59 LLDs and 48 functionals. A total of 724 audio features were removed from the IS-2012 feature set, and 972 were added. New functionals that were related to the local extrema, such as the modeling of inter-maxima, have been introduced.

Table 18.3 summarizes the main characteristics of the used feature sets. The first three feature sets were used for the detection of overlap, and the last feature set was the official feature set for the detection of conflict.

18.4 Interruption Detection

From the previous statistics analyzed in Sect. 18.2, the conflict level was shown to be highly correlated to the mean number of interruptions (cf. Fig. 18.2), the mean duration of overlap (cf. Fig. 18.3), and the percentage of overlap duration (cf. Fig. 18.4). Detecting segments of overlap is a difficult problem without individual microphones (Yamamoto et al. 2005). The main problem is due to the nonstationary characteristics of the speech signal. An alternative approach is the use of a microphone array (Quinlan and Asano 2007). In this case, the estimation of the number of signal sources allows the detection of segments that contain more than one source of speech. Another approach, which is applied for improving the speaker diarization system, is the speech segmentation by a three-class hidden Markov model (Boakye et al. 2008), with the three classes corresponding to nonspeech, speech, and overlapping speech. Mel-frequency cepstral coefficients (MFCC), root mean square (RMS) energy, and linear predictive coding (LPC) residual energy features have been used, and they provided a precision of 66 % and a recall of 26 %. In our approach, we have chosen to develop a multi-resolution framework to estimate the overlap duration percentage. This approach is based on the fusion of various overlap detectors, in which each detector is estimated on the segments of a fixed and chosen duration.

18.4.1 Clip Segmentation and Relabeling

The clips were segmented into consecutive audio segments. Three segment durations were chosen for the multi-resolution: 1, 2, and 5 s. For a given duration of segment, two segment-based detectors were designed: (1) the first detector is a two-class classifier that is referred to as an {N, O}-detector; it classifies a segment into Non-Ov (N) or Ov (O), and (2) the second detector is a three-class classifier that is referred to as an {N, L, H}-detector, which classifies a segment into Non-Ov (N), LLC-Ov (L), or HLC-Ov (H). Then, for multi-resolution detection, six SVM-based overlap detectors have been developed: (1) three two-class SVM classifiers, which we called {N, O}_1, {N, O}_2, and {N, O}_5, for the three durations, and (2) three three-class SVM classifiers, which we called {N, L, H}_1, {N, L, H}_2, and {N, L, H}_5, for the three durations. These labels (N, O, H, and L) were computed from the SSPNet corpus metadata using speaker segmentation and conflict metadata. The Train and Development sets were relabeled using the multi-resolution framework of overlap localization. For each clip, diarization and conflict information are now represented by 102 labels: 60 labels for {N, O}_1 and {N, L, H}_1, 30 labels for {N, O}_2 and {N, L, H}_2, and 12 labels for {N, O}_5 and {N, L, H}_5. These new labels will be used for the training and testing of the various overlap detectors.

In Figs. 18.5 and 18.6, the row called *Time* gives the time in seconds in the range from 1 to 30 (i.e., the clip duration), and the row *Segmentation* is the representation of the diarization metadata of the clip: N-segments are colored in white, L-segments

Clip #Train_0001 - Conflict score -7.2 - Low-Level conflict																															
{N, L, H}_5	N					N					L					N					N					N					
{N, L, H}_2	N	N	N	N	N	N	N	N	N	L	L	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N
{N, L, H}_1	N	N	N	N	N	N	N	N	N	N	N	N	N	L	L	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N
{N, O}_5	N					N					O					N					N										
{N, O}_2	N	N	N	N	N	N	N	N	N	O	O	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	
{N, O}_1	N	N	N	N	N	N	N	N	N	N	N	N	N	O	O	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N
Segmentation																															
Time (s)	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	

Fig. 18.5 Train set relabeling for the Train_0001 clip of low-level conflict

Clip #Train_0006 - Conflict score 7.3 - High-Level conflict																															
{N, L, H}_5	N					N					H					H					N					H					
{N, L, H}_2	N	N	N	N	N	N	N	N	N	H	H	H	H	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	H
{N, L, H}_1	N	N	N	N	N	N	N	N	N	N	N	N	N	H	H	H	H	H	N	N	N	N	N	N	N	N	N	N	N	H	H
{N, O}_5	N					N					O					O					N					O					
{N, O}_2	N	N	N	N	N	N	N	N	N	N	O	O	O	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	O
{N, O}_1	N	N	N	N	N	N	N	N	N	N	N	N	N	O	O	O	O	N	N	N	N	N	N	N	N	N	N	N	N	N	O
Segmentation																															
Time (s)	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	

Fig. 18.6 Train set relabeling for the Train_0006 clip of high-level conflict

in gray, and H-segments in black. The other rows contain the relabeling according to the various detectors. For the three rows $\{N, O\}_x$ ($x \in \{1, 2, 5\}$), a segment is labeled O when it contains a part of overlap and, otherwise, N. For the three rows $\{N, L, H\}_x$ ($x \in \{1, 2, 5\}$), overlap segments are labeled according to the conflict level of the clip: L for LLC-Ov and H for HLC-Ov.

Figure 18.5 gives an instance of metadata relabeling for the LLC clip #Train_0001. For this clip, an LLC-Ov occurs over 13.01 and 14.4 s. The relabeling is O for the segments 14 and 15 of $\{N, O\}_1$, the segments 7 and 8 of $\{N, O\}_2$, and the segment 3 of $\{N, O\}_5$. The relabeling is L for the segments 14 and 15 of $\{N, L, H\}_1$, the segments 7 and 8 of $\{N, L, H\}_2$, and the segment 3 of $\{N, L, H\}_5$.

Figure 18.6 gives an instance of metadata relabeling for the HLC clip #Train_0006. For this clip, HLC-Ovs occur over 14.9 and 18.9 s and over 28.3 and 30 s. The relabeling is O for the segments 15, 16, 17, 18, 19, 29, and 30 of $\{N, O\}_1$, for the segments 8, 9, 10, and 15 of $\{N, O\}_2$, and for the segments 3, 4, and 6 of $\{N, O\}_5$. The relabeling is H for the segments 15, 16, 17, 18, 19, 29, and 30 of $\{N, L, H\}_1$, for the segments 7, 8, 9, 10, and 15 of $\{N, L, H\}_2$, and for the segments 3, 4, and 6 of $\{N, L, H\}_5$.

18.4.2 Two-Class $\{N, O\}$ Classifiers

Using relabeling, three two-class SVMs ($\{N, O\}_1$, $\{N, O\}_2$, $\{N, O\}_5$) were estimated on the Train set. Each SVM classifies a segment of a given duration (1, 2, and 5 s) into overlap (O) or Non-Ov (N). To account for the imbalanced class distribution, the upper-represented category (N) was down-sampled by a given factor. A factor of 4 was applied for the $\{N, O\}_1$ detector, a factor of 3 for the $\{N, O\}_2$ detector, and a factor of 2 for the $\{N, O\}_5$ detector. We investigated the effects of different feature sets on the accuracy of the overlap speech detection. Table 18.4 gives the accuracy rates (N-Acc. and O-Acc. in %) of the two-class

Table 18.4 Accuracy rates of the detectors $\{N, O\}$ on the Development set according to the feature sets. In bold, the best feature set

Detectors $\{N, O\}$	Feature set	N-Acc. (%)	O-Acc. (%)	UAR (%)
$\{N, O\}_1$	IS-2010	86.7	73.9	80.3
$\{N, O\}_1$	IS-2011	87.7	72.3	80.0
$\{N, O\}_1$	IS-2012	87.8	71.6	79.7
$\{N, O\}_2$	IS-2010	85.1	75.1	80.1
$\{N, O\}_2$	IS-2011	87.3	71.6	79.5
$\{N, O\}_2$	IS-2012	87.4	71.7	79.5
$\{N, O\}_5$	IS-2010	82.7	78.7	80.7
$\{N, O\}_5$	IS-2011	84.9	75.3	80.1
$\{N, O\}_5$	IS-2012	84.0	75.7	79.8

classifiers on the two classes (N and O) on the Development set. Three audio feature sets were compared: IS-2010, IS-2011, and IS-2012 (cf. Sect. 18.3.1). For all two-class classifiers, the overlaps are more difficult to detect than the Non-Ovs, and IS-2010 was the best feature set, with a UAR (in %) of slightly over 80 % for the three detectors. For the architecture of the conflict detector, we chose to use only the best two-class classifiers $\{N, O\}_1$, $\{N, O\}_2$, and $\{N, O\}_5$ with the IS-2010 audio feature set. Our assumption is that only the best overlap classifiers are relevant for the detection of conflict.

18.4.3 Three-Class $\{N, L, H\}$ Classifiers

Previous studies presented different typologies of overlaps: overlap and backchannel with overlap (Gravano and Hirschberg 2011) and competitive and collaborative overlaps (Oertel et al. 2012). A backchannel indicates that the speaker producing them follows and understands the other speaker. They are generally words, onomatopoeias, or other sounds produced in the background (Clancy et al. 1996). Collaborative or competitive interruptions are manifested by speech overlap, but only overlap from a competitive interruption can it be related to a conflict (Kurtić et al. 2012). In competitive overlaps, the incoming speaker attempts to forcefully take over the turn. In collaborative overlaps, the incoming speaker assists the current speaker in his or her speech. We chose to build classes of LLC-Ovs and HLC-Ovs by making the hypothesis that they would be separable acoustically and useful for conflict detection. This choice is supported by the observation that some of the LLC-Ovs of the Train set were backchannel with overlaps and/or collaborative overlaps.

Using relabeling, three three-class SVM classifiers ($\{N, L, H\}_1$, $\{N, L, H\}_2$, and $\{N, L, H\}_5$) were estimated on the Train set. Each SVM classifies a segment of a given duration (1, 2, and 5 s) into an H, L, or N. To account for the imbalanced class distribution, the upper-represented category (N) was down-sampled by a given factor. A factor of 8 was applied for the $\{N, L, H\}_1$ detector, a factor of 6 for the $\{N, L, H\}_2$ detector, and a factor of 3 for the $\{N, L, H\}_5$ detector. We investigated the effects of different feature sets on the accuracy rate of the overlap speech detection. Table 18.5 gives the accuracy rates of the three-class classifiers on the Development set. Three audio feature sets were compared: IS-2010, IS-2011, and IS-2012. IS-2010 was the best feature set for $\{N, L, H\}_1$, having a UAR of 61.1 %. IS-2011 was the best feature set for $\{N, L, H\}_2$, with a UAR of 61.3 %. IS-2010 was the best feature set for $\{N, L, H\}_5$, with a UAR of 63.5 %. The LLC-Ovs are more difficult to detect than the HLC-Ovs. Furthermore, the detection rate of the LLC-Ovs appears to decrease with the duration of the analyzed segment: 44.7 % for $\{N, L, H\}_5$ (5 s), 35.9 % for $\{N, L, H\}_2$ (2 s), and 31.7 % for $\{N, L, H\}_1$ (1 s). A possible explanation would be that the detector $\{N, L, H\}_5$ allows a better estimation of the overlap durations than the other detectors and, consequently, a better discrimination of the LLC- and HLC-Ovs. Indeed, the duration of the LLC-Ovs is lower on average than the HLC-Ovs (1.98 s vs. 2.75 s). For the architecture of the conflict detector,

Table 18.5 Accuracy rates of the detectors {N, L, H} on the Development set according to the feature sets. In bold, the best detector

Detectors	Feature set	N-Acc. (%)	L-Acc. (%)	H-Acc. (%)	UAR (%)
{N, L, H}_1	IS-2010	78.0	31.7	73.5	61.1
{N, L, H}_1	IS-2011	79.9	32.7	70.5	61.0
{N, L, H}_1	IS-2012	79.4	31.4	71.4	60.7
{N, L, H}_2	IS-2010	79.5	32.6	71.2	61.2
{N, L, H}_2	IS-2011	78.1	35.9	70.0	61.3
{N, L, H}_2	IS-2012	80.5	31.5	68.0	60.0
{N, L, H}_5	IS-2010	77.5	44.7	68.3	63.5
{N, L, H}_5	IS-2011	76.4	40.0	67.7	61.4
{N, L, H}_5	IS-2012	80.8	38.2	67.4	62.1

we chose to use only the best three-class classifier: {N, L, H}_5 with the IS-2010 audio feature set. Our assumption is that only the best overlap classifier is relevant for the detection of conflict.

18.4.4 Audio Characteristics of Overlaps

Previous studies (Smolenski and Ramachandran 2011; Shokouhi et al. 2013) have shown that the audio characteristics of overlapping speech are different from speech in which a lonely speaker occurs. We looked for the discriminating cues (1) between Ov and Non-Ov and (2) more specifically between HLC-Ov and LLC-Ov. For these investigations, we chose to study the segments that had a 5-s duration in the Train set for the best accuracy results of the 5-s-based {N, O} and {N, L, H} detectors (see, respectively, Tables 18.4 and 18.5 in Sect. 18.4). The 38 low-level descriptors (LLDs) of the IS-2010 feature set have been used as audio characteristics. The relevance of the LLDs was analyzed with respect to the classes Non-Ov/Ov, which are referred to as {N, O}, and the HLC-Ov/LLC-Ov, which are referred to as {H, L}. For each LLD, the relevance is given by the information gain (Rauber and Steiger-Garcia 1993), which is computed on the segments of 5-s duration with the following formula: $H(\text{class}) - H(\text{class}/\text{LLD})$, where H is the Shannon entropy. Four steps were defined to compute the entropy: (1) filtering of the IS-2010 features according to a given LLD, (2) clustering of the segments of the Train set using the filtered features, (3) computation of the contingency table from the class and the cluster associated with each segment, and (4) estimation of the entropy from the table of contingency. Table 18.6 gives the information gain computed on the Train set of the five best-ranked LLDs (over 38 LLDs) in discriminating LLC-Ovs and HLC-Ovs. The most relevant LLDs are the logarithmic powers of mel-frequency bands and, more precisely, the high-frequency bands and the normalized loudness. These results show that various acoustic differences exist between the two types of overlaps.

Table 18.6 Information gain of the five best-ranked LLDs of the IS-2010 audio feature set in discriminating LLC-Ovs and HLC-Ovs

Low-level descriptors (LLD)	Inf. gain	Rank (/38)
Log power [3,934–5,649 Hz]	0.130	1
Log power [2,682–3,934 Hz]	0.119	2
Log power [1,768–2,682 Hz]	0.107	3
Normalized loudness	0.102	4
Log power [5,649–8,000 Hz]	0.102	5

Table 18.7 Information gain of the five best-ranked LLDs of the IS-2010 audio feature set in discriminating Ovs and Non-Ovs

Low-level descriptors (LLD)	Inf. gain	Rank (/38)
Fundamental frequency (F0)	0.141	1
Log power [614–1,101 Hz]	0.129	2
Log power [0–259 Hz]	0.127	3
Jitter (DDP)	0.124	4
First mel-frequency cepstral coef.	0.121	5

Table 18.7 gives the information gain that is computed on the Train set of the five best-ranked LLDs (over 38 LLDs) in discriminating Ovs and Non-Ovs. According to the information gain rank, the most relevant LLDs are the fundamental frequency, the logarithmic powers, especially in low-frequency bands, the jitter, and the first mel-frequency cepstral coefficient. The usual representation techniques and algorithms are designed and interpreted for speech signals in which a lonely subject is speaking. In the case of overlapping speech in which two or more subjects are speaking, the usual algorithms are not adapted (e.g., the pitch algorithm); the computation of one fundamental frequency has no sense, and its computation was shown to be the most discriminant cue for detecting Ov/Non-Ov. For a speech representation such as the logarithmic power in the mel-frequency bands, the low-frequency bands in which the first two formants of the speaker occur were also shown to be discriminant. Last, the jitter DDP (difference of differences of periods) related to the pitch and the first mel-frequency cepstral coefficient related to the energy of the segment were also shown to be relevant for the discrimination Ov/Non-Ov.

18.5 Conflict Detector

Overlap detectors have been developed and assessed, to incorporate their knowledge in an improved conflict detector (conflict/nonconflict). Incorporating prior knowledge (Krupka and Tishby 2007; Li et al. 2008) in classification systems allowed an increase in the performance in many applications of pattern recognition (e.g., biomedical image, pathological voice). Various methods have been developed for neural network systems (Chen et al. 2000) and SVM classifiers (Decoste and Scholkopf 2002; Lauer and Bloch 2008). As defined by Schölkopf and Smola (2001), the methods developed for including prior knowledge in an

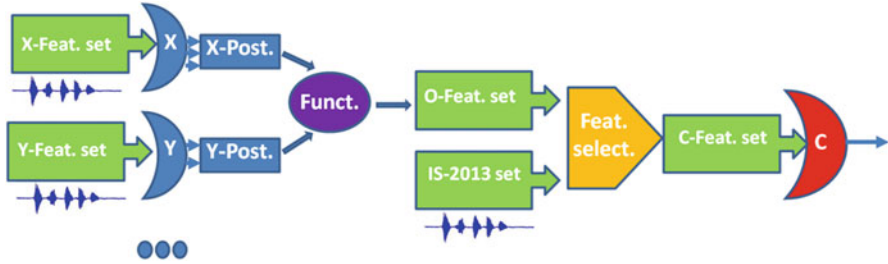


Fig. 18.7 Multi-expert architecture scheme of the conflict detector

SVM classifier can be divided into three categories: (1) the kernel methods with selection of the most appropriate kernel or the creation of a new kernel, (2) the optimization methods with the addition of constraints, and (3) the sample methods with data generation or modification of data representations. We have chosen the last category by developing an SVM-based detector, using as input a composite feature set. This feature set is a concatenation of selected audio features and posterior-based features that are computed from the posterior probabilities of the overlap detectors. The architecture characteristics of this classification system are close to those used in a mixture of experts (Jordan and Jacobs 1994). These approaches have theoretical advantages, such as a reduction in the hypothesis space and learning consistency. As described in Fig. 18.7, the multi-expert architecture scheme of the conflict detector has consisted of a set of overlap detectors (e.g., X, Y) and a conflict/nonconflict detector (C). A specialized audio feature set (e.g., X-Feat. set) was associated with each overlap detector (e.g., X), to represent the utterances. A conflict audio feature set (Cf-Feat. set) was associated with the conflict detector. This feature set consisted of the selection of the relevant features (Feat. Select.) that were extracted from the overlap feature set (Ov-Feat. set) and the IS-2013 feature set (cf. Sect. 18.3.1). A set of functionals (Funct.) was applied to the posterior probabilities of the overlap detectors (e.g., X-Post and Y-Post) to obtain the Ov-Feat. set.

We chose the overlap detectors giving the best UAR on the Development set (cf. Table 18.4 and 18.5): three two-class (Non-Ov/Ov) SVM-based detectors ($\{N, O\}_1$, $\{N, O\}_2$, $\{N, O\}_5$) and one three-class (Non-Ov/LLC-Ov/HLC-Ov) SVM-based detector ($\{N, L, H\}_5$).

18.5.1 Posterior Probabilities

Logistic regression models (Hosmer and Lemeshow 2000) were used to obtain the posterior probabilities from the four overlap detectors ($\{N, O\}_1$, $\{N, O\}_2$, $\{N, O\}_5$, and $\{N, L, H\}_5$). These posterior probabilities of the overlap detectors provide information about the uncertainty of belonging to one class: for example,

the probability of 60 % of a segment to be an overlap and 40 % to be a nonoverlap. There are various strategies for computing these probabilities, such as Platt’s method (Platt 2000), isotonic regression (Zadrozny and Elkan 2002), and Bayesian methods (Sollich 2002). These probabilities are useful to integrate expert classifiers such as overlap classifiers in a global decision process. This approach is a flexible architecture for making decisions without global optimization. The method of computation of the posterior probabilities depends on the chosen set of clips. The goal is to obtain a consistent computation of the posterior probabilities from the different corpora (Train, Development, and Test sets). For the Train and Development sets, the posterior probabilities have been computed by performing cross-predictions on the union of these two sets. This process consists of splitting the data set into s disjoint folds and predicting class posterior probabilities of each instance of a fold from a model trained on the $s - 1$ other folds. Sixteen folds have been chosen that have participant independence between two folds. For the Test set, the posterior probabilities have been computed from a model trained on the union of the Train and Development sets. A total of 120 posterior probabilities were computed for each clip: 60 for the $\{N, O\}_1$ detector, 30 for $\{N, O\}_2$, 12 for $\{N, O\}_5$, and 18 for $\{N, L, H\}_5$.

Figures 18.8 and 18.9 give an instance of the posterior probabilities from the four overlap detectors ($\{N, O\}_1$, $\{N, O\}_2$, $\{N, O\}_5$, and $\{N, L, H\}_5$), respectively, for the LLC clip #Train_0001 and the HLC clip #Train_0006. The row called *Time* gives

$\{N, L, H\}_5$ (N)	77			73			11			54			99			97														
$\{N, L, H\}_5$ (H)	00			01			51			21			00			02														
$\{N, L, H\}_5$ (L)	03			06			38			23			01			01														
$\{N, O\}_5$ (O)	01			08			60			80			01			01														
$\{N, O\}_2$ (O)	10	07	02	08	00	06	36	77	65	10	00	01	06	13	03															
$\{N, O\}_1$ (O)	03	08	01	01	04	01	03	04	06	01	04	10	10	76	55	40	19	24	03	29	02	01	03	02	35	03	14	04	02	03
Segmentation																														
Time (s)	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30

Fig. 18.8 Overlap posterior probabilities as percentages for the Train_0001 clip with low-level conflict

$\{N, L, H\}_5$ (N)	98			09			72			01			03			02														
$\{N, L, H\}_5$ (H)	01			84			26			99			94			89														
$\{N, L, H\}_5$ (L)	01			07			02			00			03			09														
$\{N, O\}_5$ (O)	03			62			12			99			95			99														
$\{N, O\}_2$ (O)	03	11	15	05	54	20	20	98	99	99	98	68	54	75	97															
$\{N, O\}_1$ (O)	01	38	35	05	02	07	04	13	12	89	68	17	19	06	02	99	99	99	99	16	87	99	58	19	05	46	21	82	29	95
Segmentation																														
Time (s)	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30

Fig. 18.9 Overlap posterior probabilities as percentages for the Train_0006 clip with high-level conflict

the time from 1 to 30 s (clip duration); the row *Segmentation* is the representation of the diarization metadata of the clip: N-segments are colored in white, L-segments in gray, and H-segments in black. The other rows contain the posterior probabilities presented as a percentage. For the three rows $\{N, O\}_x$ ($x \in \{1, 2, 5\}$), a segment of posterior probabilities that was higher than 50 was detected as O; otherwise, it was detected as N. The posterior probabilities that were associated with the $\{N, L, H\}_5$ detector are presented in the three other rows $\{N, L, H\}_5$ (N), $\{N, L, H\}_5$ (L), and $\{N, L, H\}_5$ (H) for, respectively, nonoverlap (N), low-level-conflict (L), and high-level-conflict (H). For a given segment, the higher probability (in bold) corresponds to the class that was detected.

In Fig. 18.8, the class O was detected for the segments 14 and 15 of $\{N, O\}_1$, the segments 8 and 9 of $\{N, O\}_2$, and the segments 3 and 4 of $\{N, O\}_5$. Class H was detected for the segment 3 for $\{N, L, H\}_5$. There are three wrong detections: the class O instead of N for the segment 9 of $\{N, O\}_2$ and the segment 4 of $\{N, O\}_5$, and the class H was detected instead of L for segment 3 for $\{N, L, H\}_5$.

In Fig. 18.9, the class O was detected for the segments 10, 11, 16, 17, 18, 19, 21, 22, 23, 28, and 30 of $\{N, O\}_1$, for the segments 5 and 8 through 15 of $\{N, O\}_2$, and for the segments 2, 4, 5, and 6 of $\{N, O\}_5$. Class H was detected for the segments 2, 4, 5, and 6 of $\{N, L, H\}_5$. There are ten wrong detections: the class O instead of N for the segments 10, 11, and 28 of $\{N, O\}_1$, the segments 5, 13, and 14 of $\{N, O\}_2$, the segment 2 of $\{N, O\}_5$, the class N instead of O for the segments 15 and 29 of $\{N, O\}_1$, and the class H instead of N for the segment 2 of $\{N, L, H\}_5$. We note that there was no wrong decision for segments 21, 22, and 23 of $\{N, O\}_1$, for the segments 11 and 12 of $\{N, O\}_2$, and for segment 5 of $\{N, O\}_5$ and $\{N, L, H\}_5$; after listening, an overlap occurs effectively from 20.3 to 22.2 s but was not labeled in the metadata.

18.5.2 *Overlap Feature Sets*

One hundred and twenty posterior probabilities were computed for each clip. These values depend on the time and represent the temporal shape of a conflict in terms of the overlap. There are specific temporal shapes for conflict escalation (Kim et al. 2012c), but the 797 clips of the Train set are insufficient to model these temporal shapes. We have chosen to apply statistical functionals to the posterior probabilities; the purpose was to obtain an overlap feature set that is related to the percentage of overlap duration. Three functionals have been chosen: mean, correlation, and covariance. The mean functional was applied to the posterior probabilities of $\{N, O\}_1$, $\{N, O\}_2$, $\{N, O\}_5$ for the class O and to the posteriors of $\{N, L, H\}_5$ for the classes N, L, and H. The correlation functional was applied between the posterior probabilities of the class O for all combinations of $\{N, O\}_1$, $\{N, O\}_2$ and $\{N, O\}_5$. Table 18.8 gives a list of the ten features that were computed by the mean and correlation functionals.

Functional covariance is a functional of a functional. It was applied to the mean and correlation functionals. The interest of this functional is to reveal the cofactors.

Table 18.8 List of the features computed by the mean and correlation functionals

Mean and correlation functionals	Feature name
Mean (post ($\{N, O\}_1(O)$))	O1
Mean (post ($\{N, O\}_2(O)$))	O2
Mean (post ($\{N, O\}_5(O)$))	O5
Correlation (post ($\{N, O\}_1(O)$), post ($\{N, O\}_2(O)$))	O12
Correlation (post ($\{N, O\}_1(O)$), post ($\{N, O\}_5(O)$))	O15
Correlation (post ($\{N, O\}_2(O)$), post ($\{N, O\}_5(O)$))	O25
Correlation (post ($\{N, O\}_1(O)$), post ($\{N, O\}_2(O)$), post($\{N, O\}_5(O)$))	O125
Mean (post ($\{N, L, H\}_5(N)$))	N5
Mean (post ($\{N, L, H\}_5(L)$))	L5
Mean (post ($\{N, L, H\}_5(H)$))	H5

Table 18.9 Information gain of the 15 best-ranked LLDs of the audio feature set, including the Ov-2 feature set and the IS-2013 feature set

Features	Information gain	Rank (/6,428)
Cov_O125_O125	0.43862	1
Cov_O12_O125	0.43758	2
Cov_O1_O12	0.43586	3
Cov_O1_O125	0.43177	4
Cov_O15_O125	0.42914	5
Cov_O12_O12	0.42858	6
Cov_O25_O125	0.41965	7
Cov_O12_O15	0.41957	8
Cov_O12_O25	0.41431	9
Cov_O15_O15	0.41429	10
Cov_O15_O25	0.41325	11
Cov_H5_O1	0.40915	12
Cov_H5_O15	0.40849	13
Cov_O1_O25	0.40705	14
Cov_H5_O12	0.40509	15

Two overlap feature sets have been defined. The first feature set, called Ov-1, consisted of 28 features; it was computed by the covariance functional applied to the features that are related to the $\{N, O\}$ detectors (O1, O12, O15, O2, O25, and O125). The second feature set, called Ov-2, consisted of 55 features; it was computed by the covariance functional applied to the features that are related to the $\{N, O\}$ and $\{N, L, H\}$ detectors (O1, O12, O15, O2, O25, O125, N5, L5, and H5). These two feature sets will allow a contrastive test to measure the contribution of the $\{N, L, H\}_5$ detector in the detection of conflict. The method of information gain was used to analyze the feature relevance of the Ov-2 set in comparison with those of the IS-2013 set. Table 18.9 gives the information gain computed on the Train set and the rank on 6,428 features (55 features from the Ov-2 set and 6,373 features from the IS-2013 set) of the most relevant features for the conflict detection. The best feature is the Cov_O125_O125 feature (which is equal to O125 multiplied by O125).

Table 18.10 Characteristics of the conflict feature sets

Feature set	Selected feat. set	# of selected feat.	# of selected feat. from Ov features	# of selected feat. from IS-2013
Ov-1 and IS-2013 (6,428 features)	Cf-1	315	15	300
Ov-2 and IS-2013 (6,401 features)	Cf-2	335	45	290

Table 18.11 UAR in the conflict detection task on the Development set according to the conflict feature sets

Feature set	IS-2013	Cf-1	Cf-2
# of features	6,373	315	335
UAR (devel. set) (%)	79.1	87.4	88.3

The 12th rank of the Cov_H5_O1 feature shows that the {N, L, H}_5 detector is relevant for the detection of conflict. A total of 36 out of 55 features of the Ov-2 set have better information gain than those of the IS-2013 set. These results show the interest of the overlap feature sets for the detection of conflict.

18.5.3 Conflict Feature Sets

From two initial feature sets (Ov-1 and IS-2013 and Ov-2 and IS-2013), two conflict feature sets (Cf-1 and Cf-2) were selected by a backward selection algorithm when maximizing UAR on the Development set for the conflict detection task. Table 18.10 gives the characteristics of the Cf-1 and Cf-2 sets of the conflict detector using these feature sets. The Cf-1 feature set consists of 315 features (15 features from the Ov-1 set and 300 features from the IS2013 set). The Cf-2 feature set consists of 335 features (45 features from Ov-2 set and 290 features from the IS2013-set).

Table 18.11 gives the accuracy (UAR in %) of the conflict detection on the Development set using the various feature sets (IS-2013, Cf-1, Cf-2). The results show an improvement of 8.3 % using the Cf-1 set and 9.2 % using the Cf-2 set on the Development set compared to the baseline results that use the IS-2013 set (UAR of 79.1 %). These results show also that the majority of the features of the Cf-2 set are relevant and not redundant. It confirms that the two types of detectors ({N, L, H} and {N, O}) are relevant for the detection of conflict.

18.5.4 Conflict Detectors

Two conflict detectors have been developed. Figure 18.10 resumes the architecture characteristics of the first conflict detector, called the simple overlap-based conflict detector (SO-conflict detector). This detector was based on a set of overlap detectors

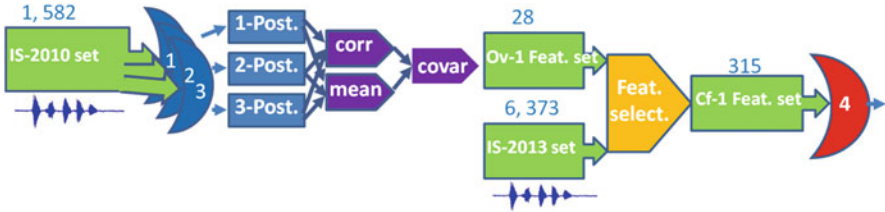


Fig. 18.10 Architecture scheme of the SO-conflict detector

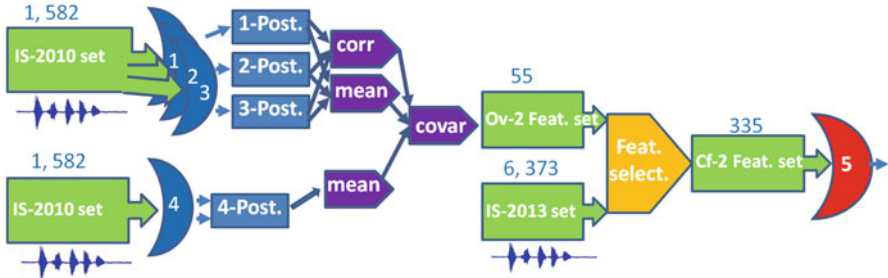


Fig. 18.11 Architecture scheme of the AO-conflict detector

(1, 2, and 3) that correspond to the three multi-resolution-based $\{N, O\}$ detectors and a conflict detector (4). The IS-2010 feature set (1,582 features) was used for the overlap detectors. The Cf-1 feature set (315 features) was associated with the conflict detector. The Cf-1 feature set was obtained by a backward selection algorithm from the Ov-1 feature set (28 features) and the IS-2103 set (6,373 features).

Figure 18.11 resumes the architecture characteristics of the second conflict detector, called the advanced overlap-based conflict detector (AO-conflict detector). This detector was based on a set of overlap detectors (1, 2, 3, and 4) and a conflict detector (5). The IS-2010 audio feature set (1,582 features) was used for the overlap detectors. The Cf-2 feature set (335 features) was associated with the conflict detector. The Cf-2 feature set was obtained by a backward selection algorithm from the Ov-2 feature set (55 features) and the IS-2103 set (6,373 features).

18.5.5 Conflict Detection on the Test Set

The Test set of the Interspeech 2013 Conflict Challenge (Schuller et al. 2013) consisted of 397 clips with no information or metadata available. Table 18.12 gives the results obtained on the Test set during the Conflict Challenge. Experiments gave a UAR of 83.4 % for the SO-conflict detector and a UAR of 85.3 % for the AO-conflict detector. These results show an improvement of 2.6 % (SO-conflict detector)

Table 18.12 Assessment on the Test set. In bold, the best conflict detector

Conflict detector	# of features	UAR (%) on Test set
SO-conflict detector	315	83.4
AO-conflict detector	335	85.3
IS-2013 baseline system	6,373	80.8
Grèzes et al. (2013)	1	83.1
Räsänen and Pohjalainen (2013)	349	83.9

and 4.5 % (AO-conflict detector) on the Test set compared to the baseline results with the IS-2013 set (UAR of 80.8 %) for the conflict detection task. These results confirm also that the two types of overlap detectors ($\{N, L, H\}$ and $\{N, O\}$) are relevant for the detection of conflict. The other results are those obtained by the other participants. In Grèzes et al. (2013), a UAR of 83.1 % was obtained on the Test set using a unique feature: the percentage of overlap predicted by an SVM-based regression model. In Räsänen and Pohjalainen (2013), a UAR of 83.9 % was obtained on the Test set using 349 relevant features selected from the IS-2013 feature set. Feature relevance was computed by a random process. We notice that the two better results were obtained by a similar number of features (335 vs. 349).

18.6 Conclusions

This article presents and assesses a detection system of conflict in group discussions from voice analysis. The system was based on a multi-expert architecture and detected two states (conflict/nonconflict). The analysis of the Train set of the SSPNet database has demonstrated that the conflict level was highly correlated with the mean number of interruptions, the mean duration of overlap, and the percentage of overlap duration. The multi-expert architecture enabled knowledge regarding overlaps to be used in the conflict detector.

The concept of LLC-Ovs and HLC-Ovs has been introduced and investigated. Two types of overlap detectors have been developed: the first type aims at detecting whether a speech segment contains overlap, and the second type aims at detecting whether a speech segment contains an LLC-Ov or HLC-Ov. The accuracy of the detectors shows that the LLC-Ovs and HLC-Ovs can be modeled. The high-frequency mel bands and the normalized loudness are shown to be the audio characteristics that are relevant to discriminating these two types of overlap. A multi-resolution framework has been developed for the overlap detectors, to improve the robustness of the detection. Three segment durations have been chosen (1, 2, and 5 s). The experiments have shown that these detectors were not redundant.

A composite set of 335 features, which consist of audio-based features and overlap detector-based features, has been defined for the conflict detection task of the Interspeech 2013 Conflict Challenge. The performance obtained for the Test set

gave a UAR of 85.3 %. These results show an improvement of 4.5 % compared to the results of the baseline system of the Conflict Challenge (UAR of 80.8 %).

These experiments have shown the capability of a multi-expert architecture to integrate a piece of conflict knowledge. Other knowledge that is related to the turn-taking patterns, such as the modeling of the moderator role (Vinciarelli 2007), or that is related to the nonverbal interactions, such as the movements of the body, the head, and the arms, could be integrated into the conflict detector.

Acknowledgments Many thanks to Björn Schuller (TUM, Germany), Stefan Steidl (FAU Erlangen-Nuremberg, Germany), and Anton Batliner (TUM, Germany) for the organization of the Interspeech 2013 Conflict Challenge and special thanks to Alessandro Vinciarelli (University of Glasgow, UK) for the SSPNet Conflict Corpus.

References

- Atkinson JM, Drew P (1979) Order in court: the organisation of verbal interaction in judicial settings. Humanities Press, Atlantic Highlands, NJ
- Barzilay R, Collins M, Hirschberg J, Whittaker S (2000) The rules behind the roles: identifying speaker roles in radio broadcasts. Paper presented at 17th National conference on artificial intelligence, Austin, USA, 30 July–3 Aug, pp 679–684
- Beattie GW (1982) Turn-taking and interruption in political interviews: Margaret Thatcher and Jim Callaghan compared and contrasted. *Semiotica* 39(1–2):93–114
- Boakye K, Trueba-Hornero B, Vinyals O, Friedland G (2008) Overlapped speech detection for improved diarization in multi-party meetings. Paper presented at ICASSP Conference, Las Vegas, USA, 31 Mar–4 Apr, pp 4353–4356
- Boden D (1994) The business of talk. Organizations in action. Polity Press, London
- Brinson SL, Winn JE (1997) Talk shows' representations of interpersonal conflicts. *J Broadcast Electron Media* 41(1):25–39
- Chen Z, Feng TJ, Houkes Z (2000) Incorporating a priori knowledge into initialized weights for neural classifier. Paper presented at international joint conference on neural networks (IJCNN), Como, Italy, 24–27 July, pp 291–296
- Clancy PM, Thompson SA, Suzuki R, Tao H (1996) The conversational use of reactive tokens in English, Japanese and Mandarin. *J Pragmat* 26:355–387
- Daly TM, Lee JA, Soutar GN, Rasmi S (2010) Conflict-handling style measurement: a best-worst scaling application. *Int J Confl Manag* 21(3):281–308
- De Ruiter JP, Mitterer H, Enfield NJ (2006) Projecting the end of a speaker's turn: a cognitive cornerstone of conversation. *Language* 82(3):515–535
- Decoste D, Scholkopf B (2002) Training invariant support vector machines. *Mach Learn* 46(1–3):161–190
- Eyben F, Wöllmer M, Schuller B (2010) openSMILE the Munich versatile and fast open-source audio feature extractor. Paper presented at the ACM multimedia conference (MM), Florence, Italy, 25–29 Oct, pp 1459–1462
- Garcia A (1991) Dispute resolution without disputing: how the interactional organization of mediation hearings minimizes argumentative talk. *Am Sociol Rev* 56:818–835
- Gravano A, Hirschberg J (2011) Turn-taking cues in task oriented dialogue. *Comput Speech Lang* 25(3):601–634
- Grèzes F, Richards J, Rosenberg A (2013) Let me finish: automatic conflict detection using speaker overlap. Paper presented at the Interspeech conference, Lyon, France, 25–29 Aug, 5 pages

- Hall M, Frank E, Holmes G, Pfahringer B, Reutemann P, Witten I (2009) The WEKA data mining software: an update. *SIGKDD Explor* 11:10–18
- Heath C, Luff P (2007) Ordering competition: the interactional accomplishment of the sale of art and antiques at auction. *Br J Sociol* 58:63–85
- Hosmer DW, Lemeshow S (2000) *Applied logistic regression*, 2nd edn. Wiley, New York
- Jordan MI, Jacobs RA (1994) Hierarchical mixtures of experts and the EM algorithm. *Neural Comput* 6:181–214
- Kim S, Filippone M, Valente F, Vinciarelli A (2012a) Predicting the conflict level in television political debates: an approach based on crowdsourcing, nonverbal communication and Gaussian processes. Paper presented at the ACM conference on multimedia, Nara, Japan, pp 793–796
- Kim S, Valente F, Vinciarelli A (2012b) Automatic detection of conflicts in spoken conversations: ratings and analysis of broadcast political debates. Paper presented at ICASSP, Kyoto, Japan, 25–30 Mar, pp 5089–5092
- Kim S, Yella SH, Valente FA (2012c) Automatic detection of conflict escalation in spoken conversations. Paper presented at Interspeech Conference, Portland, USA, OR, 9–13 Sept, 4 pages
- Korabik K, Baril GL, Watson C (1993) Managers' conflict management style and leadership effectiveness: the moderating effects of gender. *Sex Roles* 29(5–6):405–418
- Krupka E, Tishby N (2007) Incorporating prior knowledge on features into learning. *J Mach Learn Res* 2:227–234
- Kurtié E, Brown GJ, Wells B (2012) Resources for turn competition in overlapping talk. *Speech Comm* 55:1–23. doi:[10.1016/j.specom.2012.10.002](https://doi.org/10.1016/j.specom.2012.10.002)
- Lauer DF, Bloch G (2008) Incorporating prior knowledge in support vector machines for classification: a review. *Neurocomputing* 71(7–9):1578–1594
- Lerner GH (1995) Turn design and the organization of participation in instructional activities. *Discourse Process* 19(1):111–131
- Li Y, de Ridder D, Duin RPW, Reinders MJT (2008) Integration of prior knowledge of measurement noise in Kernel density classification. *Pattern Recogn* 41:320–330
- Mac Houl A (1978) The organization of turns at formal talk in the classroom. *Lang Soc* 7:183–213
- Macintosh G, Stevens CJ (2008) Personality, motives and conflict strategies in everyday service encounters. *Int J Confl Manag* 19(2):112–131
- Mehan H (1985) The structure of classroom discourse. In: Dijk TA (ed) *Handbook of discourse analysis*, vol 3. Academic, New York, pp 120–131
- Mondada L (2012) The dynamics of embodied participation and language choice in multilingual meetings. *Lang Soc* 41:1–23
- Mondada L (2013) Embodied and spatial resources for turn-taking in institutional multi-party interactions: participatory democracy debates. *J Pragmat* 46(1):39–68
- Oertel C, Włodarczak M, Tarasov A, Campbell N, Wagner P (2012) Context cues for classification of competitive and collaborative overlaps. Paper presented at Speech Prosody Conference, Shanghai, China, 22–25 May, 4 pages
- Pesarin A, Cristani M, Murino V, Vinciarelli A (2012) Conversation analysis at work: detection of conflict in competitive discussions through semi-automatic turn-organization analysis. *Cogn Process* 13(2):533–540
- Platt JC (2000) Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. In: Bartlett PJ, Schölkopf B, Schuurmans D, Smola AJ (eds) *Advances in large margin classifiers*. MIT Press, Cambridge, pp 61–74
- Quinlan A, Asano F (2007) Detection of overlapping speech in meeting recordings using the modified exponential fitting test. Paper presented at the European signal processing conference, Poznan, Poland, 3–7 Sept, pp 2360–2364
- Rahim MA (1983) A measure of styles of handling interpersonal conflict. *Acad Manag J* 26(2):368–376

- Räsänen O, Pohjalainen J (2013) Random subset feature selection in automatic recognition of developmental disorders, affective states, and level of conflict from speech. Paper presented at the Interspeech conference, Lyon, France, 25–29 Aug, 5 pages
- Rauber TW, Steiger-Garcia AS (1993) Feature selection of categorical attributes based on contingency table analysis. Paper presented at the Portuguese conference on pattern recognition, Porto, Portugal
- Sacks H, Schegloff EA, Jefferson G (1974) A simplest systematics for the organization of turn-taking for conversation. *Language* 50(4):696–735
- Schegloff EA (1987) Between macro and micro: contexts and other connections. In: Alexander J, Giesen B, Munch R, Smelser N (eds) *The micro-macro link*. University of California Press, Berkeley, pp 207–234
- Schölkopf BAJ, Smola AJ (2001) *Learning with Kernels: support vector machines, regularization, optimization, and beyond*. MIT Press, Cambridge, MA
- Schuller B, Wimmer M, Moesenlechner L, Kern C, Arsic D, Rigoll G (2008) Brute-forcing hierarchical functional for paralinguistics: a waste of feature space? Paper presented at the ICASSP conference, pp 4501–4504
- Schuller B, Steidl S, Batliner A, Burkhardt F, Devillers L, Müller C, Narayanan S (2010) The Interspeech 2010 paralinguistic challenge. Paper presented at the Interspeech conference, Makuhari, Japan, 26–30 Sept, pp 2794–2797
- Schuller B, Batliner A, Steidl S, Schiel F, Krajewski J (2011) The Interspeech 2011 speaker state challenge. Paper presented at the Interspeech conference, Florence, Italy, 28–31 Aug, 4 pages
- Schuller B, Steidl S, Batliner A, Noth E, Vinciarelli A, Burkhardt F, van Son R, Wenginger F, Eyben F, Bocklet T, Mohammadi G, Weiss B (2012) The Interspeech 2012 speaker trait challenge. Paper presented at the Interspeech conference, Portland, OR, USA, 9–13 Sept, 4 pages
- Schuller B, Steidl S, Batliner A, Vinciarelli A, Scherer K, Ringeval F, Chetouani M, Wenginger F, Eyben F, Marchi E, Mortillaro M, Salamin H, Polychroniou A, Valente F, Kim S (2013) The Interspeech 2013 computational paralinguistics challenge: social signals, conflict, emotion autism. Paper presented at the Interspeech conference, Lyon, France, 25–29 Aug, 5 pages
- Shokouhi N, Sathyanarayana A, Sadjadi SO, Hansen JHL (2013) Overlapped-speech detection with applications to driver assessment for in-vehicle active safety systems. Paper presented at ICASSP conference, Vancouver, Canada, 26–31 May, pp 2834–2838
- Smolenski B, Ramachandran R (2011) Usable speech processing: a filterless approach in the presence of interference. *Circuits Syst Mag IEEE* 11(2):8–22
- Sollich P (2002) Bayesian methods for support vector machines: evidence and predictive class probabilities. *Mach Learn* 46:21–52
- Svennevig J (2008) Exploring leadership conversations. *Manag Commun Q* 21:529–536
- Thomas KW, Thomas GF, Schaubhut N (2008) Conflict styles of men and women at six organization levels. *Int J Conf Manag* 19(2):148–166
- Valente F, Vinciarelli A (2010) Improving speech processing through social signals: automatic speaker segmentation of political debates using role based turn-taking patterns. Paper presented at the International workshop on social signal processing, Firenze, Italy, 25–29 Oct, pp 29–34
- Vinciarelli A (2007) Speakers role recognition in multiparty audio recordings using social network analysis and duration distribution modeling. *IEEE Trans Multimed* 9(6):1215–1226
- Vinciarelli A (2009) Capturing order in social interactions. *Signal Process Mag IEEE* 26(5):133–152
- Vogt T, André E (2005) Comparing feature sets for acted and spontaneous speech in view of automatic emotion recognition. Paper presented at the ICME conference, Amsterdam, The Netherlands, 6–8 July, pp 474–477
- Yamamoto K, Asano F, Yamada T, Kitawaki N (2005) Detection of overlapping speech in meetings using support vector regression. Paper presented at the international workshop on acoustic echo and noise control (IWAENC), Eindhoven, The Netherlands, 12–15 Sept, pp 2158–2165
- Zadrozny B, Elkan C (2002) Transforming classifier scores into accurate multiclass probability estimates. Paper presented at the international conference on knowledge discovery and data mining, Edmonton, Canada, 23–25 July, pp 694–699

Chapter 19

Be at Odds? Deep and Hierarchical Neural Networks for Classification and Regression of Conflict in Speech

Raymond Brueckner and Björn Schuller

19.1 Introduction

The field of computational paralinguistics is dedicated to the study of non-verbal elements of speech that convey information about human affect, emotion, personality, and speaker states and traits by applying mathematical models of the underlying mechanisms, and there is an increasing amount of research in that field (Vinciarelli et al. 2009; Schuller 2012; Schuller and Batliner 2013).

One important use case of computational paralinguistics is the automatic detection of *conflict*. Reliably detecting conflict is of high interest for the deployment of artificial intelligence agents, customer centers, security and intelligence, and general data mining applications, where it may be helpful to automatically find and extract conflictual scenes in audio recordings, both on-line and off-line.

Conflict may be defined as a *mode of interaction [where] the attainment of the goal by one party precludes its attainment by the others* (Judd 1978). The main subjects of conflict are typically finite resources or attitude differences with respect to an issue of interest. Conflicts might result in attempts to damage or limit the opportunities of others, with potentially disruptive effects on the life of any group where conflicts take place (Levine and Moreland 1998).

R. Brueckner (✉)
Machine Intelligence & Signal Processing Group, MMK, Technische Universität München,
Munich, Germany
e-mail: raymond.brueckner@web.de

B. Schuller
Department of Computing, Imperial College London, London, UK
Machine Intelligence & Signal Processing Group, MMK, Technische Universität München,
Munich, Germany
e-mail: bjoern.schuller@imperial.ac.uk

One of the earliest works of the automatic detection of “hot spots” in multi-party conversations was presented by Wrede and Shriberg (2003). They investigated whether involvement can be judged reliably by human listeners and found that despite the subjective nature of the task, raters show significant agreement in distinguishing involved from non-involved utterances. Furthermore, they found temporal trajectories of fundamental frequency (F0) and energy to be reliable acoustic cues for involvement.

Bousmalis et al. (2009) presented a survey of audio-visual cues of agreement and disagreement. The most relevant features were found to be visual cues, such as head gestures, facial and hand actions. However, also auditory cues were shown to be important, such as sighing and throat clearing, but also utterance lengths, interruptions, delays, and pauses in speech.

A semi-automatic approach aimed at detecting conflict in conversations was proposed in Pesarin et al. (2012). In that study the authors adopted a generative statistical technique based on Markov chains, capable of identifying turn-organization regularities associated with conflict.

Kim et al. (2012) instead adopted a careful selection of features and employed three different types of regression models, namely Bayesian Linear Regression, Gaussian Processes for Regression, and Support Vector Regression, for both manual and automatic diarization. They later extended their work including the detection of conflict escalation during the course of conversation (Kim et al. 2012).

The Conflict Sub-Challenge of the Interspeech 2013 Computational Paralinguistics Challenge (Schuller et al. 2012) introduced a benchmark data set to allow for the objective comparison of approaches on the detection of conflict and the prediction of conflict level. Based on this data set Grèzes et al. (2013) studied the effect of overlap for the automatic detection of conflict. They found the *overlap ratio*, i.e. the ratio of overlapping speech to non-overlapping speech, to be the single best feature for predicting the conflict level; using the *predicted* overlap ratio improved the detection performance even more. They further investigated the effect of conflict escalation or de-escalation on the prediction, but found that this did not lead to any improvements.

Given the importance of overlapping speech a number of studies have presented approaches on how to robustly estimate overlapping speech segments in multi-party conversations.

Yamamoto et al. (2006) employed Support Vector Machines (SVMs) and Support Vector Regression using microphone arrays in order to estimate the number of sound sources. Boakye et al. (2011) applied a feature analysis technique called Discriminant Capability Analysis achieving almost oracle performance on their database. Zelenák and Hernando (2011) successfully adopted a set of prosody-based long-term features as a complement to short-term spectral parameters, followed by a feature selection process according to a minimal-redundancy-maximal-relevance (mRMR) criterion.

A completely different approach was followed by Geiger et al. (2012) using a combination of features derived from convolutive non-negative sparse coding within a conventional Hidden Markov Model (HMM) system. This reduced the

overall diarization error rate by improving the overlap detection rate. Later they improved their results by combining Long Short-Term Memory (LSTM) recurrent neural networks (RNNs) with their baseline HMM system (Geiger et al. 2013). We will draw inspiration from this idea by exclusively employing neural networks to estimate overlap from acoustic features.

Neural networks have previously been utilized successfully in the field of computational paralinguistics. Stuhlsatz et al. (2011) introduced *generalized discriminant analysis* using deep neural networks (DNNs) for the task of acoustic emotion recognition and achieved highly significant improvements over previously reported baselines on a number of frequently used emotional speech corpora. Brueckner and Schuller (2012) showed good performance on the Interspeech 2012 Speaker Trait Likability Sub-Challenge (Schuller et al. 2012) adopting a moderately DNN. Using a hierarchical DNN the same authors only recently obtained the best results reported in the literature on the ComParE Social Signals Sub-Challenge, outrivaling the baseline results by 9.1 % (Brueckner and Schuller 2013). These results have further been excelled recently by adopting deep bi-directional LSTM RNNs (Brueckner and Schuller 2014).

Encouraged by this success we investigate and demonstrate how deep and hierarchical neural networks, which have become the new mainstream paradigm in automatic speech recognition over the last few years, can be leveraged to automatically classify and predict levels of conflict purely based on audio recordings. To this end, we resort to a hierarchical DNN to predict the conflict level on the Conflict Sub-Challenge of the Interspeech 2013 Computational Paralinguistics Challenge (Schuller et al. 2013). We further utilize a bi-directional Long-Short Term Memory (BLSTM) RNN to predict overlapping speech segments and demonstrate that a DNN fed with this predicted overlap achieves state-of-the-art performance. Ultimately, we show that by integrating this predicted overlap into a conversational-prosodic feature set we can improve the results even further, both for classification and regression. Using this combined feature set we obtain the best results reported so far in the literature on this data set for both the classification and the regression task.

In Sect. 19.2 we describe how to pre-train and build a DNN using Restricted Boltzmann Machines (RBMs) and how to handle real-valued input using Gaussian-Bernoulli RBMs. We further outline two recent advances to DNNs, rectified linear units and dropout. In Sect. 19.3 we briefly describe RNNs, in particular LSTM models and their bidirectional variant, and show how we will use them to generate predictions of overlapping speech segments. The underlying database used in this study is described in Sect. 19.4 and the derived feature sets are discussed in Sect. 19.5. We describe and discuss our experiments and results in Sect. 19.6 and present our conclusions in Sect. 19.7.

19.2 Deep Neural Networks

DNNs composed of multiple hidden layers were already proposed decades ago, but they were difficult to train. Neural networks are typically trained using stochastic gradient descent (SGD), e.g. using the backpropagation algorithm. However, with large initial weights the network parameters converge towards poor local minima, while with small initial weights the gradients in the lower layers vanish, making it infeasible to train networks with many hidden layers. Moreover, deep networks with many hidden layers and many hidden units in each layer tend to overfit the data sets, especially in the case of the relatively small data sets often found in computational paralinguistics.

In order to overcome these difficulties Hinton et al. (2006) proposed an efficient method to pre-train DNNs, one layer at a time, using an undirected graphical model called a *Restricted Boltzmann Machine*. One of the most interesting properties of this approach is that this pre-training is fully accomplished in an unsupervised fashion, i.e. without the need for target labels, which require (often costly) annotated data.

The principal task of this pre-training stage is to move the network parameters near to a local optimum in parameter space. The network parameters can then further be optimized by running a few iterations of SGD on the pre-trained network to discriminatively fine-tune the network to the task at hand.

19.2.1 Restricted Boltzmann Machines

An RBM is an undirected graphical model with a layer of observed (*visible*) variables and a layer of latent (*hidden*) variables, each layer forming one part of a bipartite graph; i.e., each visible unit (node) is connected to each hidden unit, but there are no intra-visible or intra-hidden connections. The graph of an RBM is depicted in Fig. 19.1.

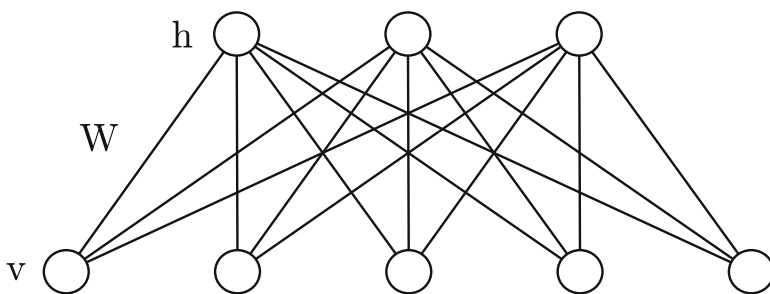


Fig. 19.1 Undirected acyclic graph of a Restricted Boltzmann Machine. W denote the weights between the visible layer v and the hidden layer h

An RBM assigns an energy to every configuration of visible and hidden state vectors, denoted v and h , respectively. For binary visible units, an RBM with V visible units and H hidden units is governed by the energy function

$$E(v, h) = - \sum_{i=1}^V \sum_{j=1}^H v_i h_j w_{ij} - \sum_{i=1}^V v_i b_i^v - \sum_{j=1}^H h_j b_j^h, \quad (19.1)$$

where v_i and h_j are the binary states of visible unit i and hidden unit j , b_i^v and b_j^h are their biases, and w_{ij} is the weight between them.

Under this energy function, the conditional probabilities for each visible and hidden unit given the others are

$$p(h_j = 1 | \mathbf{v}) = g \left(b_j^h + \sum_i v_i w_{ij} \right) \quad (19.2)$$

$$p(v_i = 1 | \mathbf{h}) = g \left(b_i^v + \sum_j h_j w_{ij} \right) \quad (19.3)$$

where

$$g(x) = \frac{1}{1 + e^{-x}} \quad (19.4)$$

is the logistic or sigmoid function.

The network assigns a probability to every possible joint configuration (v, h) via the energy function as

$$p(v, h) = \frac{e^{-E(v, h)}}{Z} = \frac{e^{-E(v, h)}}{\sum_{u, g} e^{-E(u, g)}}, \quad (19.5)$$

where Z is called the partition function. The marginal distribution of the visible units is then given as

$$p(v) = \sum_h p(v, h) \quad (19.6)$$

and the gradient of the average log-likelihood is

$$\frac{\partial \log p(v)}{\partial w_{ij}} = \langle v_i h_j \rangle_0 - \langle v_i h_j \rangle_\infty. \quad (19.7)$$

The average $\langle \cdot \rangle_0$ can be readily computed using the sample data v , but the average $\langle \cdot \rangle_\infty$ involves the normalization constant Z , which cannot generally be computed efficiently (being a sum of an exponential number of terms).

To avoid the difficulty in computing the log-likelihood gradient, Hinton (2002) proposed the *Contrastive Divergence* (CD) algorithm which approximately follows the gradient of the difference of two divergences:

$$\frac{\partial \log p(v)}{\partial w_{ij}} \approx \langle v_i h_j \rangle_0 - \langle v_i h_j \rangle_k \quad (19.8)$$

The expectation $\langle \cdot \rangle_k$ represents a distribution from running a Gibbs sampler (cf. (19.2) and (19.3)) initialized at the data for k full steps. This process is shown in Fig. 19.2. In practice, we typically choose $k = 1$. This is a rather crude approximation of the true log maximum likelihood gradient, but it works well in practice.

19.2.2 Gaussian-Bernoulli RBM

In most speech applications the input data are real-valued. A popular approach to modeling this kind of data is normalizing each input variable to fall into the range $[0;1]$ and treating it as a probability. However, even though this approach might seem appropriate at first glance, it bears serious difficulties, as it poorly models the underlying data distribution of true real-valued processes (Wang et al. 2013). In order to cope with this property we use a slightly modified RBM in the input layer, referred to as Gaussian-Bernoulli Restricted Boltzmann Machine (GBRBM). In this model we substitute the binary visible units with visible units sampled from a Gaussian distribution, where we use a modified energy function:

$$E(v, h) = \sum_{i=1}^V \frac{(v_i - b_i^v)^2}{2\sigma_i^2} - \sum_{i=1}^V \sum_{j=1}^H \frac{v_i}{\sigma_i^2} h_j w_{ij} - \sum_{j=1}^H h_j b_j^h \quad (19.9)$$

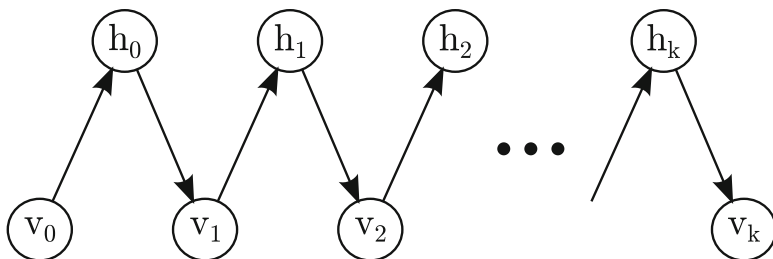


Fig. 19.2 Illustration of k -step Gibbs sampling for approximating the model data distribution, initialized at the data v_0

Under this modified energy function, the conditional probabilities for each visible and hidden unit given the others become

$$p(v_i = v | \mathbf{h}) = \mathcal{N} \left(v | b_i^v + \sum_j h_j w_{ij}, \sigma_i^2 \right) \quad (19.10)$$

$$p(h_j = 1 | \mathbf{v}) = g \left(b_j^h + \sum_i \frac{v_i}{\sigma^2} w_{ij} \right) \quad (19.11)$$

where $\mathcal{N}(\cdot | \mu, \sigma^2)$ denotes the Gaussian probability density function with mean μ and variance σ .

Training GBRBMs, even using well-defined gradients, is often difficult and takes a long time. One of the main difficulties is learning the variance parameters σ^2 , which are, unlike other parameters, constrained to be positive. Therefore, in many existing works, those parameters are set to unity (Salakhutdinov 2009), in order to simplify training. Further note that σ^2 is not the variance of the input data, but rather a probabilistic noise component added to the model.

19.2.3 Deep Belief Networks

Once the weights of an RBM have been learned, the outputs of the hidden nodes can be used as input data for training another RBM that learns a more complex representation of the input data. Proceeding this way a deep belief network (DBN) can be constructed, stacking one RBM on top of the preceding layer. Building a deep generative model one layer at a time is much more efficient than trying to learn all of the layers at once.

The parameters of this stacked network model are often close to an optimum and hence gradient descent techniques can be used to fine-tune the DBN. The limited information contained in the labels is then used to only slightly adjust the pre-trained weights in order to improve the discriminative power of the generative model.

One important property of DBNs is that their hidden states can be inferred very efficiently by a single bottom-up pass in which the top-down generative weights are used in the reverse direction. Another important property is that each time an extra layer of learned features is added to a DBN, the new DBN has a variational lower bound on the log probability of the training data that is better than the variational bound for the previous DBN, provided the extra layer is learned in the right way (Hinton et al. 2006). The weights and biases of a DBN can be used to initialize the hidden layers of a feedforward neural network, which is given an additional output layer. For a wide variety of tasks, discriminative fine-tuning of a DBN-initialized neural network gives much better performance than the same neural network initialized with small random weights (Erhan et al. 2010). Many

of the generatively learned features may be irrelevant for the discrimination task, but those that are relevant are usually much more useful than the input features, because they capture the complex higher-order statistical structure that is present in the input data.

It was shown in Erhan et al. (2010) that greedy layer-wise unsupervised pre-training is crucial in deep learning by introducing a useful prior to the supervised fine-tuning training procedure. The regularization effect is claimed to be a consequence of the pre-training procedure establishing an initialization point of the fine-tuning procedure inside a region of parameter space in which the parameters are henceforth restricted. Furthermore, overfitting can be substantially reduced if a generative model is used to find sensible features without making any use of the labels.

Strictly speaking, a DBN is a generative model consisting of several RBM layers. However, a DBN can be used to initialize the hidden layers of a standard feed-forward DNN. An additional output layer is then built on top of the DNN, typically a *softmax* layer for classification tasks or a *linear* layer for regression tasks. In the literature the terms DBN and DNN are often used interchangeably.

19.2.4 Dropout

Despite their big success, DNNs suffer from a major weakness. Due to their many non-linear hidden layers they are very expressive models and are thus very prone to the phenomenon of overfitting. This term describes the effect that a large feed-forward neural network typically performs poorly on held-out test data when trained on a small training set, as is often the case in computational paralinguistics.

Dropout was introduced by Hinton et al. (2012) as a powerful technique for reducing overfitting and for improving generalization of large neural networks. It prevents complex co-adaptations in which a hidden unit is only helpful in the context of several other specific hidden units by randomly omitting each hidden unit from the network with a probability p for each training case, so that a hidden unit cannot rely on other hidden units being present. Instead, each unit learns to detect a feature that is generally helpful for producing the correct answer given the combinatorially large variety of internal contexts in which it must operate.

This is equivalent to adding a particular type of noise to the hidden unit activations during the forward pass in training, similar to the noise added to the input units in the denoising auto-encoder approach presented in Vincent et al. (2008). However, unlike in the auto-encoder pre-training, dropout can be used in all hidden and input layers of a network and even during the fine-tuning stage of training.

An interesting way to view dropout is to consider it as a very efficient method of model averaging. Averaging the predictions of a large number of different networks is a well-known approach to reduce the test error. With neural networks this can be achieved by training many separate networks and then applying each of them to the test data, but especially with deep networks this is computationally very

expensive both during training and testing. Using the dropout technique training a huge number of neural networks in reasonable time becomes feasible. By randomly dropping out a certain percentage of hidden units, almost certainly a different network is used for each training case. Note that all of these networks share the same weights for the hidden units that are not omitted, which explains the strong regularization effect of dropout.

At test time dropout is not used and we use the “average network” with all hidden units active. However, at this stage more hidden units are active than during training. In order to compensate for this fact, during training we multiply the net input from the layer below by a factor of $\frac{1}{1-p}$ as in Dahl et al. (2013), where p is the probability of the hidden units in the lower layer being dropped out. Thus the activation y_λ of layer λ during the forward pass becomes

$$y_\lambda = g_\lambda \left(\frac{1}{1-p} y_{\lambda-1} \odot M \cdot W_\lambda + b_\lambda \right), \quad (19.12)$$

where $g_\lambda(\cdot)$ is the activation function of layer λ , W_λ and b_λ are the weights and biases of the layer, respectively, \odot denotes element-wise multiplication, and M is a binary mask matrix, whose elements are sampled i.i.d. from a Bernoulli($1-p$) distribution. The factor $\frac{1}{1-p}$, which is used during training, ensures that at test time the layer inputs are scaled correctly.

As mentioned above, dropout strongly reduces overfitting and leads to more robust models. However, applying dropout also increases the training time of the networks. The advantage is that larger networks can be used to obtain better results. This observation will be confirmed in our experiments. In the past, dropout has resulted in substantial improvements on many benchmark tasks in speech and object recognition and we will see that dropout also yields improvements in the automatic prediction of conflict levels.

19.2.5 Rectified Linear Units

The key computational unit of a deep network is a linear projection followed by a point-wise non-linearity, typically a logistic sigmoid or tanh function. Substituting this function with the recently proposed *rectified linear unit* (ReLU) has been shown to improve generalization and to make training of deep networks faster and simpler (Zeiler et al. 2013; Maas et al. 2013), and has become state of the art in speech and object recognition. A ReLU is linear when its input is positive and zero otherwise and is given by

$$g(x) = \max(x, 0) = \begin{cases} x, & \text{if } x > 0 \\ 0, & \text{else} \end{cases} \quad (19.13)$$

When a ReLu is activated above 0, its partial derivative is 1. Thus vanishing gradients do not exist along paths of active hidden units in an arbitrarily deep network. Additionally, ReLus saturate at exactly 0, which is potentially helpful when using hidden activations as input features for a classifier.

ReLus recently have been shown to yield state-of-the-art results on a number of tasks in speech recognition, for example on large vocabulary tasks, achieving lower word error rates than using a logistic network with the same topology (Zeiler et al. 2013; Maas et al. 2013; Dahl et al. 2013). To our knowledge they have not yet been applied in paralinguistics research.

19.3 Recurrent Neural Networks

A RNN is a class of neural networks whose connections between units form a directed cycle. This creates an internal state of the network, so that the network exhibits a dynamic temporal behavior and allows RNNs to process arbitrary sequences of inputs, unlike feed-forward neural networks. More precisely, given an input sequence $\mathbf{x} = (x(1), \dots, x(T))$ with $x(t) \in \mathbb{R}^D$, D being the dimensionality of vector $x(t)$, a standard RNN computes the sequence of hidden vectors $\mathbf{h} = (h(1), \dots, h(T))$ and output vectors $\mathbf{o} = (o(1), \dots, o(T))$ by recursively evaluating the following equations from time steps $t = 1, \dots, T$:

$$h(t) = g_h(W_{hx}x(t) + W_{hh}h(t-1) + b_h) \quad (19.14)$$

$$o(t) = g_o(W_{oh}h(t) + b_o) \quad (19.15)$$

where W_{hx} denotes the weight matrix from the input to the hidden layer, W_{hh} the weight matrix connecting the hidden units with each other, W_{oh} the hidden to output weight matrix, and b_h and b_o the bias vectors of the hidden and the output layer, respectively. Further, g_h and g_o are the activation functions of the hidden layer and output layer, respectively, commonly chosen to be the *sigmoid* or *tanh* function.

19.3.1 Long Short-Term Memory

RNNs are able to model a certain amount of context by using cyclic connections and can, in principle, map the entire history of previous inputs to each output. However, an analysis of the error flow in conventional recurrent neural nets reveals that they tend to suffer from the vanishing gradient problem (Hochreiter et al. 2001), i.e. the backpropagated error needed for training the network parameters either blows up or decays over time. This effect essentially limits the access to long time lags. Various attempts have been made in the past to solve this problem,

including time-delay neural networks (Waibel et al. 1989), hierarchical sequence compression (Schmidhuber 1992), or echo state networks (Jaeger 2001; Jaeger et al. 2007).

One of the most effective models, however, is the LSTM architecture (Hochreiter and Schmidhuber 1997; Gers et al. 2002). LSTM networks can be interpreted as RNNs in which the hidden neurons are replaced by special, linear memory blocks. Similar to the cyclic connections in RNNs, these memory blocks are recurrently connected. Every memory block consists of self-connected memory cells and three multiplicative units, the input, output, and forget gates, which control the information flow inside the memory block. The surrounding network can only interact with the memory cells via the gates. Since these gates allow for write, read, and reset operations within a memory block, an LSTM block can be seen as a (differentiable) memory chip in a digital computer. The overall effect of the gate units is that the LSTM memory cells can store and access information over long periods of time and thus avoid the vanishing gradient problem. For instance, as long as the input gate remains closed (corresponding to an input gate activation close to zero), the activation of the cell will not be overwritten by new inputs and can therefore be made available to the net much later in the sequence by opening the output gate. This allows to bridge long time lags between relevant inputs and outputs, which would not be possible with standard RNNs. Figure 19.3 illustrates a single LSTM memory block containing one memory cell.

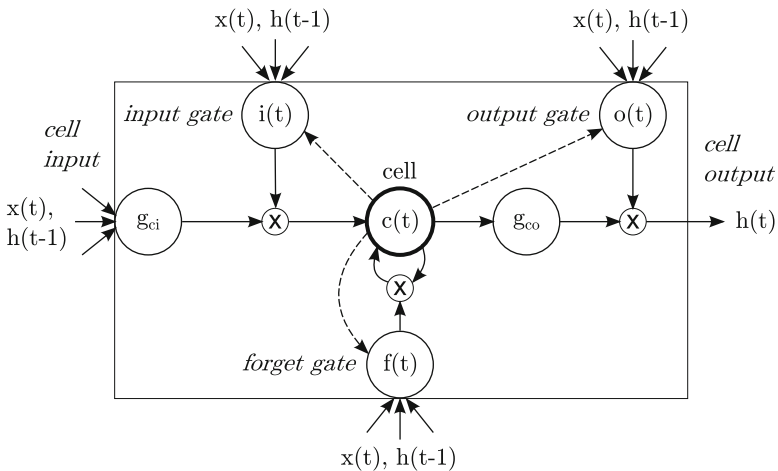


Fig. 19.3 Long Short-Term Memory block containing one memory cell. The state of cell $c(t)$ is controlled by input gate $i(t)$, its own state via a recursive connection modulated by the forget gate $f(t)$, and the cell input $g_i(x(t), h(t-1))$. The cell output emitted by the memory block, $h(t)$, is modulated by the output gate $o(t)$. The dashed arrows are termed “peephole” connections (see text for details)

The input gate activation $i(t)$ at time t is computed by applying the (non-linear) input gate activation function $g_{ig}(\cdot)$ on its inputs as

$$i(t) = g_{ig}(W_{ix} x(t) + W_{ih} h(t-1) + W_{ic} c(t-1) + b_i), \quad (19.16)$$

where W_{ix} , W_{ih} , and W_{ic} correspond to the weight matrices which project the input $x(t)$, all (hidden) memory block outputs $h(t-1)$ and the internal cell states $c(t-1)$ from the previous time step, respectively, to the input gate; b_i denotes the input gate bias. Usually, the input gate activation function g_{ig} is chosen to be the *sigmoid* function (19.4). The activation $i(t)$ of the input gate multiplies the input to all cells in the memory block, and thus determines which activity patterns are stored (added) into it. During training, the input gate learns to open ($i(t) \approx 1$) so as to store relevant inputs in the memory block, respectively close ($i(t) \approx 0$) so as to shield it from irrelevant ones.

Similarly, the activations of the forget gates $f(t)$ can be calculated as

$$f(t) = g_{fg}(W_{fx} x(t) + W_{fh} h(t-1) + W_{fc} c(t-1) + b_f), \quad (19.17)$$

where g_{fg} is commonly chosen to be the *tanh* activation function.

To determine the current state of a cell $c(t)$, we scale the previous state $c(t-1)$ by the activation of the forget gate $f(t)$ and the cell input activations g_{ci} by the activation of the input gate $i(t)$:

$$c(t) = f(t) c(t-1) + i(t) g_{ci}(W_{cx} x(t) + W_{ch} h(t-1) + b_c), \quad (19.18)$$

where g_{ci} is a logistic sigmoid function with range $[0;1]$. At $t = 0$, the cell state of a memory cell is initialized to zero, i.e. $c(0) = 0$. Subsequently, the cell accumulates a sum, discounted by the forget gate, over its input. Hence, activity circulates in the cell $c(t)$ as long as the forget gate remains open ($f(t) \approx 1$). Just as the input gate learns what to store in the memory block, the forget gate learns for how long to retain the information, and—once it is outdated—to erase it by resetting the cell state to zero. This prevents the cell state from growing to infinity and enables the memory block to store new data without undue interference from prior operations (Gers et al. 2002).

The computation of the output gate activations $o(t)$ follows the same principle as the calculation of the input gate activation. However, in this case the current cell states $c(t)$ are considered, rather than the states from the previous time step:

$$o(t) = g_{og}(W_{ox} x(t) + W_{oh} h(t-1) + W_{oc} c(t) + b_o) \quad (19.19)$$

Here, g_{og} denotes the output gate activation function, which is typically chosen to be the *sigmoid* function as for the input gate.

Finally, the memory block output $h(t)$ is computed by applying $g_{co}(\cdot)$, typically the logistic *sigmoid* squashing function (19.4) in the range $[0;1]$, to the cell state $c(t)$, multiplied by the output gate activation $o(t)$:

$$h(t) = o(t) g_{co}(c(t)) \quad (19.20)$$

The output gate activation $o(t)$ modulates the output of the current memory block and hence determines which activity patterns are transmitted to the other memory blocks, and to itself, in time step $t + 1$, effectively controlling read access to the memory block.

When first proposed, LSTMs did not have any connections from either input, output, or forget states to the cell states $c(t)$, which they are supposed to control (cf. dashed arrows in Fig. 19.3). Thus, each gate could only observe the memory block output directly, which is close to zero as long as the output gate is closed. The same problem occurs for multiple cells in a memory block—none of the gates have access to the cells they control if the output gate is closed. This lack of information may lead to sub-optimal network performance.

A solution to this problem was presented in Gers et al. (2002) with the introduction of the so-called *peepholes*: these weighted connections from the cell to all the gates in the memory block allow them to inspect the current cell state $c(t)$, even when the output gate is closed. These peephole connections were found to be necessary in order to obtain well-working network solutions.

19.3.2 Bidirectional LSTM

A shortcoming of standard RNNs is that they have access to past but not to future context. A solution to this problem are *bidirectional* RNNs (Schuster and Paliwal 1997). Here, two separate recurrent hidden layers are operating on the input sequence in opposite directions, one in forward direction, the other in backward direction. Both hidden layers are connected to the same output layer, thus providing access to long-range context in both input directions. The amount of context information that the network actually uses is learned during training, and does not have to be specified beforehand. In BLSTMs the principle of bidirectional networks and the LSTM idea are combined. Of course, by resorting to bidirectional networks true on-line processing is impossible. This may be approximated by a truncated version of BLSTM; however, in many applications it is sufficient to obtain an output at the end of an utterance so that both passes, forward and backward, can be used fully during decoding.

19.3.3 BLSTM as Overlap Prediction Generators

In the introduction we have outlined that overlap is a very informative feature for conflict level prediction and we will confirm this observation in Sect. 19.6.3. In real-world applications manual speech overlap annotations are not available and thus must be reliably estimated from the speech signal itself. We extend an approach presented in Geiger et al. (2013) by using a BLSTM model as a non-linear classifier to generate frame-wise overlap predictions. To this end, we feed the input feature vector

$$X = [x(1), \dots, x(T)] \quad (19.21)$$

into the network, where T is the total number of frames in the audio sequence, and obtain an output $y(t)$ at the *sigmoid* output layer for each time step t . Due to the BLSTM nature of our network the output $y(t)$ is dependent on both past and future input, up to time t :

$$y(t) = g_f(x(1), \dots, x(t)) + g_b(x(T), \dots, x(t)), \quad (19.22)$$

where g_f and g_b denote the function computed by the forward and backward part of the BLSTM, respectively.

For training the network, the targets are defined as

$$\hat{y}(t) = \begin{cases} 1, & \text{if } x(t) \in \text{overlap} \\ 0, & \text{else} \end{cases} \quad (19.23)$$

As in Geiger et al. (2013) the predictions $y(t)$ of the trained network are used for classification by adopting the threshold θ as follows:

$$c(t) = \begin{cases} 1, & \text{if } y(t) \geq \theta \\ 0, & \text{if } y(t) < \theta \end{cases} \quad (19.24)$$

The threshold θ can be varied in order to select a specific operating point with a different trade-off between precision and recall.

19.4 Database

The experiments and results presented in this study are based on the *SSPNet Conflict Corpus* (SC^2) (Kim et al. 2012), which was also used in the Conflict Sub-Challenge of the Interspeech 2013 Computational Paralinguistics Challenge (Schuller et al. 2013). It contains 1,430 clips, each 30 s long, extracted from the Canal9 Corpus (Vinciarelli et al. 2009), a publicly available corpus of broadcasted Swiss

political debates in French language. It includes a rich set of socially relevant annotations, such as turn-taking (who speaks when and how much), agreement and disagreement between participants, and the role played by the people involved. Each debate includes one moderator and two coalitions opposing one another on the issues of the day. A subset of this database, composed of 45 debates with four guests (two guests in each group) plus one moderator, has been annotated in terms of conflict level. The debates have been segmented into 30-s long uniform, non-overlapping clips, assuming that the levels of conflict are stationary within the time period.

The SC^2 corpus includes 138 subjects in total, 23 females (1 moderator and 22 participants) and 133 males (3 moderators and 120 participants). The clips were annotated in two ways in terms of their conflict level by approximately 550 assessors recruited via Amazon Mechanical Turk: First, a continuous conflict score in the range $[-10, +10]$ was assigned to each clip, which allows to perform a straightforward regression task (*score*). Second, based on these score labels each clip was classified to be either of *high* conflict or *low* conflict, depending if the score value assigned to it being ≥ 0 or < 0 , respectively, thus giving rise to a classification task (*class*).

As several subjects occur in debates with different moderators, a truly speaker-independent partitioning of the data is not possible. Since all participants (apart from the moderators) do not occur more than a couple of times (most of them only once), the following strategy was followed to reduce speaker dependence to a minimum: All broadcasts with the female moderator (speaker no. 50) were assigned to the training set. The development set consists of all broadcasts moderated by the (male) speaker no. 153 and the test set comprises the rest of the broadcasts, containing all remaining male moderators. This further ensures that the development and test sets are similar in case the gender of the moderator should have an influence.

The resulting distribution of the data is shown in Table 19.1 along with the respective binary class labels. Histograms of the continuous score ratings over the partitions are depicted in Fig. 19.4.

Table 19.1 Partitioning of the SSPNet Conflict Corpus into train, devel(opment), and test sets for binary classification (“low” $\equiv [10, 0]$, “high” $\equiv [0, +10]$)

#	train	devel	test	Σ
Low	471	127	226	824
High	322	113	171	606
Σ	793	240	397	1,430

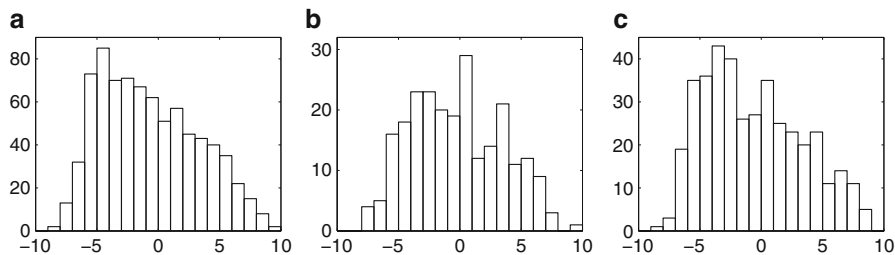


Fig. 19.4 Level of conflict (in the range $[-10, +10]$) histograms for the Challenge partitions of the SSPNet Conflict Corpus. (a) train, (b) dev, (c) test

Table 19.2 Low-level descriptors (LLD) of features set I

<i>Energy-related LLD</i>
Sum of auditory spectrum (loudness)
Sum of RASTA-style filtered auditory spectrum
RMS energy
Zero-crossing rate
<i>Spectral LLD</i>
RASTA-style auditory spectrum, bands 1–26 (0–8 kHz)
MFCC 1–14
Spectral energy 250–650 Hz, 1k to 4kHz
Spectral roll off point 0.25, 0.50, 0.75, 0.90
Spectral flux, entropy, variance, skewness, kurtosis,
Slope, psychoacoustic sharpness, harmonicity
<i>Voicing-related LLD</i>
F0 by SHS + Viterbi smoothing, probability of voicing
logarithmic HNR, spectral harmonicity
Psychoacoustic spectral sharpness
Jitter (local, delta), shimmer (local)

19.5 Feature Sets

In this paper we investigate the use of three fundamentally different sets of features: **Feature set I**, also referred to as *baseline feature set*, is a supra-segmental feature set, in our case one segment per utterance. This is an approach often followed in emotion recognition and paralinguistic analysis. For this purpose we adopt the baseline acoustic feature set used in the ComParE Conflict Sub-Challenge (Schuller et al. 2013). We use the open-source feature extractor openSMILE (Eyben et al. 2010), developed by the Technical University of Munich, to extract the features on a per-chunk level. The extracted features are the ones already proposed in the Interspeech 2012 Speaker Trait Challenge (Schuller et al. 2012) with some additional modifications and features. First, so-called low-level descriptors (LLD) are extracted, which are listed in Table 19.2.

Table 19.3 Applied functionals of feature set I

<i>Functionals applied to LLD/Δ LLD</i>
Quartiles 1–3, 3 inter-quartile ranges
1 % percentile (\approx min), 99 % percentile (\approx max)
Position of min/max
Percentile range 1–99 %
Arithmetic mean, ^a root quadratic mean
Contour centroid, flatness
Standard deviation, skewness, kurtosis
Rel. duration LLD is above/below 25/50/75/90 % range
Rel. duration LLD is rising/falling
Rel. duration LLD has positive/negative curvature ^b
Gain of linear prediction (LP), LP Coefficients 1–5
Mean, max, min, std. dev. of segment length ^c
<i>Functionals applied to LLD only</i>
Mean of peak distances
Standard deviation of peak distances
Mean value of peaks
Mean value of peaks—arithmetic mean
Mean/std.dev. of rising/falling slopes
Mean/std.dev. of inter maxima distances
Amplitude mean of maxima/minima
Amplitude range of maxima
Linear regression slope, offset, quadratic error
Quadratic regression a , b , offset, quadratic error
Percentage of non-zero frames ^d

^a Arithmetic mean of LLD/positive Δ LLD

^b Only applied to voice related LLD

^c Not applied to voice related LLD except F0

^d Only applied to F0

The set includes energy, spectral, cepstral (MFCC) and voicing related low-level descriptors (LLDs) as well as a few LLDs including logarithmic harmonic-to-noise ratio (HNR), spectral harmonicity, and psychoacoustic spectral sharpness.

On these LLD a number functionals are applied in order to extract higher level statistics. These are listed in Table 19.3. Altogether, the final feature set I contains 6,373 features.

Feature set II is extracted on a frame-wise basis and in order to keep the amount of features under control only a relatively small set of descriptors is calculated for each frame. Using the openSMILE toolkit again, we extract low-level descriptors (LLDs) and functionals every 10 ms adopting a frame size of 20 ms. Specifically, we compute frame-wise logarithmic energy and Mel-frequency cepstral coefficients (MFCC) 1–12 along with their first and second order delta (Δ) regression coefficients as typically used in automatic speech recognition. These are augmented by voicing probability, HNR, F0 and zero-crossing rate, as well as their

respective first order Δ . Then, for each frame-wise LLD the arithmetic mean and standard deviation across the frame itself and eight of its neighbouring frames (four before and four after) are appended. This results in $47 \times 3 = 141$ descriptors per frame. This feature set has been used with great success on the ComParE Vocalization Sub-Challenge data in previous work (Brueckner and Schuller 2013, 2014).

Finally, **feature set III** is a modification of the feature set proposed by Kim et al. (2012) and consists of a conversational and a prosodic part: The first corresponds to turn duration statistics, namely mean, median, maximum, variance and minimum of speaker turns duration in each audio clip as well as the number of turns. It further includes total speaking time statistics, i.e. mean, median, maximum, variance and minimum of the total speaking time for individual speakers in the clips as well as the number of people speaking. We finally add the overlap ratio described in Sect. 19.6.3 and conclude the conversational feature part with the turn keeping/turn stealing ratio in the clip, defined as the ratio between the number of times a speaker change happens and the number of times a speaker change does not happen after an overlap. This conversational part is complemented by prosodic features including clip-based statistics: mean, median, standard deviation, maximum, minimum and quantiles (0.01, 0.25, 0.75 and 0.99) of pitch and intensity statistics obtained from the entire clip, with the pitch and intensity LLDs being identical to the ones in feature set I. These general prosodic features are complemented by speaker turn-based statistics, i.e. mean, median and standard deviation of pitch and intensity obtained over individual speaker turns (similarly to the clip-base statistics). Note that the statistics above are estimated not only on single-talker segments, but also over overlapping speech segments. Altogether, feature set III contains 38 features.

19.6 Experiments and Results

19.6.1 Challenge Baseline Comparison

For the ComParE Conflict Sub-Challenge baseline, results were supplied by the challenge organizers (Schuller et al. 2013) for both a classification task, where each utterance is classified as being either non-conflictual (*low*) or conflictual (*high*), and a regression task, trying to predict the rater's score value in the range $[-10; +10]$.

For the classification task the primary evaluation measure was chosen to be the unweighted average recall (UAR), which had been used since the Interspeech's 2009 Emotion Challenge (Schuller et al. 2011). The motivation to use the unweighted rather than the weighted average recall ("conventional" accuracy) is that it is also meaningful for highly unbalanced distributions of instances among classes.

For the regression task the Pearson correlation coefficient was selected as evaluation criterion, being a measure for the prediction quality of continuous-valued score labels. It is defined as follows:

$$CC = \frac{1}{N\sigma\hat{\sigma}} \sum_{n=1}^N (\hat{y}_n - \hat{\mu})(y_n - \mu), \quad (19.25)$$

where N is the number of samples, \hat{y}_n represents the prediction for the n -th data point with target value y_n . Further, $(\hat{\mu}, \hat{\sigma}^2)$ and (μ, σ^2) are the corresponding means and variances for the predictions and targets across the data set, respectively.

The Challenge’s baseline results were computed adopting a linear kernel SVM trained using Sequential Minimal Optimization (SMO) (Bishop 2006). The SVM complexity parameter $C \in \{10^{-3}, 10^{-2}, 10^{-1}, 1\}$ which achieved the best UAR on the development set was chosen for the reference results and logistic models were fit to the SVM hyperplane distance based on the training set to obtain (pseudo) class posteriors. In the Conflict Sub-Challenge, the simple baseline already delivers a remarkable performance on the binary classification and real-valued regression tasks as shown in Table 19.4.

19.6.2 Baseline Features (Feature Set I)

In a first experiment we compared the traditionally used sigmoid hidden units to ReLU hidden units, the latter being used in combination with dropout training. For this purpose, we trained standard one-hidden layer neural networks, usually termed multi-layer perceptrons (MLP), feeding the supra-segmental feature set I as input. This comprised all 6,373 features as described in Sect. 19.5. All input features were normalized to zero mean and unit variance, with the means and standard deviations being computed on the training set. After normalization the feature set is approximately Gaussian distributed, which turns out to be beneficial when using GRBMs.

Using these normalized data as input we trained the MLPs varying the sizes of the hidden layers and using either traditional sigmoid units or ReLUs trained

Table 19.4 Challenge baseline results

[%]	C	devel	test
CC [score]	0.001	81.6	82.6
UAR [class]	0.1	79.1	80.8

C : complexity parameter in SVM training (tuned on development set).
devel: result on development set by training on training set. test: result on test set by training on the training and development sets

with the dropout technique. Further, in order to check whether pre-training helps to combat overfitting we trained two sets of networks, one with and one without pre-training. In all cases, the weights were initialized with values sampled from a uniform distribution in the range

$$\left[-4 \cdot \sqrt{\frac{6}{N_{\text{in}} \cdot N_{\text{out}}}}, +4 \cdot \sqrt{\frac{6}{N_{\text{in}} \cdot N_{\text{out}}}} \right], \quad (19.26)$$

where N_{in} is the number of input units and N_{out} the number of output units.

The networks were trained on the training set using standard SGD with momentum and early stopping determined on the development set. As the cost function we used cross entropy (CE) for the classification task and the mean squared error (MSE) for the regression task. As soon as the cost function started to raise on the development set the training was stopped.

Figure 19.5 shows the best test set results obtained on the regression task for different network sizes.

Several conclusions can be drawn from this figure: first, the ReLU networks outperform the sigmoid networks for all hidden layer sizes, except for small networks with hidden unit sizes < 512 . This might be due to the strong regularization effect of dropout that surmounts the regularization effect of pre-training. Second, pre-training helps in all cases, regardless of whether sigmoid units or rectified linear

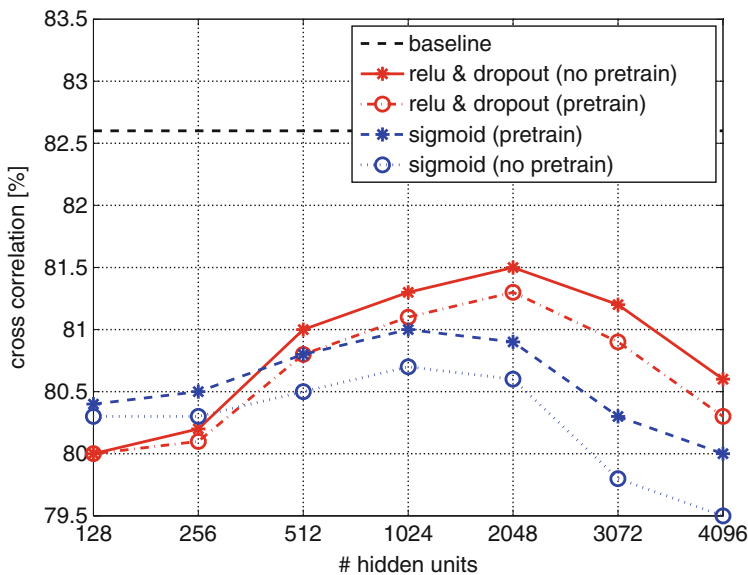


Fig. 19.5 Regression task: test set results based on the baseline feature set I for a one-hidden layer MLP for varying hidden layer sizes. Shown are the graphs for networks trained with rectified linear units and dropout vs. sigmoid hidden units and with vs. without pre-training the networks

Table 19.5 Results on the classification (class) and regression task (score) using feature set I varying the number of hidden layers

[%]	Class		Score	
	devel	test	devel	test
MLP (6373-2048-1)	78.3	79.8	80.9	81.5
DNN (6373-1024-1024-1)	79.7	80.9	81.5	82.1
DNN (6373-1024-1024-1024-1)	78.9	80.2	81.1	81.8

Shown are the best results obtained on the development set (devel) and on the test set (test). The percentages reported denote UAR for the classification task and CC for the regression task

units were employed. Sigmoid units seem to profit more from pre-training than ReLus, which again may be attributed to the effect of dropout. Pre-training further seems to be more helpful for bigger hidden layer sizes than for smaller ones. This observation can be explained by the regularization effect of pre-training, which has a bigger impact on big networks with their high number of parameters. Third, due to the adoption of dropout, ReLu networks show best performance for bigger network sizes than sigmoid networks. This effect was already explained in Sect. 19.2.4. Last but not least, even though the performance of the best network of $CC = 81.5\%$ indicates rather good modelling power, it still underperforms the baseline system ($CC = 82.6\%$).

Based on these insights we trained a number of DNNs varying the number of hidden layers and using the training procedure described above. In particular, we used ReLus trained adopting the dropout technique. The best results are reported in Table 19.5. For this feature set and setup we achieve best results for a DNN with two hidden layers with 1,024 hidden units for each layer.

19.6.3 *Overlap Ratio*

As outlined in the introduction a number of studies have indicated that overlap and features derived from it serve as very good indicators for the conflict level of discourse. Hence, we were interested in how this feature can be leveraged by deep, hierarchical networks. The SC^2 corpus is equipped with hand-labeled meta-data containing speaker-turn information as well as overlapping speech segment annotation. This allows us to compute the true overlap ratio, i.e. the relative percentage of overlap with respect to the utterance length; in the following, we refer to this reference value as the *oracle overlap ratio*.

We first examined this ratio as a single feature and trained several feed-forward neural networks for 1–3 hidden layers and varying the number of ReLu hidden units from 2 to 128. As before we trained the networks using SGD on the training set adopting dropout and early stopping.

Table 19.6 Results for the **oracle** overlap ratio as a single feature (top) and alongside baseline feature set I (bottom) varying the number of hidden layers on the classification (class) and regression task (score)

[%]	Class		Score	
	devel	test	devel	test
MLP (1-32-1)	75.5	76.7	78.6	79.1
DNN (1-32-32-1)	76.4	77.2	79.2	79.6
DNN (1-32-32-32-1)	75.4	76.7	78.9	79.2
MLP (6374-2048-1)	80.6	81.8	81.3	81.8
DNN (6374-1024-1024-1)	81.4	82.5	81.9	82.5
DNN (6374-1024-1024-1024-1)	80.5	81.9	81.3	81.7

Shown are the best results obtained on the development set (devel) and on the test set (test). The percentages reported denote UAR for the classification task and CC for the regression task

In order to see if the overlap ratio adds additional information to the baseline features, we also trained a series of networks on the overlap ratio being added to the feature set I, varying the number of hidden units from 64 to 4,096.

Table 19.6 shows the results for the respective optimal number of hidden units, $H_{\text{hid}} = 32$ for oracle overlap ratio as a single feature and $H_{\text{hid}} = 1,024$ for oracle overlap ratio alongside feature set I, trained with a different number of hidden layers.

Even when used alone, the oracle overlap ratio predicts conflict class and level of conflict with high accuracy. When combined with feature set I it yields even better performance than feature set I alone. It further outperforms the baseline classification task results and stays on par with the baseline regression task results.

As described in Sect. 19.3.3, in a real-world application overlap must be reliably estimated from the speech signal itself. To this end, we used a BLSTM classifier to generate predictions of overlapping speech. We used the frame-wise feature set II (cf. Sect. 19.5) to train our network, which contained one recurrent hidden layer consisting of 50 BLSTM memory blocks. We trained this network using the backpropagation through time (BPTT) algorithm with a learning rate of 10^{-5} and momentum 0.9. We initialized the weights sampling from a uniform distribution in the range $[-0.1;0.1]$. We further added noise sampled from a zero-mean Gaussian distribution with $\sigma = 0.1$ to the network inputs. Training was run until the cross-entropy (CE) error measure did not show an improvement on the development set for 10 epochs.

We then used the trained BLSTM network to classify each frame as either overlapping or non-overlapping speech and computed the predicted overlap ratio for the full audio clip from the resulting classification result. Using this predicted overlap ratio either as a single feature or alongside the baseline feature set I, just as before, we obtained the results shown in Table 19.7. Using the predicted overlap ratio as a single feature a DNN with two hidden layers, each containing 32 ReLus

Table 19.7 Results for the **predicted** overlap ratio as a single feature (top) and alongside baseline feature set I (bottom) varying the number of hidden layers on the classification (class) and regression task (score)

[%]	Class		Score	
	devel	test	devel	test
MLP (1-32-1)	80.5	82.3	81.3	82.1
DNN (1-32-32-1)	80.8	82.9	81.9	82.7
DNN (1-32-32-32-1)	80.5	82.5	81.2	81.9
MLP (6374-2048-1)	82.0	83.2	82.0	82.6
DNN (6374-1024-1024-1)	82.3	83.7	82.5	83.2
DNN (6374-1024-1024-1024-1)	82.0	83.1	82.1	82.6

Shown are the best results obtained on the development set (devel) and on the test set (test). The percentages reported denote UAR for the classification task and CC for the regression task

trained with dropout, yielded the best results. Alongside feature set I, again a DNN with two hidden layers, but $H_{\text{hid}} = 1,024$, was found to be optimal.

Our finding is that using a prediction of overlapping speech as the sole feature of a DNN classifier produces models with better performance than the baseline experiment, confirming the results in Grèzes et al. (2013). Even for the DNN regressor we obtained slightly higher cross-correlation results. When added to feature set I the predicted overlap ratio further improves results on both tasks.

19.6.4 Conversational-Prosodic Features

Encouraged by the high impact of the overlap ratio feature we drew inspiration from Kim et al. (2012) and investigated the performance of DNNs on feature set III. As described in Sect. 19.6.4 this feature set contains prosodic as well as conversational features, including the overlap ratio examined in the previous section as well as speaker-turn based features. In order to facilitate the experiments we computed the speaker-turn features from the manual annotation provided with the data set. However, as it gave better results, we used the predicted overlap ratio instead of the oracle overlap ratio, as described in Sect. 19.6.3.

As before we trained a number of ReLu networks with dropout for different network topologies, varying the number of hidden units as well as the number of layers. All networks were trained on the training set with SGD and momentum, stopping training as soon as the cost function started to raise on the development set. Again, we used CE as the cost function for the classification task and MSE for the regression task.

The results for the best performing network varying the number of hidden layers are shown in Table 19.8.

Table 19.8 Results for feature set III varying the number of hidden layers on the classification (class) and regression task (score)

[%]	Class		Score	
	devel	test	devel	test
MLP (38-512-1)	82.5	83.8	82.2	83.2
DNN (38-512-512-1)	83.1	84.3	83.0	83.8
DNN (38-512-512-512-1)	82.6	84.0	82.7	83.4
Challenge baseline (Schuller et al. 2012)	79.1	80.8	81.6	82.6
Räsänen and Pohjalainen (2013)	–	83.9	–	–
Grèzes et al. (2013)	–	83.1	–	–

Shown are the best results obtained on the development set (devel) and on the test set (test). The percentages reported denote UAR for the classification task and CC for the regression task. For comparison the baseline results and the highest published competition results of the Conflict Sub-Challenge are shown as well

The results reveal that using the conversational-prosodic feature set III we obtain best results for a two-layer DNN with each hidden layer containing 512 ReLu units. This value is smaller than the 1,024 hidden units per layer from the results above and is due to the smaller number of features in feature set III.

On the classification task we achieve a $UAR = 84.3\%$, which outperforms the baseline result by 3.5% and the best result in the Conflict Sub-Challenge reported by Räsänen and Pohjalainen (2013) by 0.4% . On the regression task the relative improvements are smaller, still raising the benchmark of the Challenge correlation coefficient $82.6\text{--}83.8\%$, measured on the test set. To our knowledge, these numbers represent the best results reported in the literature to date.

19.7 Conclusions

This study presents an approach for the detection of conflict during spontaneous, multi-party conversations employing deep, hierarchical neural networks. The experiments have been performed on the *SSPNet Conflict Corpus* (SC^2) (Kim et al. 2012), which was also used in the Conflict Sub-Challenge of the Interspeech 2013 Computational Paralinguistics Challenge (Schuller et al. 2013). Investigating different feature sets we show that replacing the traditionally used sigmoid hidden units with rectified linear units and pre-training the networks using RBMs—combined with dropout as an advanced regularization method—improves performance and allows us to obtain results almost as good as the already high baseline results reported in the challenge.

We then show that the use of the oracle overlap ratio, i.e. the ratio of overlapping speech to non-overlapping speech obtained from manual segmentation, as a single feature already allows to predict the conflict level to a good degree. Combined

with the baseline feature set it further improves upon the results. Using BLSTM RNNs based on frame-wise features, we predict the overlap ratio and show that this prediction outperforms the reference feature set, even more when combined with it. This confirms the findings in Grèzes et al. (2013).

Encouraged by these results we add the predicted overlap ratio to a carefully constructed feature set, which combines conversational and prosodic features, and train DNN models on it. Our best models outperform the Conflict Sub-Challenge baseline and the best challenge contributions on both the classification task, predicting if an utterance is conflict or non-conflict, and the regression task, predicting the level of conflict in the range $[-10.0, +10.0]$.

Adopting a DNN architecture with two hidden layers of rectified linear units, pre-trained and fine-tuned on the conversational-prosodic feature set using the dropout technique, we outperform all previously reported results. On the classification task we achieve a UAR = 84.3 %, which improves the baseline result by 3.5 % and the best result reported for the Conflict Sub-Challenge by Räsänen and Pohjalainen (2013) by 0.4 %. On the regression task the relative improvements are smaller, still raising the benchmark of the Challenge correlation coefficient of 82.6–83.8 %, measured on the test set. It is interesting to note that while for the baseline the best cross-correlation percentage is higher than the best UAR percentage, this is different in our study. It should be noted, however, that the best results of the baseline were obtained for different SVM complexity parameters C , as shown in Table 19.4.

These results are very promising; however, they are partly based on the manual speaker-turn annotations provided with the database. For a fully automatic system we intend to continue our study of conflict detection by deploying an automatic speaker turn detection and diarization system.

Furthermore, the feature set that led to the best results reported in this study was hand-tuned and while showing its high potential the selection of features might still be sub-optimal. We therefore intend to complement our current approach by adopting some sort of feature selection algorithm, e.g. as suggested in Räsänen and Pohjalainen (2013). This way we hope to further improve upon the current results.

Acknowledgements The research presented in this publication was conducted while the first author was employed by Nuance Communications Deutschland GmbH.

References

- Bishop CM (2006) Pattern recognition and machine learning. Springer, Berlin
- Boakye K, Vinyals O, Friedland G (2011) Improved overlapped speech handling for speaker diarization. In: Proceedings of interspeech, ISCA, Florence, Aug 2011, pp 941–944
- Bousmalis K, Mehu M, Pantic M (2009) Spotting agreement and disagreement: a survey of nonverbal audiovisual cues and tools. In: Proceedings of the 3rd international conference on affective computing and intelligent interaction and workshops, AII 2009, vol 2. IEEE Computer Society Press, Los Alamitos
- Breckner R, Schuller B (2012) Likability classification - a not so deep neural network approach. In: Proceedings of interspeech, Portland, OR, Sep 2012

- Brueckner R, Schuller B (2013) Hierarchical neural networks and enhanced class posteriors for social signal classification. In: Proceedings of ASRU, IEEE, Olomouc, Dec 2013, pp 361–364
- Brueckner R, Schuller B (2014) Social signal classification using deep BLSTM recurrent neural networks. In: Proceedings of ICASSP, IEEE, Florence, May 2014
- Dahl G, Sainath T, Hinton G (2013) Improving deep neural networks for LVCSR using rectified linear units and dropout. In: Proceedings of ICASSP, IEEE, Vancouver, May 2013, pp 8609–8613
- Erhan D, Bengio Y, Courville A, Vincent PAMP, Bengio S (2010) Why does unsupervised pre-training help deep learning? *J Mach Learn Res* 11:625–660
- Eyben F, Wöllmer M, Schuller B (2010) openSMILE – the Munich versatile and fast open-source audio feature extractor. In: Proceedings of ACM multimedia, MM 2010, ACM, Florence, Oct 2010. ACM, New York, pp 1459–1462 (acceptance rate short paper: about 30 %)
- Geiger JT, Vipplera R, Bozonnet S, Evans N, Schuller B, Rigoll G (2012) Convolutional non-negative sparse coding and new features for speech overlap handling in speaker diarization. In: Proceedings of interspeech, Portland, OR, Sept 2012
- Geiger J, Eyben F, Schuller B, Rigoll G (2013) Detecting overlapping speech with long short-term memory recurrent neural networks. In: Proceedings of interspeech, ISCA, Lyon, Aug 2013, pp 1668–1672
- Gers F, Schraudolph N, Schmidhuber J (2002) Learning precise timing with LSTM recurrent networks. *J Mach Learn Res* 3:115–143
- Grèzes F, Richards J, Rosenberg A (2013) Let me finish: automatic conflict detection using speaker overlap. In: Proceedings of interspeech, ISCA, Lyon, Aug 2013, pp 200–204
- Hinton GE (2002) Training products of experts by minimizing contrastive divergence. *Neural Comput* 14(8):1771–1800
- Hinton GE, Osindero S, Teh YW (2006) A fast learning algorithm for deep belief nets. *Neural Comput* 18(7):1527–1554
- Hinton G, Srivastava N, Krizhevsky A, Sutskever I, Salakhutdinov R (2012) Improving neural networks by preventing co-adaptation of feature detectors. *CoRR*. abs/1207.0580
- Hochreiter S, Schmidhuber J (1997) Long short-term memory. *Neural Comput* 9:1735–1780
- Hochreiter S, Bengio Y, Frasconi P, Schmidhuber J (2001) Gradient flow in recurrent nets: the difficulty of learning long-term dependencies. In: Kremer SC, Kolen JF (eds) *A field guide to dynamical recurrent neural networks*. IEEE Press, New York
- Jaeger H (2001) The “echo state” approach to analysing and training recurrent neural networks. GMD Report 148, GMD - German National Research Institute for Computer Science
- Jaeger H, Maass W, Principe JC (2007) Special issue on echo state networks and liquid state machines. *Neural Netw* 20(3):287–289
- Judd CM (1978) Cognitive effects of attitude conflict resolution. *J Conflict Resolut* 22(3):483–498
- Kim S, Filippone M, Valente F, Vinciarelli A (2012) Predicting the conflict level in television political debates: an approach based on crowdsourcing, nonverbal communication and Gaussian processes. In: Babaguchi N, Aizawa K, Smith JR, Satoh S, Plagemann T, Hua XS, Yan R (eds) *Proceedings of ACM international conference on multimedia*, Nara. ACM, New York, pp 793–796
- Kim S, Yella SH, Valente F (2012) Automatic detection of conflict escalation in spoken conversations. In: Proceedings of interspeech, ISCA, Portland, OR, Sept 2012
- Levine JM, Moreland RL (1998) Small groups. In: Gilbert D, Lindzey G (eds) *The handbook of social psychology*, vol 2. Oxford University Press, Oxford
- Maas A, Hannun A, Ng A (2013) Rectifier nonlinearities improve neural network acoustic models. In: Proceedings of ICML workshop on deep learning for audio, speech, and language processing, WDLASL, Atlanta, GA, Jun 2013
- Pesarin A, Cristani M, Murino V, Vinciarelli A (2012) Conversation analysis at work: detection of conflict in competitive discussions through automatic turn-organization analysis. *Cogn Process* 13(2):533–540

- Räsänen O, Pohjalainen J (2013) Random subset feature selection in automatic recognition of developmental disorders, affective states, and level of conflict from speech. In: Proceedings of interspeech, Lyon, Aug 2013, pp 210–214
- Salakhutdinov R (2009) Learning deep generative models. Ph.D. thesis, University of Toronto
- Schmidhuber J (1992) Learning complex extended sequences using the principle of history compression. *Neural Comput* 4(2):234–242
- Schuller B (2012) The computational paralinguistics challenge. *IEEE Signal Process Mag* 29(4):97–101
- Schuller B, Batliner A (2013) Computational paralinguistics: emotion, affect and personality in speech and language processing. Wiley, New York
- Schuller B, Batliner A, Steidl S, Seppi D (2011) Recognising realistic emotions and affect in speech: state of the art and lessons learnt from the first challenge. *Speech Commun* 53(9/10):1062–1087 [Special Issue on Sensing Emotion and Affect – Facing Realism in Speech Processing]
- Schuller B, Steidl S, Batliner A, Nöth E, Vinciarelli A, Burkhardt A, van Son R, Wenginger F, Eyben F, Bocklet T, Mohammadi G, Weiss B (2012) The interspeech 2012 speaker trait challenge. In: Proceedings of interspeech, Portland, OR
- Schuller B, Steidl S, Batliner A, Vinciarelli A, Scherer K, Ringeval F, Chetouani M, Wenginger F, Eyben F, Marchi E, Mortillaro M, Salamin H, Polychroniou A, Valente F, Kim S (2013) The interspeech 2013 computational paralinguistics challenge: social signals, conflict, emotion, autism. In: Proceedings of interspeech, Lyon, Aug 2013
- Schuster M, Paliwal K (1997) Bidirectional recurrent neural networks. *IEEE Trans Signal Process* 45:2673–2681
- Stuhlsatz A, Meyer C, Eyben F, Zielke T, Meier G, Schuller B (2011) Deep neural networks for acoustic emotion recognition: raising the benchmarks. In: Proceedings of ICASSP, Prague, pp 5688–5691
- Vincent P, Larochelle H, Bengio Y, Manzagol PA (2008) Extracting and composing robust features with denoising autoencoders. In: Proceedings of ICML, New York, NY, 2008, pp 1096–1103
- Vinciarelli A, Dielmann A, Favre S, Salamin H (2009) Canal9: a database of political debates for analysis of social interactions. In: Proceedings of the international conference on affective computing and intelligent interaction, Sept 2009, pp 1–4
- Vinciarelli A, Pantic M, Bourlard H (2009) Social signal processing: survey of an emerging domain. *Image Vis Comput* 27(12):1743–1759
- Waibel A, Hanazawa T, Hinton G, Shikano K, Lang K (1989) Phoneme recognition using time-delay neural networks. *IEEE Trans Acoust Speech Signal Process* 37(3):328–339
- Wang N, Melchior J, Wiskott L (2012) An analysis of Gaussian-binary restricted Boltzmann machines for natural images. In: Proceedings of ESANN, Bruges, Apr 2012, pp 287–292
- Wrede B, Shriberg E (2003) Spotting “hot spots” in meetings: human judgments and prosodic cues. In: Proceedings of Eurospeech, ISCA, Geneva, Sept 2003, pp 2805–2808
- Yamamoto K, Asano F, Yamada T, Kitawaki N (2006) Detection of overlapping speech in meetings using support vector machines and support vector regression. *IEICE Trans Fundam Electron Commun Comput Sci* 89-A(8):2158–2165
- Zeiler M, Ranzato M, Monga R, Mao M, Yang K, Le QV, Nguyen P, Senior A, Vanhoucke V, Dean J, Hinton G (2013) On rectified linear units for speech processing. In: ICASSP, IEEE, Vancouver, May 2013, pp 3517–3521
- Zelenák M, Hernando J (2011) The detection of overlapping speech with prosodic features for speaker diarization. In: Proceedings of interspeech, ISCA, Florence, Aug 2011, pp 1041–1044

Chapter 20

Conflict Cues in Call Center Interactions

Maria Koutsombogera, Dimitrios Galanis, Maria Teresa Riviello,
Nikos Tseres, Sotiris Karabetos, Anna Esposito, and Harris Papageorgiou

20.1 Introduction

This work explores the multimodal nature of conflicts occurring in call center dyadic interactions from a multidisciplinary perspective, including paralinguistic and conversation analysis cues. It is based on a Greek phone company's corpus which consists of conversations revealing how customers interact with call center operators to express concerns in terms of efficiency, provided services, argumentation and negotiation issues, and expressed emotions during the interactional exchanges. In this setting, we consider conflictual the interaction between speakers that pursue individual and at times incompatible goals (Allwood 2007). Conflict holds between the beliefs and goals of two individuals involved in the conversation, which, namely, represent the consumer and service provider roles. Instantiations of emotional behavior related to conflict and to the speakers' roles are, e.g., the expression of frustration or anger on the customers' side, stress detection, disappointment mitigation, and failure in providing the requested services or solutions on the operators' side.

Understanding conflict cues and being able to model them enables the development of technologies that can deal effectively and efficiently with the complexity of

M. Koutsombogera (✉) • D. Galanis • S. Karabetos • H. Papageorgiou
Institute for Language and Speech Processing, 'Athena' Research Centre,
Artemidos 6 & Epidavrou, 15125 Athens, Greece
e-mail: mkouts@ilsp.gr

M.T. Riviello • A. Esposito
Department of Psychology, Second University of Naples, Caserta, Italy

International Institute for Advanced Scientific Studies (IIASS), Vietri sul Mare, Italy

N. Tseres
Department of Linguistics, University of Athens, University Campus, 15784 Athens, Greece

expressions and understandings of human behavior and respond to a growing need of applications related to human behavior analysis (Narayanan and Gregoriou 2013). For example, the telecommunications industry suffers from approximately 30 % of churn rate, while it is of great importance to keep a high percentage of customer retention (Jahromi et al. 2010). In this context, it is crucial to detect emotional traits providing information about the speakers' intentions and emotional states. These traits are multimodal, in the sense that they can be inferred from the paralinguistic properties of the utterances, from the structural units of the interaction and their flow, as well as from the linguistic content. At the same time, the perception of the speakers' emotional states and the definition of the appropriate values describing them are more than a trivial issue. This work focuses on most of the aforementioned aspects, keeping aside for the time being the investigation of the linguistic content.

For the needs of our task, we extracted emotionally colored units from our call center corpus. These units were in turn labeled by a human annotator with the values of positive or negative. A large number of speech and other context-related features were extracted for each unit to train mathematical models that can be used to predict the label of an unseen emotional unit. A subset of the corpus was further annotated in terms of turn management types, and the resulting annotations were then associated to the emotional labels. In the next section, we describe the collected corpus and the procedures applied to annotate it as well as a small-scale experiment aimed to assess the perception of emotions from conversations extracted from this domain. In Sect. 20.3 we describe the automatic feature extraction process and the machine learning models proposed for the automatic classification tasks together with the obtained results. Section 20.4 is dedicated to the turn management annotation process, the study of overlapping speech in the corpus, and the association of turn management values to the emotional ones. Finally Sect. 20.5 concludes the presented work and provides future directions.

20.1.1 Related Work on Automatic Emotion and Conflict Detection

A rich set of combined speech features has been investigated and it is often employed, usually related to the temporal, prosodic, as well as the spectral content of the speech signal, to capture any underlying emotional pattern reflected upon these features (Schuller et al. 2010, 2011; Morrison et al. 2007). Previous works have studied emotion recognition and more specifically anger in speech (Neiberg and Elenius 2008; Lee and Narayanan 2005; Burkhardt et al. 2009; Polzehl et al. 2011; Erden and Arslan 2011). On a task of discriminating five emotions (fear, anger, sadness, neutral, and relief) in real-world audio data from a French human-human call center corpus, an average detection rate of 45 % was reported with only 107 acoustic features (Vidrascu and Devillers 2007). Support vector machines (SVMs) and Gaussian mixture models (GMMs) have shown a reasonable accuracy

in detecting anger in voice-controlled telephone services (Neiberg and Elenius 2008; Burkhardt et al. 2009) (average recall of 83 % and 69 %, respectively). Duration measures seem not to play an important role in emotion detection (Burkhardt et al. 2009). Incorporating linguistics and training of “emotion salient” words seems promising (Lee and Narayanan 2005), with fusion of acoustic and linguistic cues slightly improving overall scores as in Polzehl et al. (2011). Approaches to automatic conflict detection rely on extracting and analyzing nonverbal behavioral informative cues (Kim et al. 2012) and exploit turn organization features (Pesarin et al. 2012).

20.2 Data Collection

Our data collection consists of 135 audio files corresponding to call center human-human dyadic conversations between an operator and a customer. The conversations come from a customer support service of a Greek telecommunications company and are classified according to their content in six major categories: (1) churns, outgoing calls to contractual customers that have requested to cancel their contract with the company (the operators ask the customers the reason for their choice and attempt to change their mind); (2) customers, incoming calls about any issue related to customer service (technical problems, bills, complaints, etc.); (3) telesales, outgoing calls to customers aiming to sell regular phone contracts; (4) upgrade, outgoing calls informing customers about new offers; (5) mobile, outgoing calls aiming to sell mobile phone contracts; and (6) welcome, outgoing calls to welcome the customers to the company’s network. Each category consists of unequal number of files.

The overall duration of the corpus is approximately 9.5 h. Each conversation corresponds to a unique customer, while the operator might be the same in more than one conversation. The distribution of audio files per content, duration, and number of speakers of the corpus is shown in Table 20.1.

Table 20.1 Corpus details

Categories	# of files	Duration (min)	# of speakers	
			Operator	Customer
Customers	43	169	18	43
Churns	13	63	6	13
Upgrade	23	91	7	23
Telesales	24	121	12	24
Welcome	17	64	4	17
Mobile	15	52	9	15
Total	135	560	56	135

20.2.1 Data Annotation

The audio files were next annotated with a twofold aim to (a) identify instances of emotional behavior as expressed in the participants' conversation and subsequently (b) assign an emotional label to them. The data annotation was performed by an expert annotator. The selection of the appropriate utterances relied on the annotator's perception of verbal and paralinguistic cues expressing the speakers' sentiments and feelings. Specifically, the annotator's task was to detect units that are emotionally colored, i.e., that deviate from a nonemotional and/or neutral way of speaking in terms of either linguistic expressions or prosodic and paralinguistic properties of speech (such as loudness, intensity, etc.) and carry information about emotions the speakers are actually experiencing. These units may be of varying lengths, i.e., interjections, words, phrases, or utterances. In this respect, the conversational segments that are judged neutral or not emotionally colored by the annotator were left unmarked and unlabeled.

The identified units were in turn annotated as positive or negative. This set of values seems to be appropriate for the goals of the specific task, i.e., to describe the attitude of the speakers toward each other as well as their evaluation on the provided services or on the reported problems. It is important to assess in this domain, on the one hand, whether the customers are eventually satisfied with the services they get or seem to evaluate them negatively and, on the other hand, whether operators express the intent to resolve problematic issues, soothe possible negative effects, or are unable to fulfill the customers' requests.

In this sense, variation or scaling of similar emotions pertaining to either the positive or the negative spectrum is considered to be grouped under one of the two values. For example, no matter if speakers express helplessness, frustration, or anger, the essential part is that they eventually express a negative stance; thus, emotions of the aforementioned values will be labeled as negative.

A second reason for selecting this binary set of values was to avoid ambiguity that is expected to affect the automatic processing phase as well as the evaluation of the data by human judges. Specifically, the higher the number or the granularity of the labels is, the more complex the recognition task becomes, especially when certain labels are semantically close to each other.

Initially, the selected annotation labels consisted of a set of 25 categorical values tailored to the needs of the call center domain and inspired by inventories of categories representing emotions and related states as suggested in the EmotionML (Schröder 2013; Schröder and Pelachaud 2012). In practice, annotating the data with this fine-grained set of labels proved to be a hard task due to the difficulty to assign an appropriate label to speech units showing relatively insufficient perceptual cues in order to disambiguate between labels of semantically similar values. For example, though it was easy to discern between units expressing opposite emotional states, such as satisfaction and anger, there were lots of ambiguous units which were perceptually considered representative of more than one single emotional label (e.g., anger/irritation/frustration).

Furthermore, this binary set of positive and negative values would facilitate the inter-annotator agreement experiments, where different annotators are expected to make faster and more reliable judgments when using binary labels. In addition, this approach will facilitate the evaluators’ task when the emotion is not distinctly expressed (as in the case of phone calls, where noise is a constant factor).

The distinct 25 categorical labels initially selected were thus mapped to two values, namely, positive and negative. Table 20.2 shows the diversity of emotion types that positive and negative classes refer to.

The aforementioned process resulted in the annotation of 1,396 speech units. Their distribution according to content categories, label type, gender, and speaker role is depicted in Figs. 20.1 and 20.2 below. The call types where an emotional behavior is detected more frequently are churns and customers. The negative emotional label prevails over the positive one, suggesting that when speakers exhibit an emotional behavior, it is usually targeted to expressions of complaints and dissatisfaction. Furthermore, the majority of emotional units are uttered by customers.

Table 20.2 Coarse categorical emotion labels used and mapping to detailed values

Coarse	Fine grained
Positive	Pleasure, satisfaction, excitement, interest, politeness, certainty, relief, trust, surprise, reassurement
Negative	Anger, annoyance, irritation, disappointment, frustration, anxiety, worry, helplessness, confusion, doubt, uncertainty, irony, indifference, surprise, suspicion

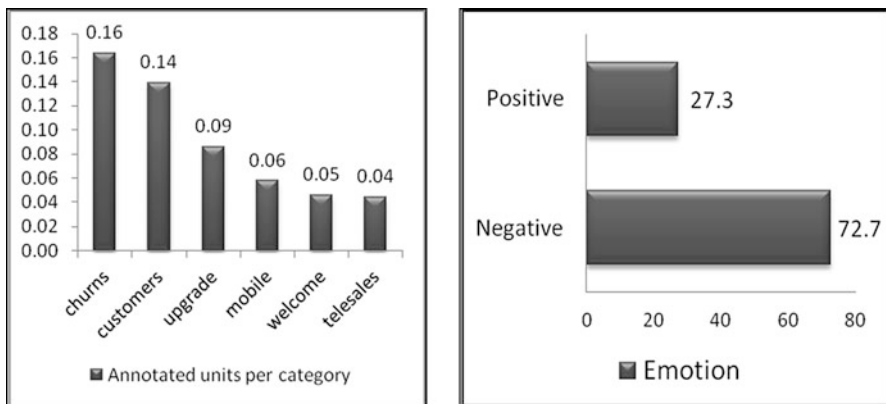


Fig. 20.1 Annotated units per file type—ratio of the total duration of the annotated units over the total duration of the audio files according to the call types (*left*) and the binary emotional labels (*right*)

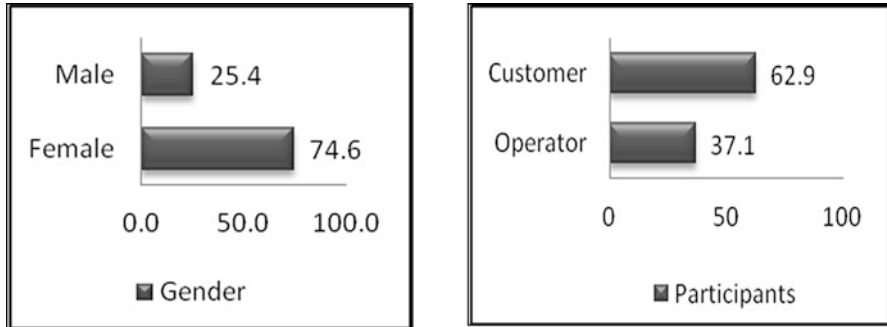


Fig. 20.2 Distribution of emotionally annotated units according to gender (*left*) and speaker role (*right*)

Table 20.3 Agreement results between annotators within the same language (GReek and ITalian) and between the naive annotators and the expert one

Agreement results			
<i>GR1 vs. GR2</i>	<i>Expert vs. GR1</i>	<i>Expert vs. GR2</i>	<i>Average</i>
80.4 %	89.3 %	79.8 %	83.2 %
<i>IT1 vs. IT2</i>	<i>Expert vs. IT1</i>	<i>Expert vs. IT2</i>	<i>Average</i>
40 %	44.6 %	65.30 %	50 %

20.2.2 Perceptual Experiments and Inter-annotator Agreement

To measure inter-annotator agreement, the extracted and annotated units were assigned to two Greek nonexpert annotators (GR1, GR2) to label them as positive, negative, or neutral. The “neutral” value was given as an option so as not to bias the annotators toward the positive or negative direction. The annotators had no prior knowledge of (a) the context of the discussion and (b) the content of the files the units were extracted from. The two annotators agreed on the 74 % of speech units, showing a high agreement with the expert annotator. The out-of-context assessment was then compared to the expert annotator’s assessment, resulting in an agreement average of 83.2 %, cf. first row of Table 20.3.

In parallel, driven by works reporting on psychological experiments carrying out a comparative analysis of subjective perceptions of emotional states (Riviello et al. 2011; Esposito and Riviello 2011), a cross-cultural pilot experiment was carried out to investigate the role of paralinguistic information and language in the perception of emotional information. Specifically, our goal was to test through a small-scale experiment the human ability to infer emotional information through only perceptual cues and how effective it is compared to the knowledge of the linguistic content.

The experiment involved the assessment of the speech units by two Italian annotators (IT1, IT2) who do not have any knowledge neither of the Greek language nor of the context of the discussions. Inter-annotator agreement was measured among the two Italian annotators as well as between each of them and the expert annotator (cf. Table 20.3).

Table 20.4 Inter-annotator agreement between the Greek (left) and the Italian (right) subjects

GR inter-annotator agreement			IT inter-annotator agreement				
	Positive	Negative	Neutral		Positive	Negative	Neutral
Positive	4.2 %	17.6 %	2.7 %	Positive	3.8 %	1.2 %	0
Negative	0.4 %	69.5 %	1.2 %	Negative	9.4 %	36.1 %	0
Neutral	0.1 %	4 %	0.3 %	Neutral	24.2 %	25.2 %	0.1 %

The low agreement score (40 %) between the Italian annotators is mostly because one of them used the “neutral” label frequently, while the other did not, as can be shown in the detailed inter-annotator agreement scores in Table 20.4. The “neutral” label was mainly attributed to units whose paralinguistic properties could not drive the annotator to infer whether those units have a positive or negative value. This also explains why the neutral instances annotated by one of the Italian subjects are equally attributed to a 50 % of positive and 50 % of negative labels (cf. Table 20.4) by the other Italian subject.

This pilot experiment suggests that paralinguistic cues are essential for the perception of emotions in speech as well as that lexical or linguistic information drastically improve the annotation’s accuracy. Thus, these preliminary results show that the decoding of positive and/or negative emotion in speech units largely depends on the native language knowledge and the communication context. Native speakers seem to be favored in comparison to the nonnative ones because of their ability to infer linguistic and the semantic contents in addition with the exploitation of prosodic and paralinguistic information. This assumption, however, needs to be verified by further experimentation including more elaborate conditions as well as an adequate number of nonnative subjects.

20.3 Automatic Emotion Classification Experiments

In order to automatically classify the data at hand, the speech units were shuffled and grouped into a training (TR) and a testing (TE) set, respectively, in such a way that the resulting sets refer to disjoint speakers. Also, to avoid bias toward one or another category during the training and the testing phases, the corpus splitting resulted in parts that contain a similar proportion of positive/negative, operator/customer, and male/female speech units (cf. Table 20.5).

The TR set (1,150 units) was used for training two different machine learning algorithms to discriminate between emotionally positive and negative speech units. The TE set (246 units) was used for assessing the algorithms’ performance on unseen positive and negative speech examples.

Table 20.5 Corpus annotation statistics—percentage of annotated units per label, speaker role, and gender

	Negative (%)	Positive (%)	Operator (%)	Customer (%)	Female (%)	Male (%)
Train	72	28	38.5	61.5	78.7	21.3
Test	76	24	30.5	69.5	55.3	44.7
Total	72.7	27.3	37.1	62.9	74.6	25.4

20.3.1 Automatic Feature Extraction

Audio Features. To extract audio features, the extended set of speech features as proposed by the Interspeech 2010 Paralinguistic Challenge was exploited (Schuller et al. 2010). The speech features are computed using openSMILE, the audio feature extraction front-end component of the open-source Emotion and Affect Recognition (openEAR) toolkit (Eyben et al. 2010). A total of 1,582 acoustic features were extracted for each speech unit, including mainly descriptive statistical functionals (DSFs) computed over low-level descriptors (LLDs), i.e., speech features derived on a frame-level analysis.

Given the varying lengths of our speech units, the DSF extraction provides static feature vectors for speech units of different sizes, and therefore, their use is beneficial with respect to the nature of the problem (Schuller et al. 2007, 2011). The resulting feature vector acoustically describing each speech unit includes many speech feature types and statistical functionals, to cover prosodic (e.g., loudness, pitch, pitch envelope, etc.), spectral (e.g., energy, log Mel frequency bands, MFCCs, line spectral pairs, etc.), as well as voice quality (e.g., jitter, shimmer, etc.) quantities. Specifically, the 1,582 first level functionals are obtained from 21 DSFs applied over 34 LLDs plus their deltas and 19 DSFs applied over 4 LLDs and their deltas together with two more features regarding pitch onsets and turn durations. The LLDs are computed on a frame rate of 10 ms with a window size of 40 ms (except for MFCCs and LSPs where the window size was set to 25 ms) and then are smoothed with a moving average low-pass filter. The DSFs, among others, include lower-order moments, extremes, percentiles, quartiles, regression coefficients, peaks, etc. Additional details can be found in Schuller et al. (2007; 2011).

Additional Features. For each speech unit, the three following features were also taken into account: (a) the speaker role, i.e., whether a unit is uttered by a customer or an operator, (b) the gender (male/female), and (c) the speech unit's duration that may vary from 0.5 to 12 s.

20.3.2 Classification Experiments

LIBSVM (Chang and Lin 2011), a popular open-source implementation of support vector machines (SVM) learning method, was used for the classification experiments. Given a set of training instances labeled as positive (+1) or negative (−1), the

SVM selected model learns an optimal hyperplane that separates these classes. The learned model can then be used to predict the label (category) to be attributed to an unseen speech unit. In our experiments, the SVM model selected uses a radial basis function (RBF) kernel which, as shown in previous works, is effective in similar emotion classification tasks (Mower et al. 2011).

The SVM model was initially trained considering only the 1,582 features that openSMILE extracted from each unit. We used WEKA’s implementation (Hall et al. 2009) of information gain (IG) and Pearson correlation (PC) to select the N most relevant to the task features ($N = 1, 2, 3, 1,582$). The feature selection was performed on the whole training set. At a next level, we repeated the same experiments using the additional abovementioned features (role, gender, and duration). To distinguish among the two different sets of features used, the first method using 1,582 features was indicated with SVM-RBF1, and the second one using 1,585 features was named SVM-RBF2. Both methods use PC as a feature selection method, since it performed slightly better than IG. A majority classifier was used as baseline, i.e., a classifier that assigns the label that dominates in the training set, the negative in our case, to all instances of the test set.

20.3.3 Results

The obtained classification results from the SVM-RBF1 and SVM-RBF2 procedure are illustrated in cf. Fig. 20.3. The x -axis indicates the number of the features

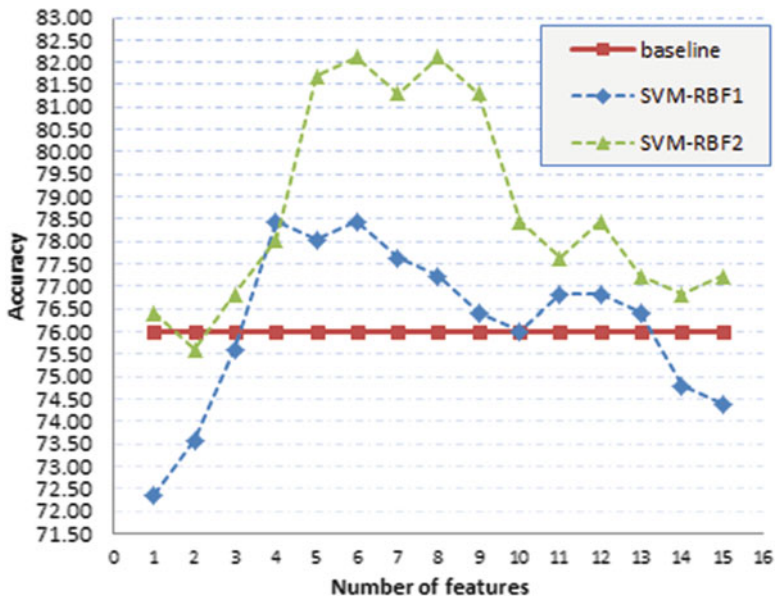


Fig. 20.3 Accuracy results obtained with different feature sets

Table 20.6 Precision and recall scores for SVM-RBF2 ($N = 6$)

	Negative (%)	Positive (%)
Recall	97.33	33.90
Precision	82.35	80.00

Table 20.7 Best features

Ten best features (PC measure)	
1	Speaker role
2	logMelFreqBand_sma[6]_amean
3	logMelFreqBand_sma[7]_amean
4	lspFreq_sma[1]_quartile1
5	logMelFreqBand_sma[1]_percentile99.0
6	logMelFreqBand_sma[6]_quartile1
7	logMelFreqBand_sma[6]_quartile2
8	logMelFreqBand_sma[7]_quartile1
9	logMelFreqBand_sma[7]_quartile2
10	logMelFreqBand_sma[5]_amean

used, i.e., the N best according to PC, and the y-axis shows the accuracy, i.e., the proportion of the correctly classified instances from the TE. More specifically, the baseline method achieves a 76.01 % accuracy, while SVM-RBF1 and SVM-RBF2 achieve a maximum accuracy of 78.45 % (for $N = 6$) and 82.11 % (for $N = 6$), respectively. In general, SVM-RBF2 has a higher accuracy for almost all values of N . This is most probably due to the fact that the speaker role (customer/operator) feature has the highest PC value.

The effectiveness of the best classifier SVM-RBF2 ($N = 6$) was further analyzed using precision and recall measures for the negative and positive labels. In particular, as shown in Table 20.6, the classifier achieves high precision and recall scores for the negative label. On the other hand, the recall scores for the positive label are much lower than the negative one.

Table 20.7 includes a ranked list of the ten best features according to PC measure.

A further qualitative analysis of the erroneously classified speech units shows that the majority of them are attributed to female (79 %) rather than male (21 %) speakers in accord with the different gender distribution in the corpus. In addition, operators' speech units are slightly worse classified (53 %) than customers' (47 %) ones. This is largely due to the nature of the conversations: for example, operators, in repeated attempts to reassure customers that their problems are being dealt with and a solution is underway or in explaining a misunderstanding with regard to a certain procedure, often speak loudly and in a severe manner. This may produce a clash between their positive intents, as expressed in the verbal content of the utterance and the paralinguistic properties of their speech that leads to a classification of these cases as negative instead of positive. The aforementioned cases account for the 37 % of the errors. On the other hand, there are cases of customers' negative instances, expressed, however, in a calm and quiet manner; hence, these instances are erroneously classified as positive.

20.4 Turn-Taking Structure and Conflict

Exploring the turn-taking structure in the conversation is closely linked to the investigation of conflict, since conflicting exchanges may be traced in the structure of the floor management. While compliance with the turn-taking rules guarantees a successful flow of conversation, irregularities may indicate that there is a tendency for disagreement and intensity, negotiation of opinions, and social situations and thus may be associated with conflict. In an attempt to identify turn-taking cues or patterns related to conflict, we added an extra data annotation level to enrich part of the corpus with this information.

Conversations from the churn category were manually annotated with turn management labels. Specifically, all turn transition points in the audio files were marked on the time axis and were assigned with a label describing the type of turn transition. The ELAN¹ editor was used with a set of labels mainly inspired by the MUMIN schema (Allwood et al. 2007) and with the addition of the backchannel value as shown in Table 20.8 below.

Churn files were selected as a representative content category with regard to conflict, since in this type of conversations, the participants pursue their own individual and conflicting goals. The customers have acknowledged issues which lead them to the decision of quitting their contract, while the operators attempt to change their mind after inspecting their problems and suggesting solutions. A typical attested behavior is customers complaining on noneffective services and/or company's unfair behavior. The clients generally express distrust in the company's services and the feeling of not being adequately protected in their customers' rights. On the other hand, the operators try to soothe the customers' negative feelings performing a series of planned actions devoted to resolve the inconsistencies among customers' wishes and actual services provided by the phone company. We thus believe that the churn category would be of particular interest to further study its turn-taking organization.

Table 20.8 Turn management annotation labels

Turn management labels	
Turn take	A speaker initiates the turn by introducing a new topic
Turn grab	A speaker takes the turn without being offered to do so, possibly by interrupting
Turn accept	A speaker accepts a turn that is offered
Turn offer	A speaker offers the turn to another speaker
Turn yield	A speaker yields the turn being under pressure or interrupted
Turn complete	A speaker completes a turn
Turn hold	A speaker attempts to keep the turn
Backchannel	A speaker produces acknowledgments and backchannels

¹<http://tla.mpi.nl/tools/tla-tools/elan/>

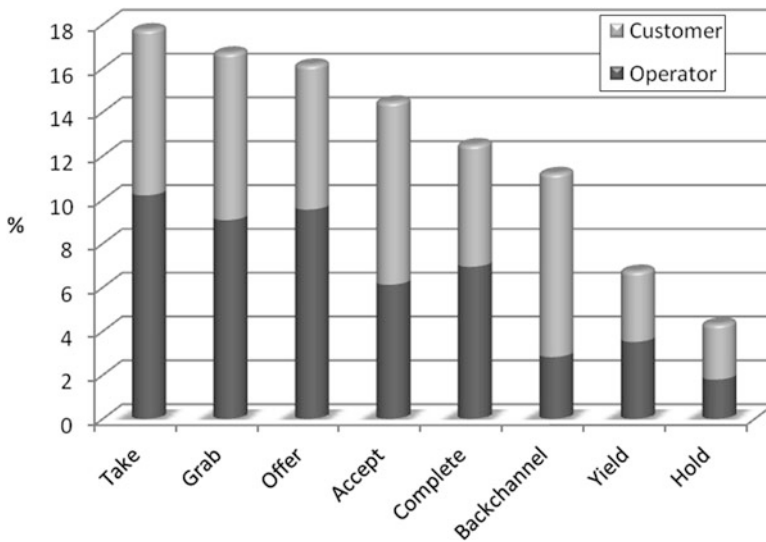


Fig. 20.4 Percentage distribution of turn management labels assigned to customers and operators

In 63 min of 13 churn files, 1,455 turn management segments were annotated with the respective set of labels and their distribution is depicted in Fig. 20.4. The second most frequent label, *turn grab*, indicates that there is a large number of interruptions performed by both participants and seems to function like an effective cue in tracing conflict. The distribution of labels per speaker role, i.e., customer and operator, may vary, as shown in Fig. 20.4. Most of the differences in speaker roles are explained by the label semantics, i.e., it is expected that the operators perform more turn offers by, i.e., asking customers questions, and that customers accept the turn more frequently than the operators. There are nevertheless interesting differences, especially regarding the *turn grab* and *turn yield* labels, the distribution of which implies that both speakers are engaged on a more or less equal basis in interruption instances and thus in conflictual situations.

Another conversational feature related to conflict according to the literature (Sacks et al. 1974; Schegloff 2000; Schuller et al. 2013) is that of overlapping speech, which is considered as a “violation” of the social rule that one party should speak at a time and therefore may be informative of speakers’ interrupting attempts to grab the floor. Recent work has outlined the role of overlaps as a reliable cue accounting for the presence of conflict and a sign of competition for having the floor, focusing on their frequency and duration (Grezes et al. 2013).

In this respect, we calculated the overlaps between turn management labels by directly exploiting the annotations, i.e., by extracting all instances where there is an overlap between the conversational actions of the two speakers in a turn transition point, as it is, for example, in the case where a speaker grabs the turn and at the same time the other speaker yields the turn. Overlapping labels account for 31.6 % of the

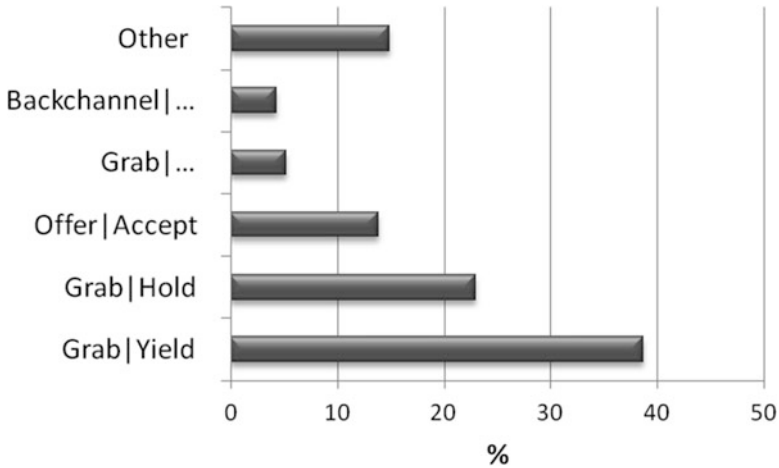


Fig. 20.5 Percentage of overlapping cases between turn management labels

annotations in the turn management layer. Annotations of turn management that do not overlap are, in their majority, cases of turn initiation, greetings, and starting or ending the discussion. In the latter cases, no overlapping speech is observed.

Most of the overlaps (67 %) occur with the *turn grab* label, which is meant to indicate interruptions. Specifically, *turn grab* overlaps mostly with *turn yield*, followed by *turn hold*, as well as with some other labels, as shown in Fig. 20.5 below. It is attested that not all instances of simultaneous talk constitute interruptions; there might also be collaborative or delayed completions, overlaps at transition relevant places (c.f. the offer-accept pair of values), signaling feedback, or backchanneling. However, the high frequency of the *turn grab* labels (paired with other labels) is an indicator of pure interruptions that express the intention to claim the turn and implies conflict between the goals and beliefs of the two speakers.

To better account for conflict cues emerging from the corpus, the turn-taking structure was associated with the expressed emotions exploring the distribution of emotionally annotated units overlapping with a turn management label at any point in the units' lengths, i.e., either at the beginning, middle, or the end of a speech unit. This was done by extracting from the annotated corpus speech segments attributed both to a turn management and an emotion label. During a given interval, it was often attested that an emotional unit overlapped with two turn management units, when simultaneous speech occurs. For example, a negative emotion is expressed by a speaker during simultaneous speech, where two turn management labels are annotated, e.g., *turn grab* by one speaker and *turn yield* by the other. In this case, two overlaps are measured for the emotional unit perspective (one for grab and one for yield, respectively) and one overlap from the turn management perspective. The results show that 83 % of the emotionally labeled speech units overlap with at least one (but also more than one) turn management label and 35 % of turn management labels overlap with emotional labels because of the high number of

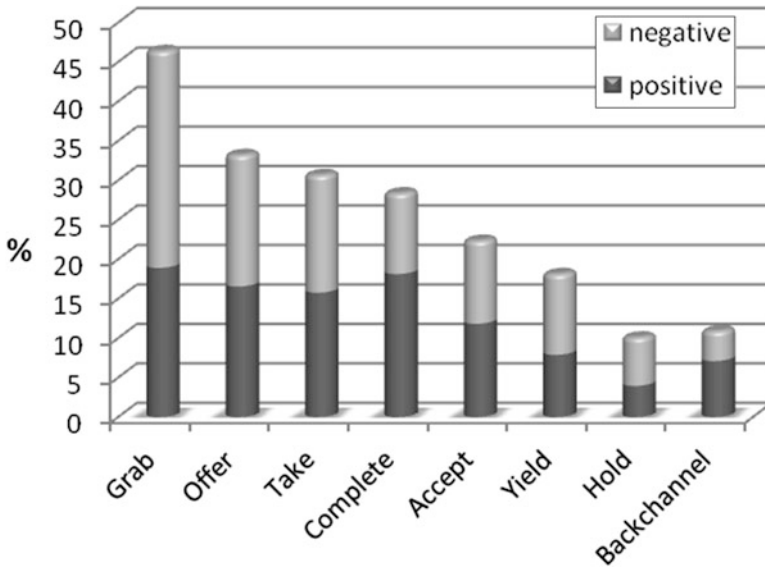


Fig. 20.6 Percentage distribution of turn management labels overlapping with an emotion label

turn managements with respect to the emotional speech units. Turn management labels that are not associated to emotions are mostly turn transitions (e.g., new topic introduction, greetings in the end of the conversation).

As far as the distribution of the labels is concerned (cf. Fig. 20.6), the findings are interesting in that labels that are related to interruption points and that may indicate conflict, such as *grab* or *yield*, overlap more with negative than with positive emotions. A t-test showed the overlaps of negative labels with turn management are statistically significant for *turn grab* with $p = 0.008$, *turn yield* with $p = 0.030$, and *turn hold* labels with $p = 0.031$. In particular, the significance of *turn hold* labels indicates that the speaker who is being interrupted attempts to hold the turn signaling a conversational cue related to conflict.

On the other hand, turn management labels related to the normal conversational flow and regular turn exchanges, i.e., *accept* and *complete*, overlap more with positive emotions. A t-test showed that, in this case, the positive emotional labels' overlaps are statistically significant for *turn accept* ($p = 0.003$) and *turn complete* ($p = 0.014$).

The reported results provide some evidence with regard to the relation between turn management labels (representing the rules and their aberrations in terms of how the exchange of turns is performed) and emotional states expressed by the speakers, indicating the whereabouts of potential conflict points within the conversational structure. Before generalizing though, these findings need to be further investigated and compared to other audio files categories, such as the welcome calls, where the distribution and the correlation between emotion and turn management labels remain to be explored.

20.5 Conclusions and Future Work

In this work, we approached the notion of conflict occurring in call center interactions as a complex problem which is decomposed in subtasks related to (a) perceiving and perceptually decoding the emotions occurring in such interactions, (b) automatically classifying them, and (c) exploring the turn-taking structure to find cues and patterns related to conflict.

With regard to the perceptual decoding of vocal emotional expressions, the high agreement scores between Greek annotators indicate the existence of salient perceptual cues allowing to adequately perceive the emotional trace of an utterance, independently of the context. The small-scale perceptual experiment involving native and nonnative raters showed that familiarity with the linguistic content largely improves the assessment of emotion in positive and negative classes. Moreover, an SVM-based algorithm that classifies emotional units extracted from the conversations as positive or negative was presented, the best version of which (SVM-RBF2 $N = 6$) obtained an accuracy score that is 6 % higher than a majority classifier. Finally, in a subset of our corpus (churns), we measured the distribution of turn-taking types, and we explored the association of overlapping speech as well as of overlapping values in turn-taking and emotion to cue the presence of conflict. In this case, it was found that overlapping speech occurs mostly with conflict-related labels in turn-taking (e.g., grab, yield) and that these labels are more correlated with negative emotions.

In future, we plan to improve the classification task by using speaker diarization and speech segmentation techniques to automatically segment recordings and come up with conversation units of variable duration that includes additional features coming from the turn-taking structure and overlapping speech points. Moreover, to allow for generalizations, a cross-lingual study on the human ability to decode perceptually emotional vocal expressions derived from call center dyadic interactions is foreseen, involving more subjects and distinct experimental conditions. Future work will also investigate various ways of incorporating and modeling the temporal sequence and transitions of emotional states, both within the same speaker, and between the two speakers, to show conflict escalation and de-escalation, and discourse structures to improve the automatic classification of conflictual conversations from a business perspective.

Acknowledgments The research leading to these results has been partially funded by POLYTROPON project (KRIPIS-GSRT, MIS: 448306). Also, the participation to Dagstuhl Seminar 13451 “Computational Audio Analysis” held from Nov 3 to 8, 2013, in Wadern, Germany, inspired Anna Esposito to contribute to this work.

References

- Allwood J (2007) Cooperation, competition, conflict and communication. *Gothenbg Pap Theor Linguist* 94:1–14
- Allwood J, Cerrato L, Jokinen K, Navarretta C, Paggio P (2007) The MUMIN coding scheme for the annotation of feedback, turn management and sequencing phenomena. *Multimodal corpora for modeling human multimodal behaviour*. *J Lang Resour Eval* 41(3–4):273–287
- Burkhardt F, Polzehl T, Stegmann J, Metz F, Huber R (2009) Detecting real life anger. In: *ICASSP 2009*, Taipei, Taiwan, 19–24 Apr
- Chang CC, Lin CJ (2011) LIBSVM: a library for support vector machines. *ACM Trans Intell Syst Technol* 2(3). <http://www.csie.ntu.edu.tw/~cjlin/libsvm>
- Erden M, Arslan LM (2011) Automatic detection of anger in human-human call center dialogs. In: *Interspeech 2011*, Florence, Italy, 28–31 Aug
- Esposito A, Riviello MT (2011) The cross-modal and cross-cultural processing of affective information. In: Apolloni B et al (eds) *Proceedings of the 2011 Conference on Neural Nets WIRN10: Proceedings of the 20th Italian Workshop on Neural Nets*. IOS Press Amsterdam, The Netherlands, pp 301–310
- Eyben F, Wollmer M, Schuller B (2010) openSMILE—the Munich versatile and fast open-source audio feature extractor. In: *ACM multimedia*, Florence, Italy, pp 1459–1462
- Grezes F, Richards J, Rosenber A (2013) Let me finish: automatic conflict detection using speaker overlap. In: *Interspeech 2013*, ISCA, Lyon, France
- Hall M, Frank E, Holmes G, Pfahringer B, Reutemann P, Witten IH (2009) The WEKA data mining software: an update. *SIGKDD Explor* 11:10–18
- Jahromi AT, Sepehri MM, Teimourpour B, Choobdar S (2010) Modeling customer churn in a non-contractual setting: the case of telecommunications service providers. *J Strateg Mark* 18(7):587–598
- Kim S, Filippone M, Valente F, Vinciarelli A (2012) Predicting the conflict level in television political debates: an approach based on crowdsourcing, nonverbal communication and Gaussian processes. In: *ACM international conference on multimedia*, Nara, Japan, pp 793–796
- Lee CM, Narayanan S (2005) Toward detecting emotions in spoken dialogs. *IEEE Trans Speech Audio Process* 13(2):293–303
- Morrison D, Wang R, De Silva LC (2007) Ensemble methods for spoken emotion recognition in call-centres. *Speech Comm* 49(2):98–112
- Mower E, Mataric MJ, Narayanan S (2011) A framework for automatic human emotion classification using emotion profiles. *IEEE Trans Audio Speech Lang Process* 19(5):1057–1070
- Narayanan S, Gregoriou P (2013) Behavioral signal processing: deriving human behavioral informatics from speech and language. *Proc IEEE* 101(5):1203–1233
- Neiberg D, Elenius K (2008) Automatic recognition of anger in spontaneous speech. In: *Interspeech 2008*, Brisbane, Australia, 22–26 Sept, pp 2755–2758
- Pesarin A, Cristani M, Murino V, Vinciarelli A (2012) Conversation analysis at work: detection of conflict in competitive discussions through semi-automatic turn-organization analysis. *Cogn Process* 13(Suppl 2):533–540
- Polzehl T, Schmitt A, Metz F, Wagner M (2011) Anger recognition in speech using acoustic and linguistic cues. *Speech Comm* 53(9–10):1198–1209, Special Issue on Sensing Emotion and Affect—Facing Realism in Speech Processing
- Riviello MT, Chetouani M, Cohen D, Esposito A (2011) Inferring emotional information from vocal and visual cues: a cross-cultural comparison. In: *IEEE 2nd international conference on cognitive computation*, Budapest, pp 1–4
- Sacks H, Schegloff E, Jefferson G (1974) A simplest systematics for the organization of turn-taking in conversation. *Language* 50:696–735
- Schegloff E (2000) Overlapping talk and the organisation of turn-taking for conversation. *Lang Soc* 29(1):1–63

- Schröder M (ed) (2013) Emotion markup language (EmotionML) 1.0, W3C Proposed Recommendation 16 Apr 2013. <http://www.w3.org/TR/emotionml/>
- Schröder M, Pelachaud C (ed) (2012) W3C vocabularies for EmotionML, W3C Working Group Note 10 May 2012. <http://www.w3.org/TR/emotion-voc/>
- Schuller B et al (2007) The relevance of feature type for the automatic classification of emotional user states: low level descriptors and functionals. In: Interspeech 2007, Antwerp, pp 2253–2256
- Schuller B et al (2010) The INTERSPEECH 2010 paralinguistic challenge—age, gender, and affect. In: Proceedings of 11th international conference on spoken language processing, interspeech 2010—ICSLP, Makuhari, Japan, 26–30 Sept, pp 2794–2797
- Schuller B, Batliner A, Steidl S, Seppi D (2011) Recognising realistic emotions and affect in speech: state of the art and lessons learnt from the first challenge. *Speech Comm* 53(9/10):1062–1087, Special Issue on Sensing Emotion and Affect—Facing Realism in Speech Processing
- Schuller B, Steidl S, Batliner A, Vinciarelli A, Scherer K, Ringeval F, Chetouani M, Weninger F, Eyben F, Marchi E, Salamin E, Polychroniou A, Valente F, Kim S (2013) The interspeech 2013 computational paralinguistics challenge: social signals, conflict, emotion, autism. In: Interspeech 2013, ISCA, Lyon, France
- Vidrascu L, Devillers L (2007) Five emotion classes detection in real-world call center data: the use of various types of paralinguistic features. In: International workshop on Paralinguistic Speech—between models and data, ParaLing, pp 11–16

Chapter 21

Serious Games for Teaching Conflict Resolution: Modeling Conflict Dynamics

Yun-Gyung Cheong, Rilla Khaled, Christoffer Holmgård,
and Georgios N. Yannakakis

21.1 Introduction

Conflict resolution skills are fundamental to navigating daily social life, but many of us acquire them only piecemeal and indirectly, over a lifetime of social interactions with others. In this chapter, we describe our game *Village Voices* (Khaled and Yannakakis 2013), an adaptive serious game designed to support children in learning about conflict resolution, in the context of conflict situations that are likely to arise in a school setting. Relying on experiential learning as an underlying learning philosophy, and based on Bodine and Crawford's six-phase model of resolving conflict (Bodine and Crawford 1998), *Village Voices* puts players in the role of interdependent villagers who need to work through the various conflicts that arise in the game world. To gradually earn conflict guru status, players must successfully complete various personalized quests, which require cooperation between players.

Our game does not represent the first instance of a game about conflict resolution. Existing games, for example, have placed players in the role of conflict mediators seeking to find a win-win solution to conflict and of mentors teaching conflict resolution skills (FableVision 2012; PlayGen 2010). What is novel about our game is that it creates situations for players to experience conflicts first-hand, and relies

Y.-G. Cheong (✉)
Sungkyunkwan University, Seoul, South Korea
e-mail: ygcheong@gmail.com

R. Khaled • G.N. Yannakakis
University of Malta, Msida, Malta
e-mail: rilla.khaled@um.edu.mt; georgios.yannakakis@um.edu.mt

C. Holmgård
IT University of Copenhagen, Copenhagen, Denmark
e-mail: holmgard@itu.dk

on an accurate player model that drives the adjustment and selection of quests for each player. As such, it provides a personalized learning experience for its players, providing them with quests appropriate to their conflict resolution abilities.

In developing this form of personalized experience, it is necessary to capture how players react to conflict, identify detectors of conflict through the game and, thereby, understand the dynamics of conflict. An innovation of *Village Voices* is the data-driven approach for assessing children's perception about conflict through a game. We therefore start this chapter by detailing the path towards designing the game based on theories about conflict resolution and learning (which are discussed in Sect. 21.2). We then present a game user survey held in Portugal involving 32 students playing *Village Voices* in a classroom setting. We log behavioural data of the players and we ask them to report (a) the conflict intensity of the game every time a conflict situation occurs (e.g., trading, message posting, stealing), (b) the current affective state, and (c) relationship status of two players involved in the conflict situation. In addition to the self-reported and behavioural data which are logged during the game, we ask students and teachers to fill in questionnaires prior to the game experience that provide information about their age, cultural background, and conflict resolution style. The data is analysed for identifying detectors of conflict intensity. It shows that gender, age, in-game behaviour, cultural tendency, and conflict management style all have a direct impact on perceived conflict in our game. In addition, the occurrences of detrimental actions in the game are associated with conflict escalation and positive actions tend to result in conflict deescalation. Finally, the children's social relationships with the other players and their emotional status within the game are key indicators of their perception about conflict.

21.2 Related Work

This section covers the central aspects of the presented study and reviews related work on serious games for conflict which involves framing conflict and learning about conflict resolution. The section ends with an outline on computational models of conflict.

21.2.1 *Serious Games for Conflict*

Serious games, games that feature non-entertainment objectives alongside entertainment ones (Winn 2008), are increasingly being recognized as effective and powerful tools for facilitating learning and encouraging behaviour change. In the case of conflict resolution in particular, games are a promising vehicle for learning. At a basic level, games provide us with the ability to create bounded worlds and realities (Huizinga 1955). Conflict falls within the class of events that is not necessarily safe or consequence-free if enacted in non-simulated environments, thus the partly

bounded nature of game worlds is a useful affordance for exploring, experiencing, and practicing conflict behaviours and resolution approaches. In addition, our focus is on the practice and internalization of skills that are highly relevant to real-world behaviour, and we note that the literature on games and role-play indicates that learning achieved within gaming contexts can be transferred to real-world knowledge and behaviours (Pedersen 1995; Raybourn 1997; Simkins and Steinkuehler 2008). Finally, games are a ubiquitous media form of our time, and are played widely within the target population we are hoping to reach with our work, namely, young people between the ages of 9 and 12. As such, games are highly promising as an intervention mechanism.

There are a number of existing serious games that deal with topics related to conflict. The *Global Conflicts* series developed by Serious Games Interactive concern games set in different locations around the world dealing with major conflicts for the purpose of challenging players' beliefs and ideas about conflict. In *Global Conflicts: Palestine*, for example, the player takes on the role of a journalist who is collecting information for a newspaper article, and must balance trust building with information collection (Serious Games Interactive 2007). *FearNOT!* is another example of a serious game about conflict, but specifically focuses on bullying. In this game, the player is an invisible friend of a virtual character who is a victim of bullying, and the player's task is to interact with the friend and advise him on how to cope with bullying-related problems (Aylett et al. 2007). *Choices and Voices* is a role-playing game in which players can experiment with peer pressure management and resistance strategies, decision making in moral dilemmas, and critical assessment of advice (PlayGen 2010). The interactive scenarios are integrated into a narrative, where players must make a range of decisions and consider different points of view. *Quandary* is a digital card-based game that presents ethical issues and conflicts involving non-player characters (NPCs) for the player to reason through from a mediator perspective, requiring critical thinking, perspective taking, and decision making (FableVision 2012).

Several design and research opportunities become apparent on examination of these games. Firstly, they are all single-player games, thus not requiring players to deal with other people in exploring and resolving conflicts. A multiplayer game, in theory, can lead to richer, meaningful, emotionally charged, and memorable game experiences and social learning, especially if the players are familiar with one another (Egenfeldt-Nielsen 2007). Secondly, many of them put players in an advisory or mediator role. On the one hand, this relieves players of encountering the effects of conflict directly, and invites them to approach decision-making in a more objective manner. On the other hand, it does not place players in situations in which they genuinely experience conflict. Formulating and enacting conflict resolution behaviours when there are personal stakes involved is considerably more difficult than simply being knowledgeable of possible response behaviours. Finally, these games present conflicts that have been pre-established and set during the game design phase. There is no adaptation to a particular player's sense of conflict: what one player perceives as a situation of high conflict may not be perceived as such by a different player.

21.2.2 Framing Conflict

Our work concerns expressing conflicts through game experiences. As there is a wide variety of literature available about conflict, and assorted perspectives on how to manage it, we have looked to perspectives that can be relatively easily interpreted for game design structures. In keeping with Thomas's (1992) views, we understand conflict as a process that is initiated when two or more parties involved in an interaction perceive that one member shows or feels strong opposition to the interaction (Thomas 1992). Notably, such an opposition can arise in relation to parties having different goals. For example, Deutsch (2006) states that "a conflict of interests occurs when the actions of one person attempting to reach his or her goals prevent, block, or interfere with the actions of another person attempting to reach his or her goals" (Deutsch 2006). Processes, characters, opposition, and goals all form a natural fit with game design.

Other broadly accepted characteristics of conflict have also assisted us in conceptualizing and representing conflict from a game design context. At base, conflict is a social interaction between opposing sides with a specific temporal duration. Laursen and Hafen (2010) explain the typical components of conflict with language we normally associate with narrative texts: "There is a protagonist and an antagonist (conflict participants), a theme (conflict topic), a complication (initial opposition), rising action (conflict behaviours), climax or crisis (conflict resolution), and denouement (conflict outcome)" (Laursen and Hafen 2010). Such a conceptualization of conflict readily maps to building blocks of game design, namely, characters, overall game objectives, obstacles that complicate achieving the objectives, game actions that supply players with means to achieve the objectives, and game outcomes.

The conflict literature reports how conflict can lead to both detrimental and beneficial consequences (Deutsch 2006; Laursen and Hafen 2010; Johnson and Johnson 1996). While detrimental consequences can include stress and other health problems, as well as emotional and behavioural difficulties, some beneficial consequences can include enhanced autonomy and individualization, and improved social, cognitive, and negotiation abilities (Laursen and Hafen 2010). It is these beneficial consequences that we set out to foster within the relatively controlled environment of a serious game played in a classroom context, in concert with post-play reflection.

Thomas suggests a taxonomy for conflict-handling modes specified in terms of two underlying dimensions: *cooperativeness* and *assertiveness* (Thomas 1992). The former describes attempts to satisfy the concerns of others, while the latter describes attempts to satisfy one's own concerns. These dimensions can be combined to describe five modes of behaviour, which in the rest of this chapter we refer to as TKI style: *competition*, *collaboration*, *compromise*, *avoidance* and *accommodation*. The competition mode is used when one party places their interests before those of another party, and thus adhere to their own solution in solving the conflict. The collaboration mode is used when solutions that are optimal for both parties are

adopted. The compromise mode is used when solutions that are acceptable for both parties are adopted. The avoidance mode is used when a party displays passive behaviour and shows no interest in conflict resolution. The accommodation mode is used when one party allows the other to control the situation. Thomas's (1992) typology suggests that in managing interpersonal conflicts, we need to consider our own goals and needs in relation to those of other parties involved, as well as how we wish to act towards other parties. We note that these conflict-handling modes also relate to conflict process and conflict outcomes.

A final point we raise with regard to our framing of conflict concerns whether conflict should be understood solely as situations in which all parties involved recognize the presence of a conflict (mutual conflicts) or whether it should also include situations in which only one party registers the presence of conflict (unilateral conflicts). Many authors define conflict in terms of an initial aggressive move from a first party, followed by an aggressive counter-move from a second party (Shantz 1987). For the applied goal of developing a broadly effective intervention tool, we believe that it is also necessary to consider situations in which the aggression of both moves is not recognized by one or both parties. This can include, for example, situations in which the first party intends a move as playful but the second party interprets the move as hostile.

21.2.3 *Learning About Conflict Resolution*

Understanding the theoretical qualities of conflict can only go so far in terms of informing approaches on how to manage conflict resolution, or how to teach conflict management. Bodine and Crawford (1998) note that the way conflict resolution is typically approached in schools includes third party settlement of disputes, reactive responses after conflict has reached a critical point, and breaking of school rules as key focal points. Additionally, our own informal studies reveal that many schools have no specific conflict education, and teachers have no guidance on how to approach conflict resolution in the classroom.

Bodine and Crawford (1998) identify the following six abilities as being critical in the constructive management of conflict. They are: *Orientation abilities*: Holding appropriate values, beliefs, and attitudes (e.g. respect for others' opinions; the desire to resolve conflicts in a way that benefits both sides), *Perception abilities*: Understanding both parties' points of view on a particular conflict situation, *Emotion abilities*: Managing negative feelings such as anger, frustration, and fear, *Communication abilities*: Listening and speaking to other parties in such a way as to enable the effective exchange of facts and feelings, *Creative-thinking abilities*: Brainstorming a variety of options for resolving the conflict, and *Critical-thinking abilities*: Choosing effective, objective criteria for judging between options.

To facilitate young people in dealing with their conflicts more proactively and effectively, and based on the six aforementioned abilities, Bodine and Crawford (1998) have developed a six-phase problem-solving process. These are: *Setting*

the stage: Assuring students that they will be listened to and not judged, and that all the parties are equally valued, *Gathering perspectives*: Collecting as many points of view as possible, *Identifying interests*: Using communication abilities to determine the underlying sources of conflict, and to focus on people's interests rather than positions, *Creating options*: Using creative-thinking abilities to come up with imaginative, mutual gain solutions to conflict-related problems, *Evaluating options*: Using critical-thinking abilities to apply objective criteria for determining the suitability of a conflict resolution option, and finally, *Generating agreement*: Coordinating integrated deployment between the two opposing parties, across all of the foundation abilities.

Importantly, Bodine and Crawford's (1998) process puts young people in the position of resolving their own conflicts. To enable young people to build up the skills necessary to resolve conflicts, we therefore opted to design a multiplayer game for the classroom that would create conflict situations for players to experience from a first-person perspective. Such a game would provide a bounded, supervised opportunity for players to experiment with conflict behaviours and their consequences.

21.2.4 Computational Models of Conflict

While the literature extensively discusses variant theoretical models of conflict (some of which are mentioned earlier), it is rather sparse with regard to computational models of conflict. In our previous work (Cheong et al. 2011) we proposed a computational model of conflict for serious games as a generic adaptive framework. The framework consists of five processes: *conflict creation*, *conflict detection*, *player modeling*, *conflict management*, and *conflict resolution*. Campos et al. (2013) present a conflict process that focuses on conflict emergence and escalation from an emotional agent. The model accounts for an agent's decision-making process when latent conflicts (caused by interference of incompatible goals) exist. The model was implemented in the FATiMA emotional agent architecture (Campos et al. 2012), which in turn was used as an architecture for NPCs appearing in the serious game *My Dream Theatre*. In *My Dream Theatre*, the player is given the task of directing a play, which involves allocating roles to different game NPCs. Each NPC may respond to a conflict-triggering event differently, depending on its TKI style (Thomas 1992).

A number of attempts have been made to model the types of conflicts found in narrative. Swanson and Jhala (2012b) present a computational model of interpersonal and group conflict. Their model proposes *cause*, *diagnosis*, *intervention*, *expectation*, and *relationship* as contextual factors that may influence the dynamics of an ongoing conflict. Via Amazon's Mechanical Turk, Swanson and Jhala (2012a) also collected a corpus of narratives recounting real conflict experiences, for the purpose of learning about the relationships and interactions between the contexts and conflict dynamics from the data. From the initial analysis of the corpus, they identified six core dimensions which can categorize the parties in a dispute: *active*, *aggressive*, *interest*, *yielder*, *solver*, and *involvement*.

Ware et al. (2013) present an AI-plan based computational model of narrative that operationalizes conflict. In their Conflict Partial order Causal Link (CPOCL) model, a conflict can be understood as a series of steps that threaten or obstruct part of another plan, i.e. negate causal links that have been established for another plan. CPOCL is based on the perspective that “conflict occurs when a character forms a plan that is thwarted by another event in the story, or would have been thwarted if the event had succeeded”. Ware et al. (2013) further define seven dimensions that characterize conflict: *participants*, *reason*, *duration*, *balance*, *directness*, *stakes*, and *resolution*. The first three dimensions (i.e. participants, reason, and duration) are used to detect if conflict exists, while the remaining dimensions are used to generate stories that embed different kinds of conflicts. Their experimental studies validated the notion that the *participants*, *reason*, and *duration* metrics could detect conflict as well as the average human subject. The *balance*, *directness*, and *resolution* metrics were also found to be in agreement with what human subjects judged. On the other hand, there was disagreement between CPOCL and subjects on evaluating the degree of *stake*. The nature of conflict in Ware et al. (2013) differs from our work in that their understanding of conflict concerns a narrative phenomenon which can potentially involve multiple participants and a structured story, while the conflict dealt with in this particular study refers to emotion experienced by an individual participant.

While the aforementioned studies take a theoretical stance on conflict modeling or approach conflict indirectly, in this paper we introduce a data-driven approach to conflict model construction which allows subjects to *experience* conflict and report it. For that purpose we collected data from children playing the *Village Voices* game, asking them to fill-in demographics and conflict style questionnaires and recording their behavioural patterns in the game. We further asked them to report the conflict intensity of each game quest during play. We then used linear mappings to approximate the underlying function between in-game behaviour, user profile (e.g. conflict style) and conflict intensity.

21.3 Game Design

Village Voices is a four-player open world game that takes place in a fictional village set in pre-industrialization times. It is designed to be played in a classroom setting by players who know one another, under teacher supervision. On the surface, the game is about survival and prosperity in the village. On closer inspection, however, the game is about friendship and reputation management in the village, and mastery of conflict resolution.

Each game world contains four player-controlled character roles: the blacksmith, the innkeeper, the alchemist, and the carpenter (see Fig. 21.1). When the game begins, each player is assigned a particular character to play. This character role is retained for the entire duration of the player’s experience with *Village Voices*. As part of daily life in the village, players are required to undertake various actions



Fig. 21.1 A screenshot from *Village Voices* where the player of the Blacksmith role is given the quest to gather three metals

related to maintenance of their characters' livelihoods, and to complete quests related to their responsibilities within the village. For example, the alchemist must tend to his crop of magic mushrooms, keep an eye on his own health, and maybe in the process of collecting and processing items, to build a wall to keep wolves out of the village. At the same time, all of the characters are interdependent, thus situations inevitably arise that trigger conflicts or exacerbate existing ones. For example, in order to complete the barrier wall, the alchemist may need to obtain an item from the innkeeper, who he is not on good terms with due to a previous theft incident involving the innkeeper helping herself to the alchemist's mushrooms. While players may initially be faced with simple quests involving no trades or only one trade with other characters (trading interface shown in Fig. 21.2), more difficult quests involve trades with all three of the other characters (see Table 21.1). Given that players have the ability to perform actions that can lead to conflict—including theft, property damage, spreading rumours, and not sharing collective resources such as food, completing multi-player quests rapidly becomes a difficult proposition in the absence of negotiation.

Many digital learning games adopt an explicitly didactic approach to conveying domain knowledge. But how one resolves conflict in a constructive manner is contextually defined; it makes little sense to teach “correct responses”. In addition, an individual's life experiences will play into their interpretations of acceptable or desirable modes of conduct. What becomes important, then, is to explore possible responses and what they may entail in specific situations and social contexts, and to make sense of them on one's own terms. As such, in terms



Fig. 21.2 The trading interface in *Village Voices*

of pedagogical approaches, *Village Voices* draws from social constructivism and experiential learning theory (Kolb 1984). Instead of explicitly instructing players how to resolve particular conflicts, the game creates situations in which players are pushed into conflicts with one another. Players therefore experience conflict from a first-person perspective, and must use the affordances of the game to enact conflict management strategies.

Crucial for learning experiences of this kind is access to learning facilitation. While any individual play session is designed to last around 15 min in itself, it is followed by a 10–15 min reflective debrief featuring the other players and a learning instructor, thus a complete game experience takes around 30 min. During the debrief, players are able to address the issues raised during the play session, relate game experiences back to life experiences, reason through and advise one another on alternative resolution strategies, and collectively negotiate rules to guide future play sessions. Game sessions are intended to take place once or twice a week, over a time period spanning approximately 6 weeks. It is important that the game is played over a number of weeks, as opposed to in a block of back-to-back sessions, as this gives players the opportunity to reflect on how their play behaviours evolve over time. Furthermore, it enables players to identify parallels between conflict episodes experienced in daily life, and the problem solving strategies employed in-game.

In terms of underlying conflict resolution philosophy, *Village Voices* draws on Bodine and Crawford (1998) previously discussed six-phase model of resolving

Table 21.1 Quest structure and narrative texts for innkeeper

Quest number	Title	Description
Q1	Farming	You are an Innkeeper. And Innkeepers can gather Grain. Try to gather 3 Grain. Find a cornfield and click on it. The Grain will then appear in your pack. Gather 3 grain
Q2	The scarecrow	Oh no! There are birds all over the cornfields. And they eat the grain. They MUST be stopped. You should build a scarecrow to keep them away. Get 3 metal (trade with Blacksmith). Make 1 food (gather 5 grain and click on the oven)
Q3	Biggest bread—ever	The mayor of the village has asked you to make the BIGGEST bread EVER. This is a big challenge, but if you get some grain and a magical potion, maybe you will be able to do it? Gather 4 grain. Get 1 potion (trade with Alchemist)
Q4	Spirits of the cloud	The spirits of the cloud need to be paid. They keep the sky clear so the sun can shine and make the grain grow. If they stop, it will turn dark and the grain will stop growing. Gather this for the spirits: 2 metal (trade with Blacksmith), 2 mushrooms (trade with Alchemist), 5 wood (trade with Carpenter), and 2 foods
Q5	The present	One of your friends has a birthday and you want to make a very special present. A kite you can sit on while it is flying! Great magic is needed for this, so get 1 potion (trade with Alchemist), 1 tool (trade with Blacksmith), 1 furniture (trade with Carpenter), and 1 food

conflict. It creates situations in which players develop their *perceptual skills*, that is, they learn to see an issue from different perspectives, as the game triggers conflicts of different types and intensities based on player relationships and conflict experiences within the game, and requires players to present conflict solutions for one another. The game encourages awareness of *emotional skills*, namely, recognizing the distortion role that strong emotions play during conflict, as it requires players to regularly update their feelings towards other characters. The game makes player develop their *communication skills*, by requiring them to express their perspectives, feelings, and strategies for conflict to other players and also to the learning facilitator. The game invites practice of *creative-thinking skills*, as there are no “correct” ways to resolve specific conflicts triggered within the game world, and players must piece together their own potential conflict resolution strategies given the resources and networks they have within the game. Finally, the game invites players to hone their *critical thinking skills*, in judging between options. Conflicts are only deemed constructively resolved once both parties feel sufficiently at ease with the conflict outcome, thus an appropriate conflict resolution approach is very much contingent on context, requiring application of critical thinking.

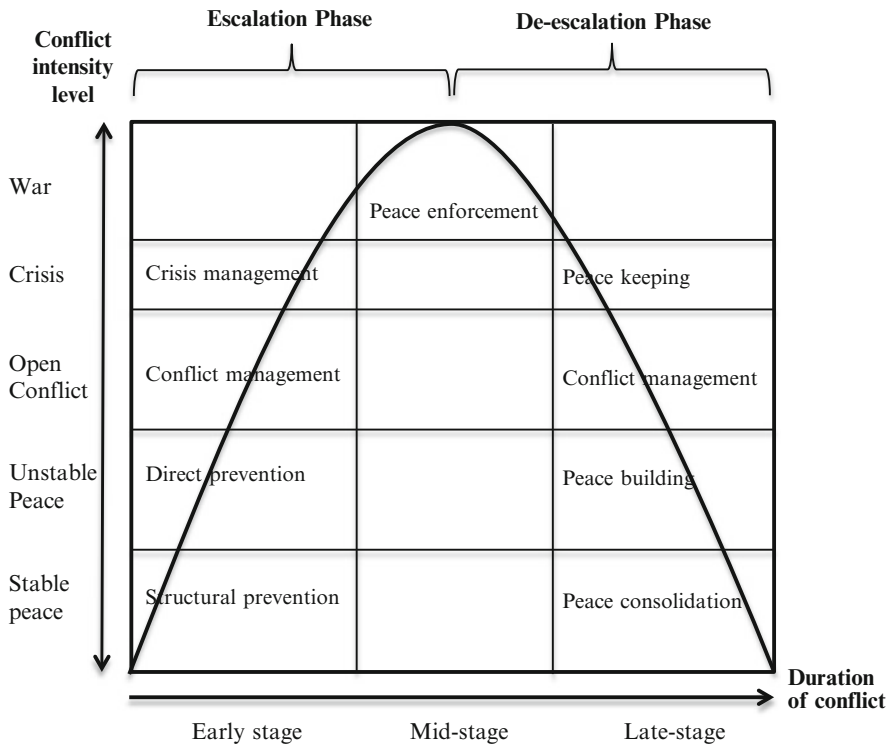


Fig. 21.3 A conflict cycle proposed by Swanstrom and Weissmann [reproduced from Swanstrom and Weissmann (2005)]

As *Village Voices* is premised around creating conflict situations, it is therefore centrally important that individual players indeed perceive game situations as actual conflicts, as without this they would lack meaningful grounds to draw on during the debrief phase. But perception of conflict can vary radically from one individual to another. Part of the design challenge therefore became to ensure that each player experienced conflict, in spite of their different conflict perception thresholds. The literature on conflict proposes that it typically takes the form illustrated in Fig. 21.3.

Village Voices attempts to create this experience curve for each individual player, by selecting quests that appropriately escalate or de-escalate the conflict they are currently experiencing. An example of quests that are given sequentially to the character *innkeeper* is shown in Table 21.1. The quests were designed so that a player’s perception of conflict rises as the quests advance. We hypothesized that the degree of intensity of conflict that players feel escalates as the number of trading partners increases, and the number of resources required grows. As such, in *Quest 1*, the simplest quest, the player is asked to collect her own resources only. In *Quest 2*, the player is required to obtain items from one other player, while in *Quest 4*

the player must obtain items from all three of the other players. Put differently, we expected that players would be more likely to experience greater conflict when pursuing *Quest 4* than when pursuing *Quest 2*.

To test our assumptions regarding quest design and in order to model the conflict escalation and deescalation process, we collect the following three metrics from players: experienced conflict intensity, emotional state, and current feelings towards other players following major game events (e.g. completion of quests, trading success and failure, stealing, message posting, and spreading rumours).

Although the game can continue indefinitely, the game has a number of badges that players can collect. These badges both serve the purpose of introducing to players to the possibilities of the game world, while also articulating play goals for them to work towards, some of which concern game progress, others of which concern conflict learning progress. Examples of badges include “Rude surprise”, which is earned by a player when another player leaves a bad comment about her, “Change of heart”, which is earned when a player registers a significant change of feelings about another player, “There’s No “I” in Team”, which is earned after successfully completing a trade with three players during a single quest, and “Wise Words to Live By”, for giving helpful advice to other players during the reflective debrief conversation.

21.4 Experimental Protocol and Data Collection

In order to test the game’s ability to induce the escalation and deescalation of conflict, as well as reflective learning of conflict resolution strategies, we conducted a game user study. The user study also served the purpose of collecting data for training the computational user models enabling the game’s scenario adaptation.

The study was run at a Portuguese school with 32 (20 male) children between the ages of 10 and 12 who played 21 game sessions in total. The maximum duration of each play session was 15 min. All sessions consisted of groups of four players who were classmates. All groups consisted of children of the same gender. Choosing classmates with prior relations meant that the actions and evaluations in the game plausibly would be influenced by these relations. However, since this would be the natural use case for the game in a school setting, we consider this advantageous, rather than detrimental, to the user study.

Every participant played multiple game sessions (2–4). In total, we collected 2,126 reports about conflict intensity, emotions, and judgement about the other players across the 21 game sessions. In order to establish a baseline for assessing the short-term impact of interacting with the game, the students were asked to report their emotions and judgements about the other players at the very start of the session. These ratings were updated whenever key events occurred in the game environment. Events included rumoring, message posting, trading, and stealing actions. The conflict intensity in the beginning of the game was assumed to be 0

Table 21.2 Categories and types of data collected during the game (online) and just before (offline)

	Category	Data type
Offline	Demographic data	Age, gender, years in Portugal
	Cultural tendency	individualist index, collectivist index
	Aggression tendency	victimization index, aggression index
	TKI conflict resolution style	accommodation, competition, compromise, collaboration, avoidance
Online	In-game actions	Trading accept and reject rumour and positive message posting stealing
	Self-reported	Conflict intensity (11-point Likert scale) affective state (angry, sad, happy,neutral) social relationship with others (5-point Likert scale)

as no in-game actions have happened yet. Subsequently, the conflict intensity was updated whenever any student reported his or her own rating about conflict, on the completion of a quest, and at the end of the game.

Additional data gathered from the participants included their demographic background data, a psychological profile, self-reported conflict resolution strategies, and behavioural data from the game session. The demographic data consisted of information about participant age, gender, and whether they had been residents of Portugal for a minimum of 5 years. This information was collected from the participants' school teachers.

Profile information consisted of a number of psychometric measures obtained from the following questionnaires (see Table 21.2): The Aggression/Victimization Scale (Dahlberg et al. 2005; Orpinas and Frankowski 2001), an adapted version of the Thomas–Kilman Conflict Mode Instrument (Thomas and Kilmann 1974) modified to facilitate comprehension by children, and a questionnaire assessing individualist/collectivist orientation (Triandis 1995). The children completed the questionnaires before interacting with the game (offline).

Children's self-reports were collected through the in-game interface during gameplay (see online data types in Table 21.2 and the reporting schemes/interfaces in Fig. 21.4). The reports include ratings (0–10) of experienced levels of conflict, ratings of conflict with other individual players (a variant of a 5-point self-assessment manikin), as well as indications of the player's immediate emotion (four affective states to pick from: angry, sad, neutral, happy). In-game behavioural data was logged automatically including all core game actions (i.e. trading, spreading rumours, message posting, and stealing).

Finally, sessions were video recorded in order to ensure supporting qualitative material for the subsequent interpretation of the game data, as necessary.

For the user study, the following protocol was followed: Background data was collected several days before the actual user study, in order to allow the children

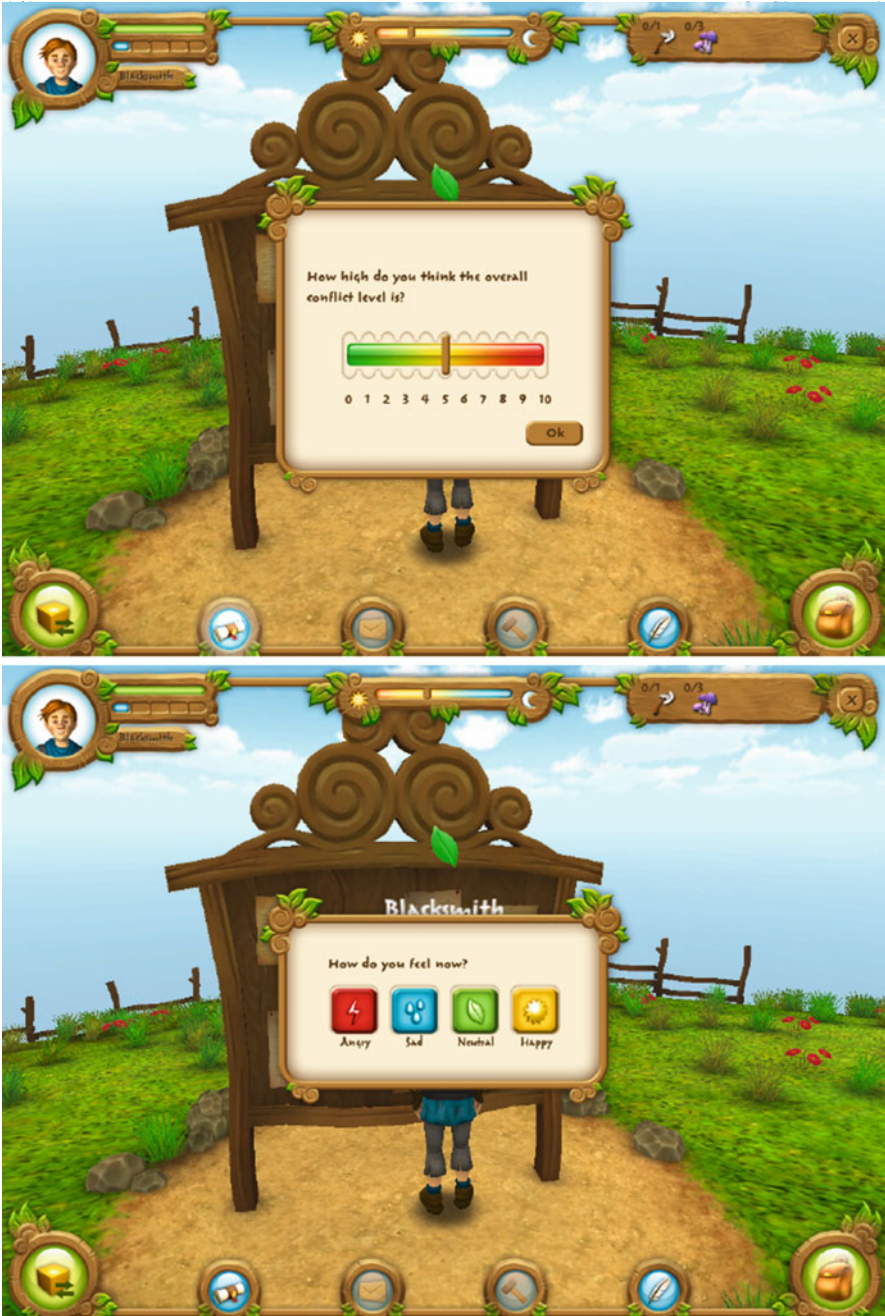


Fig. 21.4 Three interfaces to collect data on the player’s affective state. The *topmost* interface requires the player to specify the level of conflict intensity s/he currently feels on a 0–10 Likert scale. The *middle* interface requires the player to select one of four available affective states (angry, sad, neutral, happy) to describe his/her current emotional state. The interface at the *bottom* asks the player to report his/her current relationship with a particular player on a 5-point self-assessment manikin represented by emoticons (i.e. highly positive, somewhat positive, neutral, somewhat negative, highly negative)



Fig. 21.4 (continued)

to focus on the game during the gameplay sessions. The questionnaires were administered in the classroom with a regular teacher of the children's, assisted by an experimenter.

On the day of the user study, once the children entered the classroom, a common presentation of the purpose of the study and an introduction to the game was given in Portuguese. Two experimenters played through a shortened session of the game, demonstrating the interface and the object of the game, completing quests. All features of interaction were demonstrated in order to ensure the children knew all forms of interaction with and through the game. This included demonstrating quest management, resource collection, trading, and stealing. Children were allowed and encouraged to ask questions during this demonstration session.

Following the introduction, the children were divided into groups by the experimenters with the help of their regular teacher. Each session was conducted in a round table arrangement with the children seated facing each other in order to facilitate easy communication during gameplay. For the same reason, the game was played on laptop computers allowing the children to simultaneously perceive each other as well as the game. To ensure minimum technical complications, the experimenters facilitated the start of game for all groups. In the case of technological failure for one player, an experimenter would restart the session if the game had recently begun, but would allow the session to finish if the game was almost over, before including the child again. The decision of whether to restart or let the session

play out was made on a case-by-case basis in order to ensure that all children felt included throughout the study. During sessions, the experimenters interacted with the children to handle questions about the game's interface, functionality, and rules, but otherwise attempted to have minimal influence on in-game proceedings, avoiding giving the children any advice on which actions to take during gameplay. In the case that children asked experimenters to intervene or arbitrate in any conflict, the experimenters would communicate that it was the children's decision to make. Apart from fulfilling this observational and supportive role, experimenters were instructed to only intervene in the actual gameplay and mediate in the case that a conflict escalated to an unacceptable level, for ethical reasons. This did not prove necessary during the user study. At the end of the session, all children were collectively thanked for their participation.

21.5 Conflict Rating Analysis

This section outlines the key findings of our analysis of the data. In particular, we begin by describing our data processing process. Next, we present the result of analysing the conflict ratings with respect to gameplay time and quest played (Sect. 21.5.2), and age and gender (Sect. 21.5.3). We then proceed to examining the relationship between conflict ratings and the TKI style of children (Sect. 21.5.4) and the impact of children's cultural tendencies on their conflict ratings (Sect. 21.5.5), and we conclude by investigating the effects of in-game actions to conflict ratings (Sect. 21.5.6).

21.5.1 Data Preprocessing

As mentioned in the previous section, the students' profiles/TKI styles and reports were separately collected. Since we were investigating the relationships between conflict ratings and the student's demographic data, each report was annotated with the reporting student's age, gender, TKI style, and cultural tendency. The TKI conflict resolution strategies and collectivist/individualist orientation values were computed from the student's responses to the questionnaire distributed before gameplay.

Due to a technical error in the logging system, the timestamps in the logs collected during the first 2 days of the experiment were not available. As a result, of the 2,126 reports collected, 1,523 reports lacked timestamp information, while only the reports collected the last day contained timestamps. As such, the number of reports with timestamp information was reduced to 603. Worth noting is that these 603 reports all came from male students, which accounts for 50.9% of the male population involved in our dataset.

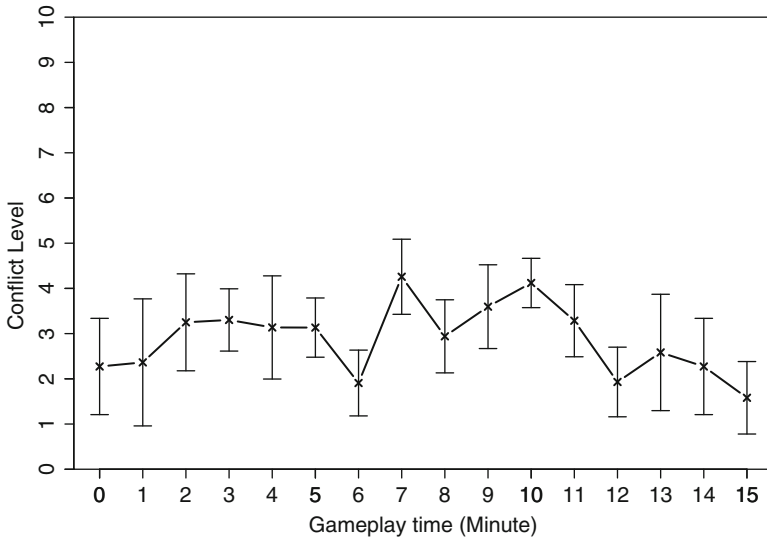
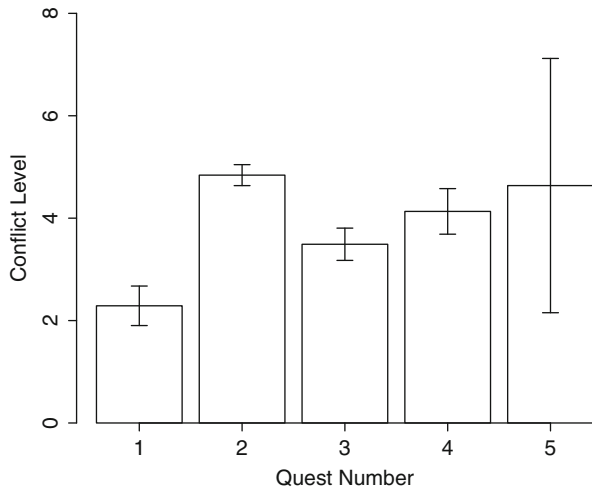


Fig. 21.5 The means of conflict ratings by minute with 95 % standard error bars (number of reports = 603). The x axis denotes the time of gameplay by minute and the y axis denotes the average of conflict ratings which are measured on a scale from 0 to 10

Fig. 21.6 Average conflict rating values over time (95 % standard error bars are included) from all the students (number of reports = 2,126). The X axis denotes the quest number and the Y axis denotes the average reported conflict rating across subjects



21.5.2 Gameplay Time and Quest Played

Figure 21.5 shows the mean values of the 603 ratings (i.e., the reports containing time information when the rating was recorded) including a 95 % of error bar for each minute of gameplay. The conflict levels that are reported during, for example, the first minute of gameplay are averaged and shown at the 0 min on the x axis.

Next, we examined whether there is a relationship between the conflict ratings and the quests given to the students. Figure 21.6 shows the average of conflict ratings submitted by students per quest, across all of the 2,126 reports, as timestamps are not necessary in this analysis. As the graph shows, students reported higher conflict levels as the quest number increased with exception of *Quest 2* which produced the highest conflict level.

Analysis and Discussion Averaging the conflict ratings across the minutes of gameplay generates a graph which consists of three bell-shaped curves. Although individual variances exist, the width of a bell curve generally corresponds to the duration of a single quest. This finding is consistent with the principles of Swanstrom and Weissmann (2005), suggesting that a conflict experience is expected to reveal a bell shape in terms of conflict intensity (see Fig. 21.3). Furthermore, an analysis on the relationship between the reported conflict ratings and quests reveals that children felt greater conflict as they received more difficult missions. Increased conflict intensity in children is in agreement with the principle of our quest design; that is, higher conflict is induced as the quests advance. It is notable that conflict intensity peaks at *Quest 2*, even though *Quest 2* should have theoretically been one of the easier quests amongst the five quests. Given that *Quest 2* is the first quest to introduce trading between two players (see the quest description in Table 21.1), we hypothesize that the high level of conflict was perceived as a result of the introduction of trading, rather than the inherent difficulty of *Quest 2*.

21.5.3 Student Profile (Age and Gender)

A correlation analysis was carried out to test if the student's profile (i.e. age and gender) had an impact on her evaluation about a conflict situation. While all the reports contained valid gender information (number of reports = 2,126), some reports lacked age information. For the analysis of the relationship between age and conflict ratings, these reports were eliminated, which resulted in 2,043 valid reports in total.

The analysis shows a statistically significant difference between genders as determined by one-way ANOVA ($F(1,2124) = 229.46$, $p = 0.00$) (see Table 21.3). On average, female students were shown to report higher conflict ratings than male students (see Table 21.4). Moreover, we observe a significant difference between age groups as determined by one-way ANOVA ($F(2,2040) = 115.01$, $p = 0.00$) (see Table 21.5). A post-hoc Tukey test showed that the conflict ratings reported by 11-year-old students were significantly lower than those reported by students in the other two age groups at $p = 0.000$, as can be seen from Table 21.6.

Analysis and Discussion Gender has been found to be a crucial factor affecting perception and experience of conflict. In particular, girls reported significantly higher conflict ratings than did boys. Statistical analysis also revealed that children of 11 years reported significantly lower ratings than children from the other two

Table 21.3 One-way ANOVA analysis on the conflict ratings by student’s gender

Source	df	SS	MS	F	Sig.
Gender	1	2,734.33	2,734.33	229.46	0.0000***
Residuals	2,124	25,310.20	11.92		

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’

Table 21.4 The mean value and standard deviation of conflict ratings reported by each gender group

Gender	N	Mean	SD
Female	941	5.50	3.77
Male	1,185	3.22	3.18
Total	2,126	4.23	3.63

Table 21.5 One-way ANOVA analysis on the conflict ratings by student’s age

Source	df	SS	MS	F	Sig.
Age	2	2,649.46	1,324.73	115.01	0.0000***
Residuals	2,040	23,496.85	11.52		

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’

Table 21.6 The mean value and standard deviation of conflict ratings for each age group

Age	N	Mean	SD
10	801	5.28	3.55
11	967	2.89	3.53
12	275	4.76	2.22
Total	2,043	4.08	3.58

groups (10 years and 12 years old). However, no clear pattern was found to explain the relationship between perception of conflict and age.

21.5.4 TKI Style

We investigated whether the student’s TKI style was related to their perception of conflict (reported via conflict ratings). We consider that the TKI category that has the maximum value is being the dominant one. We eliminated data instances in which more than one TKI style is dominant for each student resulting in 1,049 data samples. For example, we would disregard a data sample if the accommodation and the collaboration indexes for a student are the highest among the five and both equal to, e.g., 10.

As shown in Table 21.7, the majority of the TKI styles found in the conflict ratings fell under the accommodation style (70.2 % of the total report with a single dominant TKI style). The number of subjects in the competition and the compromise groups were marginal; and avoidance did not appear as a dominant style. Applying

Table 21.7 Means and SD of conflict ratings for TKI conflict style (avoidant style is omitted because it did not appear as a dominant style)

TKI style	N (percentage)	Mean	SD
Accommodation	736 (70.2 %)	3.82	3.1
Collaboration	221 (21.1 %)	2.60	3.37
Competition	41 (3.9 %)	2.15	2.69
Compromise	51 (4.9 %)	3.71	4.51
Total	1,049	3.49	3.27

Table 21.8 The result of one-way ANOVA analysis on the conflict ratings by TKI style

Source	df	SS	MS	F	Sig.
Between	3	334	111.43	10.7	0.0000***
Within	1,045	10,881.91	10.41		

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’

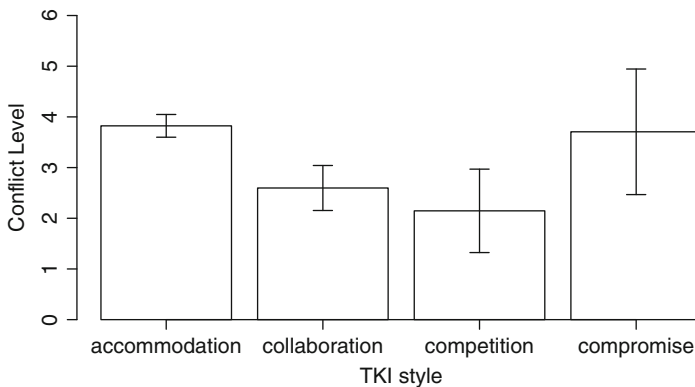


Fig. 21.7 The means of conflict ratings by the student’s dominant TKI style (number of reports = 1,049)

one-way ANOVA on the groups we find significant differences in conflict ratings among TKI style groups ($F(3,1045) = 10.7, p = 0.000$) (see Table 21.8). Since the conflict ratings do not necessarily follow a normal distribution, we ran the Kruskal–Wallis test which also showed that TKI style has a significant impact on the conflict ratings ($H(3) = 40.0453, p\text{-value} = 1.042e-08$). Moreover, a post-hoc Tukey test showed that the conflict ratings reported by the students in the accommodation group differed from those reported by the students in the competition group and those in the collaboration group significantly at $p < 0.01$.

The conflict ratings reported by the students in the compromise group were not significantly different from the other three groups, although some differences are shown (see Fig. 21.7).

Analysis and Discussion The statistical analysis presented here shows that a student’s self-evaluation of her conflict management style has an impact on her conflict ratings. It appears that the TKI style is linked to the student’s judgement about a conflict situation in terms of intensity. The students in the accommodation

and compromise groups tended to give higher ratings; and the students in the competition group tended to give lower ratings. However, it is noted that there might be a discrepancy between the self-reported TKI style and the student's real TKI style, since the majority of the students' responses to TKI questionnaire were evaluated as either accommodative or collaborative categories (91.3 % of all the valid reports).

21.5.5 Cultural Tendency

We also examined whether the student's collectivism/individualism orientation was related to their perception of conflict (reported via conflict ratings). We considered the orientation with a higher value as being the dominant one. This resulted in us eliminating the ratings associated with students for whom collectivist and individualist index values were equal. This resulted in 1,461 reports being available for analysis.

Table 21.9 reveals that the majority of the students were oriented towards the individualism orientation, and made up 87.2 % of the total reports available. As shown in Fig. 21.8, there is a clear difference in conflict ratings between the two groups, with students in the individualism group tending to give higher ratings than those in the collectivism group. The two groups' average reports are significantly different ($t = 11.94$, $df = 298.974$, $p = 0.000$).

Analysis and Discussion The statistical analysis indicates that a student's self-evaluation of her collectivist/individualist orientation has a significant impact on her conflict ratings. The students in the individualist group appeared to rate conflict situations significantly higher than the students in the collectivist group. It also appears that the students who were individualist tended to be more sensitive to conflict situations, therefore giving them higher ratings.

21.5.6 In-Game Actions

We examined potential effects between reported conflict ratings and in-game behaviour, reported emotions, and social networks updates. Table 21.10 presents the features recorded and extracted for investigation in our study.

Table 21.9 Means and SD of conflict ratings for Collectivist and Individualist style

Orientation	<i>N</i> (percentage)	Mean	SD
collectivist	188 (12.87 %)	1.59	2.66
individualist	1,273 (87.13 %)	4.20	3.60
Total	1,461	3.86	3.60

Fig. 21.8 The means of conflict ratings by the student's dominant Collectivism/Individualism orientation (number of reports = 1,461)

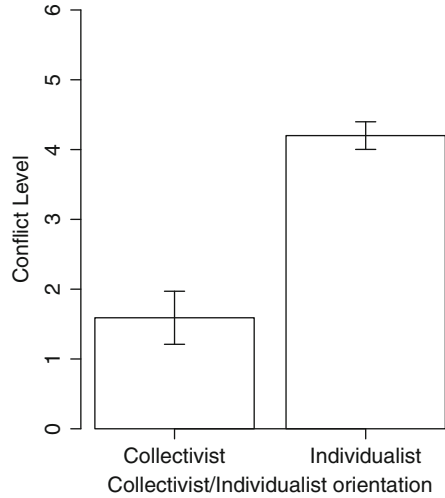


Table 21.10 Game Features where t denotes the time that the log is recorded

Feature type	Symbol	Description
Reported features	C	Conflict rating reported by the player at t
	S	The sum of all the players' reports about social relationships with the other players (where higher number denotes disliking)
	E_H	The number of players who report being happy at t
	E_S	The number of players who report being sad at t
	E_N	The number of players who report being neutral at t
	E_A	The number of players who report being angry at t
In-game actions	G_D	The number of trade rejections featuring the player up to t
	G_R	The number of incidents concerning spreading or receiving rumors featuring the player up to t
	G_M	The number of positive messages posted or received featuring the player up to t
	G_S	The number of thefts (stealing or being stolen from) featuring the player up to t
	G_T	The number of successful trades featuring the player up to t

While C represents a single student's report only, the social judgement about the other players (S) and emotion factors (E_H , E_S , E_N , and E_A) represent collective states at the moment. S is the sum of all the player's social judgements about the others participating in the game. Therefore, the range of S is between 0 (when all the players reports are highly positive) and 16 (when all the players reports are highly negative). The emotion factors define the number of the players who reported being in the particular emotion. For instance, if three players report that they are happy and the fourth one reports that she is angry, E_H is 3 and E_A is 1 and the rest emotion factors (E_N , E_S) are 0.

In addition to the features shown in the table, two complementary factors were computed from the recorded in-game actions and used in the analysis: the detrimental factor (D) and the beneficial factor (B). The detrimental factor is the summation of the counts of the harmful actions that the player was involved in either as the giver or receiver (Eq. (21.1)). The harmful actions in this study include rejecting a trade, spreading a rumour, and theft

$$D = G_D + G_R + G_S. \tag{21.1}$$

The beneficial factor (B) is the sum of the counts of the positive actions that the student was involved in, such as posting a positive message, accepting a trade, or being involved in altruistic gifting, either as giver or receiver (Eq. (21.2))

$$B = G_M + G_T. \tag{21.2}$$

The correlation analysis between conflict ratings, in-game actions, and other reported features reveals some noteworthy effects. As is evident from Table 21.11, in-game harmful actions (D) beneficial actions (B) were, respectively, positively and negatively correlated with conflict ratings. Self-reported anger (E_A) and happiness (E_H) were positively correlated whereas neutral feelings (E_N) were negatively correlated with conflict ratings. In particular, the positive association between happiness and conflict escalation is noteworthy. Since the emotion factor is the summation of all the players’ reports, we speculate that the happy emotion reported during conflict escalation may represent the view of the player who is the giver of the action rather than the view of the action’s receiver. Imagine that a high degree of conflict arises as a result of a theft. While the victim of the theft may report an angry affective state (i.e., $E_A = 1$), the other student (i.e., the thief) may report happiness (i.e., $E_H = 1$). In this case, the conflict escalation is associated with both of the angry and happy emotions. The children’s judgements (S) about other players in the game were also found to be significantly (and positively) correlated with conflict ratings. There was no significant relationship between the self-reported emotion of being sad and conflict ratings.

Analysis and Discussion Strong correlations between conflict ratings and in-game actions and self-reported responses were found. Not surprisingly, negative in-game actions tended to escalate conflict and positive in-game actions tended to de-escalate conflict. The student’s own judgement about the other three players

Table 21.11 The result of Pearson’s r analysis between reported conflict intensity, children’s reported features and in-game actions (number of reports: 2,126)

S	D	B	E_A	E_H	E_N	E_S
Social	Detrimental	Beneficial	Angry	Happy	Neutral	Sad
0.44***	0.23***	-0.15***	0.15***	0.08***	-0.10***	0.02

Significance notation: 0 “***”

and self-reported emotions were found to be direct indicators of conflict; negative judgements about other players were correlated with conflict escalation. Likewise, the angry affective state was also positively correlated with conflict escalation. Surprisingly the happy affective state was also found to be positively correlated with conflict escalation. This might have been linked to the student's role as a giver or receiver of the detrimental action.

21.6 Discussion and Conclusions

In this paper, we present a multiplayer serious game *Village Voices*, which was designed to support children between the ages of 9 and 12 in learning about conflict resolution. *Village Voices* is informed by Bodine and Crawford (1998) conflict resolution problem-solving process, and thus focuses on equipping young people with skills to address conflicts proactively and without the need for mediators, and to be reflective about conflict experiences. As our intention in *Village Voices* was to create situations of conflict for each player even though players would potentially have different perceptions of conflict from each other regarding the same game event, we set out to model conflict intensity in players during play. As such, we introduced a method for capturing conflict and conflict dynamics within a group of players via the use of serious games.

We then used the game and its available conflict quests to collect data from 32 children in Portugal each playing a number of quests in a 4-player setup in a school setting. The data consisted of demographic information, conflict resolution strategy profile information, cultural tendency, in-game behavioural data, self-reports on conflict and affect during the game. A number of features were extracted from the data and analysed in this paper. We revealed effects between conflict management style, gender, cultural tendency, and conflict intensity reports. In particular, significant effects were found among different TKI conflict resolution styles, collectivist/individualist orientation, and gender. The students in the accommodation and compromise groups gave higher ratings than the students in the competition group. Girls tend to report higher conflict ratings than boys. The students who are oriented towards individualist also tend to give higher conflict ratings while the students oriented towards collectivist tend to give lower ratings. Finally, strong correlations were found between in-game actions and conflict ratings; in particular, the self-reported emotions and social relationships indicated the degree of conflict that the students felt.

The study presented in this article is constrained by a number of limitations. First, the curved pattern of conflict experience reflects only the responses collected from the male students. A subsequent study is required to test whether a similar pattern can be clearly found from girls' reports. Second, a student's TKI style was judged by his own responses to a TKI questionnaire and not by an expert's (e.g. teacher's responses), thus, our findings on TKI style and conflict ratings might be undermined by this limitation. Third, the duration of each game session was

limited to 15 min, which can sometimes be too short for meaningful conflicts to occur, or for conflicts to occur between players with particular conflict management styles. We therefore plan longer game sessions as an immediate future study to examine the impact of gameplay time in reported conflict. Finally, while self-reports in rating format offer a potentially useful psychometrics tool for children of our age group, the study presented here treats ratings as values for a regression problem. We argue that doing so only provides an initial (and limited) perspective on the conflict phenomenon as ratings should be naturally treated as ordinal values. Several studies have already showed the clear benefits of rank-based questionnaires over rating-based questionnaires (Yannakakis and Hallam 2011; Metallinou and Narayanan 2013) which suggest the transformation of ratings into ordinal scales and the further processing of them via rank-based statistics or non-linear preference learning models.

Obtained results show that the game can successfully elicit conflict situations and that reported conflict ratings appear to follow a conflict escalation and conflict deescalation phase. While this is already an indication for the game's capacity in teaching conflict resolution, extensive studies across different countries within school settings will be required to further validate the ability of the game to transfer these skills to the real (non-digital) world. Such studies are currently running and preliminary results already suggest that both the game and its personalization mechanisms (user modeling and scenario adaptation) (Grappiolo et al. 2011; Cheong et al. 2013) offer an alternative, yet very powerful, approach to social skill learning.

Acknowledgements This research was partially supported by the FP7 ICT project SIREN (project no: 258453). The authors would like to thank Joanna Campos and Carlos Martinho at INESC-ID for organizing the user study in Portugal. We also thank all the students who participated in this study.

References

- Aylett R, Vala M, Sequeira P, Paiva A (2007) Fearnot! - an emergent narrative approach to virtual dramas for anti-bullying education. In: International conference on virtual storytelling, Saint Malo
- Bodine R, Crawford D (1998) The handbook of conflict resolution education: a guide to building quality programs in schools. Jossey-Bass, San Francisco
- Campos H, Campos J, Martinho C, Paiva A (2012) Virtual agents in conflict. In: Nakano Y, Neff M, Paiva A, Walker M (eds) Intelligent virtual agents. Lecture notes in computer science, vol 7502. Springer, Berlin/Heidelberg, pp 105–111
- Campos J, Martinho C, Paiva A (2013) Conflict inside out: a theoretical approach to conflict from an agent point of view. In: Proceedings of the 2013 international conference on autonomous agents and multi-agent systems AAMAS '13. International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC, pp 761–768
- Cheong YG, Khaled R, Grappiolo C, Campos J, Martinho C, Ingram GPD, Paiva A, Yannakakis G (2011) A computational approach towards conflict resolution for serious games. In:

- Proceedings of the 6th international conference on foundations of digital games, FDG '11. ACM, New York, NY, pp 15–22
- Cheong YG, Grappiolo C, Holmgrd C, Berger F, Khaled R, Yannakakis GN (2013) Towards validating game scenarios for teaching conflict resolution. In: Games for learning workshop at FDG (Foundations of Digital Games), Chania, Crete
- Dahlberg LL, for Injury Prevention NC, (US) C, for Disease Control C, (US) P (2005) In: Dahlberg LL et al. (eds) Measuring violence-related attitudes, behaviors, and influences among youths [electronic resource]: a compendium of assessment tools, 2nd edn. Division of Violence Prevention, National Center for Injury Prevention and Control, Centers for Disease Control and Prevention, Atlanta, GA
- Deutsch M (2006) Introduction. In: Deutsch M, Coleman PT, Marcus EC (eds) The handbook of conflict resolution: theory and practice, 2nd edn. Jossey-Bass, San Francisco, CA, pp 1–22
- Egenfeldt-Nielsen S (2007) Third generation educational use of computer games. *J Educ Multimed Hypermed* 16(3):263–281
- FableVision (2012) Quandary. <http://www.quandarygame.org/>
- Grappiolo C, Cheong YG, Togelius J, Khaled R, Yannakakis GN (2011) Towards player adaptivity in a serious game for conflict resolution. In: VS-games 2011 natural interaction and player satisfaction in games workshop, Athens
- Huizinga J (1955) *Homo Ludens: a study of the play-element in culture*. Beacon Press, Boston, MA
- Johnson DW, Johnson RT (1996) Conflict resolution and peer mediation programs in elementary and secondary schools: a review of the research. *Rev Educ Res* 66(4):459–506
- Khaled R, Yannakakis GN (2013) Village voices: an adaptive game for conflict resolution. In: Proceedings of the 8th international conference on foundations of digital games, pp 425–426
- Kolb D (1984) *Experiential learning: experience as the source of learning and development*. Prentice-Hall, Englewood Cliffs, NJ
- Laursen B, Hafen CA (2010) Future directions in the study of close relationships: conflict is bad (except when it's not). *Soc Dev* 19:858–872
- Metallinou A, Narayanan S (2013) Annotation and processing of continuous emotional attributes: challenges and opportunities. In: Workshop on emotion representation, analysis and synthesis in continuous time and space (EmoSPACE)
- Orpinas P, Frankowski R (2001) The aggression scale: a self report measure of aggressive behavior for young adolescents. *J Early Adolesc* 21:51–68
- Pedersen P (1995) Simulations: a safe place to take risks in discussing cultural differences. *Simul Gaming* 26(2):201–206
- PlayGen (2010) Choices and voices. <http://playgen.com/choices-and-voices/>
- Raybourn EM (1997) Computer game design: new directions for intercultural simulation game designers. In: *Developments in business simulation and experiential exercises*, vol 24
- Serious Games Interactive (2007) Global conflicts: Palestine. <http://globalconflicts.eu/>
- Shantz CU (1987) Conflicts between children. *Child Dev* 58:283–305
- Simkins DW, Steinkuehler C (2008) Critical ethical reasoning and role-play. *Games Cult* 3(3–4):333–355
- Swanson R, Jhala A (2012a) A crowd-sourced collection of narratives for studying conflict. In: Proceedings of the 2nd computational models of narrative workshop at LREC
- Swanson R, Jhala A (2012b) Rich computational model of conflict for virtual characters. In: Proceedings of the 12th international conference on intelligent virtual agents, IVA'12. Springer, Berlin/Heidelberg, pp 502–504
- Swanstrom N, Weissmann M (2005) Conflict, conflict prevention and conflict management and beyond: a conceptual exploration. Policy Paper
- Thomas KW (1992) Conflict and conflict management: reflections and update. *J Organ Behav* 13(3):265–274
- Thomas KW, Kilmann RH (1974) *Thomas-kilmann conflict mode instrument*. Tuxedo, NY: Xicom
- Triandis HC (1995) *Individualism & collectivism*. Westview Press, Boulder, CO
- Ware S, Young R, Harrison B, Roberts D (2013) A computational model of plan-based narrative conflict at the fabula level. *IEEE Trans Comput Intell AI Games* 6(3):271–288

- Winn B (2008) The design, play, and experience framework. In: Handbook of research on effective electronic gaming in education. IGI Global Publication, Hershey/Philadelphia, pp 1010–1024
- Yannakakis GN, Hallam J (2011) Rating vs. preference: a comparative study of self-reporting. In: Affective computing and intelligent interaction, vol 6974. Springer, Berlin, pp 437–446

Index

A

Activity and conflict, x, 57, 66–68
Ape communication, 91, 93, 94
Argumentation, xi, 3, 30, 55, 104, 117–133,
142, 211, 217, 245, 431
Audio features, 379, 380, 385–386, 389–392,
395, 397, 438
Autobiographical memories, 350, 353, 360,
364, 368, 372, 373

B

Bi-directional Long-Short Term Memory
(BLSTM), 405, 415, 416, 424, 427

C

Call centre interactions, 431–445
CMC. *See* Computer mediated communication
(CMC)
Cognitive analysis, x, 72, 76, 80
Common ground, 94, 129, 137–157, 164, 238
Competition, ix, 4, 16, 20–23, 28, 56, 83,
91–98, 104, 108, 281, 333, 384, 426,
442, 452, 461, 467, 468, 472
Computational paralinguistics, 384, 403–406,
410, 416, 426
Computer mediated communication (CMC),
159, 160, 180
Concepts, xiv, 33, 34, 39, 46, 49, 50, 138,
142–157, 160, 161, 165–168, 192, 199,
228, 296–298, 318, 319, 321, 326,
328–344, 351, 352, 358, 362, 367,
369–371, 398

Conflict, v–xv, 3–31, 33–51, 53–68, 71–87, 91,
103–112, 117–133, 137–157, 159–180,
190, 207–219, 223, 262, 267–292, 295,
318, 368, 377–399, 431–445, 449–473
detection, viii, xiv, 377–399, 427, 432–433,
454
escalation, ix, 53–68, 119, 126–131, 377,
394, 404, 445, 450, 460, 471–473
management, vii, viii, xv, 104, 108, 126,
131, 272, 300–301, 310, 311, 453, 454,
457, 468, 472
process, 271, 276, 283, 287, 292, 453, 454
talk, xi, 159–180
Conflicting interactions, 91, 138–141, 143,
144, 147, 148, 156, 157
Conversation analysis, 179, 431
Cooperation, ix, x, 3, 4, 16, 19–24, 29, 31, 56,
71, 91–98, 104, 105, 108, 110, 111,
132, 161, 191, 268, 272, 318, 333, 449
Cost/benefit analysis, 77, 85, 119, 120, 122,
128, 129
Courtroom narrative, 190–192, 199
Culture, i, vi, x, 22, 72, 77, 81, 83, 124, 164,
188, 193, 194, 209–214, 218, 219, 228,
234, 262, 310, 351, 352

D

Data analysis, 123
Decision making, xi, 41, 118, 121, 124, 130,
131, 333, 451, 454
Deep neural networks (DNNs), xiv, 405–412,
423–427
Disaffiliation, xii, 159–180

Discourse, 117, 125, 126, 143, 145, 153, 154, 157, 164, 193, 208–212, 217, 218, 247, 259, 291, 352–354, 360, 363, 371, 423, 441
 Disgust, xii, 110, 111, 224–229, 233, 235, 236, 238, 239, 253, 254, 258
 DNNs. *See* Deep neural networks (DNNs)
 Dropout, 405, 410–411, 421–423, 425–427

E

Emotional speech, 405, 444
 Emotions, vii, x, xiii–xv, 5, 56, 73, 94, 103–112, 122, 160, 188, 209, 228, 244–247, 278, 295–311, 353, 403, 431, 451, 453
 classification, 437–440
 Emotive communication, 159–180
 Empathy, xiii, 105, 109–111, 172, 295–303, 305, 307–311, 332, 333
 Escalation, viii–x, 60, 62–63, 68, 83, 85, 119, 132, 214, 259, 292, 454, 460
 Expectations, xiii, 7, 10, 15, 23, 92, 105, 112, 119, 121, 124, 129, 131, 139, 233, 234, 236, 270, 271, 408, 454

F

FACS analysis, 354, 360, 363
 Feud, 72, 77, 79, 83, 84, 132

G

Genoa G8, xii, 185–202
 Group discussion, 129, 384, 398
 Group selection, 111

H

Hierarchical neural networks, 403–427

I

Impoliteness, xii, xiii, 60, 223–239, 268, 269, 272, 329
 Intergroup conflict, 110, 207–209, 295, 377
 Intergroup relations, 295–297, 300, 301
 Internet message boards, 160, 168, 173

L

Language evolution, 91–98

M

Mental models, 318, 345
 Models of conflict escalation, 55, 57
 Multi-expert architecture, 379, 380, 392, 398, 399
 Multimodal analysis, 267–292, 349–373
 Multimodality, vii, 273
 Multi-resolution detector, 379, 387, 398

O

Obama, 349–373
 Offence, xii, 77, 172, 223–225, 227, 231–233, 236, 239
 Overlapping speech, 378, 379, 381, 386, 390, 391, 404, 405, 420, 423–426, 432, 442, 443, 445
 Overlap ratio, 404, 420, 423–427

P

Parrhesia, 349–373
 Phylogenesis, 104
 Polarization, 56, 129
 Police interview, xii, 317–345
 Politeness, vii, xii, xiii, 163, 223–239, 268–270, 325, 326, 328–330, 340, 344, 435
 Political debate, ix–xi, xiv, 54, 57–68, 138, 377, 379, 380, 417
 Political speech, 349–373
 Pragmatic markers, 161, 162, 164, 165, 168, 170, 179
 Pragmatics, vii, xi–xiii, 94, 120, 137–139, 159–180, 224, 230, 231, 235, 236, 239, 247, 253, 257, 259, 260
 Psychological motives, 72

R

RBM. *See* Restricted Boltzmann machines (RBMs)
 Rectified linear units, 405, 411–412, 422–423, 426, 427
 Reinterpretation, xi
 Replying strategies, 156
 Restricted Boltzmann machines (RBMs), 405–409, 426
 Retribution, 74, 75, 85
 Revenge, x, 71–87, 249, 254, 262
 Rhetorics, 120, 129, 160, 194, 207–219, 295–311
 Ritual, x, xiii, 92, 106–110, 191, 269, 270

S

Self-conscious pariah, 350, 351
 Semantics, xii, 138, 141, 142, 147, 151, 162, 165, 168, 179, 191, 223–239, 250, 253, 255–256, 323, 344, 354, 437, 442
 Social attributions, 295–311
 Social damage, x, 73, 78, 86
 Social delegitimization, 193–195
 Social identity, 207–208, 218
 Social interaction, x, 94, 104, 105, 129, 244–247, 249, 318, 378, 449, 452
 Sociality, 3, 16, 20–22, 24, 103–105, 109, 111, 218
 Social order, 84, 87, 106, 109
 Speech acts, xiii, 164, 165, 269, 285, 286, 332, 333
 Sport organized support, 207–208
 Stages of conflict, ix, x, 53–68
 State violence, 185–202

Struggle for interactional power, xiii, 268, 286, 292

Support vector machines (SVMs), 384, 385, 387–389, 391, 392, 398, 404, 421, 427, 432, 438–440, 445

T

Theory of mind, 97, 98
 Therapeutic jurisprudence, xii, 199–201
 Truthfulness, xiii, 295–311
 Turn-taking, xiii, xiv, 120, 165, 167, 169, 286, 288, 319, 331, 378, 379, 417, 441–445
 characteristics, 380
 Tutoring application, 345

V

Verbal aggression, xii, xiii, 268, 269, 272, 292
 Virtual agents, 317, 320, 326, 328, 329, 331, 333, 334, 345