LNAI 11939

Gábor Bella
Paolo Bouquet (Eds.)

Modeling and Using Context

11th International and Interdisciplinary Conference, CONTEXT 2019
Trento, Italy, November 20–22, 2019
Proceedings

Springer

Lecture Notes in Artificial Intelligence 11939

Subseries of Lecture Notes in Computer Science

More information about this series at http://www.springer.com/series/1244

Gábor Bella · Paolo Bouquet (Eds.)

Modeling and Using Context

11th International and Interdisciplinary Conference, CONTEXT 2019
Trento, Italy, November 20–22, 2019
Proceedings

Springer

Editors
Gábor Bella
University of Trento
Trento, Italy

Paolo Bouquet
University of Trento
Trento, Italy

ISSN 0302-9743 ISSN 1611-3349 (electronic)
Lecture Notes in Artificial Intelligence
ISBN 978-3-030-34973-8 ISBN 978-3-030-34974-5 (eBook)
https://doi.org/10.1007/978-3-030-34974-5

LNCS Sublibrary: SL7 – Artificial Intelligence

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

After 20 years, the Italian city of Trento once more hosted the CONTEXT conference for its 11th installment, during November 20–22, 2019. The success of the conference was proof that there is a scientifically diverse community "still crazy after all these years," true to the slogan of the event. Clearly there is ongoing interest in the study of various aspects of contextuality, both as a notion worthy of research per se, and as a conceptual solution to complex problems, applicable across a wide range of disciplines.

This year, 20 submissions were accepted to CONTEXT, all of which are long papers. The peer-review process involved three single-blind reviews per paper by the Program Committee (PC) members listed below. True to the interdisciplinary heritage of the conference series, the accepted papers cover a large spectrum of fields, including Philosophy of Language and of Science (10 papers), computational papers on Context-Aware Information Systems, Artificial Intelligence, and Computational Linguistics (8 papers), as well as Cognitive and Social Sciences (1 paper each). To our satisfaction, interdisciplinarity also manifested itself in the same phenomena being approached from several angles, such as the problem of contextuality in lexical semantics that was addressed from philosophical, linguistic, as well as computational points of view.

The conference also hosted invited talks by four renowned researchers: Giancarlo Guizzardi (University of Bolzano), Frank van Harmelen (Vrije Universiteit Amsterdam), François Recanati (CNRS, Paris), and Jason Stanley (Yale University). The abstracts of these talks have been included in this volume. A panel session was also held with the participation of the invited speakers as well as high-profile CONTEXT "regulars". We thank all speakers for having accepted to attend and contribute to the event.

We also thank the people who helped make the conference happen, especially the Steering Committee members: Carlo Penco, Patrick Brézillon, Roy Turner, and Fausto Giunchiglia, who contributed to the organization efforts and provided guidance throughout. We are also grateful to all PC members for all they did to ensure the high quality of accepted contributions. We thank our local organizers, Olha Vozna and Mattia Fumagalli, for handling the logistics of the entire event.

Last but not least, we thank the Department of Information Engineering and Computer Science of the University of Trento for having provided financial support.

November 2019

Gábor Bella
Paolo Bouquet

Organization

CONTEXT 2019 was organized by the Department of Information Engineering and Computer Science at the University of Trento, and was hosted at the Aula Kessler of the Department of Sociology.

General Chair

Paolo Bouquet	University of Trento, Italy

Program Chair

Gábor Bella	University of Trento, Italy

Steering Committee

Patrick Brézillon	Sorbonne Université, France
Fausto Giunchiglia	University of Trento, Italy
Carlo Penco	University of Genova, Italy
Roy Turner	University of Maine, USA

Program Committee

Varol Akman	Bilkent University, Turkey
Gábor Bella	University of Trento, Italy
Massimo Benerecetti	Università di Napoli Federico II, Italy
Marcos R. S. Borges	Federal University of Rio de Janeiro, Brazil
Paolo Bouquet	University of Trento, Italy
Patrick Brézillon	Sorbonne Université, France
Altangerel Chagnaa	National University of Mongolia
Alan Colman	Swinburne University of Technology, Australia
Biswanath Dutta	Indian Statistical Institute, India
Anita Fetzer	University of Augsburg, Germany
Thomas Forissier	Université des Antilles, France
Christian Freksa	University of Bremen, Germany
Mattia Fumagalli	University of Trento, Italy
Fausto Giunchiglia	University of Trento, Italy
Avelino Gonzalez	University of Central Florida, USA
Debra Hollister	Valencia College, USA
David Leake	Indiana University Bloomington, USA
Vincenzo Maltese	University of Trento, Italy
Fiona McNeill	Heriot Watt University, UK

Alain Mille	CNRS – Université de Lyon, France
Carlo Penco	University of Genova, Italy
François Recanati	CNRS, France
Arjun Satish	Confluent, USA
Marina Sbisà	University of Trieste, Italy
Marco Schorlemmer	IIIA-CSIC, Spain
Luciano Serafini	Fondazione Bruno Kessler, Italy
Pavel Shvaiko	Informatica Trentina, Italy
Ken Turner	University of Brighton, UK
Roy Turner	University of Maine, USA
Massimiliano Vignolo	University of Genova, Italy
Hao Xu	Jilin University, China
Roger Young	University of Dundee, UK
Arkady Zaslavsky	Deakin University, Australia
Rui Zhang	Jilin University, China

Local Organizers

Olha Vozna	University of Trento, Italy
Mattia Fumagalli	University of Trento, Italy

Sponsor

DISI—Department of Information Engineering and Computer Science, University of Trento, Italy

Invited Talks

Why Polysemy Supports Radical Contextualism

François Recanati

CNRS, France
frecanati@gmail.com

Contextualism is the view that context sensitivity generalizes to *all* expressions (whether indexical or not). All expressions are such that the content they contribute depends upon the context, in contrast to the (invariant) linguistic meaning of the expression. I will distinguish several versions of Contextualism. Moderate Contextualism appeals to the phenomenon of *pragmatic modulation*. Just as the content of an indexical depends upon the context of use, the content actually carried by an ordinary, non indexical expression also depends upon the context: it depends on *whether, and how*, the literal meaning of the expression is modulated in context. (Modulation covers processes of sense extension and sense narrowing as well as metonymies and possibly other phenomena.) What makes Moderate Contextualism moderate is the fact that modulation is *optional*: it may or may not take place. Whether or not it takes place depends upon the context, so the possibility of 'zero modulation', that is, of purely literal uses of language, is compatible with the generalization of context sensitivity characteristic of Contextualism. According to Radical Contextualism, even if no modulation takes place and words are used with their standard sense, there still is a principled difference between the linguistic meaning of an expression and the content it carries in context. Appealing to the phenomenon of *polysemy*, Radical Contextualism distinguishes between the lexical meaning of an expression (a network of senses bearing modulation relations to each other) and the particular senses it may contribute in context (whether these senses are standard or modulated). Context sensitivity thus generalizes: An expression cannot directly contribute its lexical meaning, which has the 'wrong format' for being a constituent of content. The lexical meaning must be contextually converted into an appropriate sense through various context-sensitive operations.

Empirical Studies of Context at Scale: The Case of Equality Reasoning

or: How Leibniz Got it Wrong

Frank van Harmelen

Vrije Universiteit Amsterdam
`frank.van.harmelen@vu.nl`

The rise of very large linked open datasets has allowed us over the past few years to study the structure of knowledge graphs not only in theory, but also empirically at very large scale. I will report on a number of studies that all have empirically analysed the role of context in equality reasoning in linked open data, encoded in the `owl:sameAs` predicate. All of these studies show that the standard formal semantics of equality does not suffice in practical settings, and is simply ignored and violated at a large scale. At the same time, we can show that different notions of context are very useful in making sense of what users choose to do in practice, and that within local contexts, a sensible semantics for equality reasoning does emerge.

Hustle: The Politics of Language

Jason Stanley

Yale University
jason.stanley@yale.edu

Philosophy of language and semantics have not proven to be useful in a new age of rhetoric and propaganda. Does that means its tools and resources are useless? Or do we need to think differently about these tools and resources in order to make them applicable? In this talk, part of a forthcoming book with David Beaver, I explore how to make the philosophy of language relevant to our current political moment, by rethinking the nature of presupposition and accommodation.

Objects and Events in Context

Giancarlo Guizzardi

Conceptual and Cognitive Modeling Research Group (CORE),
Free University of Bozen-Bolzano, Bolzano, Italy
giancarlo.guizzardi@unibz.it

Keywords: Events · Objects · Context · Conceptual modeling · Ontology

Introduction

"Smiles, walks, dances, weddings, explosions, hiccups, hand-waves, arrivals and departures, births and deaths, thunder and lightning: the variety of the world seems to lie not only in the assortment of its ordinary citizens…but also in the sort of things that happen to or are performed by them" [3]. This variety is also evident in our conceptualizations of reality, with, on one hand, "processes", "activities", "tasks", "events", "occurrences", "incidents" unfolding in time, and, on the other hand, "objects", "actors" and "resources" persisting through time, possibly changing in a qualitative way while maintaining their identity.

The distinction between these categories is commonplace in philosophical literature, with the former broadly referred to as "events" (or *perdurants*) and the latter broadly referred to as "objects" (or *endurants*) [3, 8]. In this spirit, I here use the term "event" broadly, including references to atomic transitions, complex processes, and even "stative occurrences" (e.g., "Mary sitting in a bench for an hour") [6]. Analogously, I use the term "object" to refer to independent entities or *substantials* (e.g., you and me, Italy, the moon, John's car), as well as parasitic "object-like" entities such as *qualities* (e.g., the objectified color of that rose; Sofia's beauty), *dispositions* (e.g., Matteo's capacity of speaking Portuguese), and *relationships* (e.g., the marriage of John and Mary, Linda's presidential mandate) [6, 8, 11].

In existing modeling frameworks in computer science and related areas, the distinction between *behavioral elements* and *structural elements* ("how" versus "what") is often invoked to account for the different nature of elements belonging to these two broad ontological categories [12, 16]. Accordingly, different modeling disciplines have been established to deal with behavioral and structural modeling, each of which with a different focus. For example, the business process modeling discipline focuses on the "event-like entities", and, in contrast, the (structural) conceptual modeling discipline focuses on "object-like entities". In each of these disciplines, entities of one of these ontological categories are first-class citizens, while the other category plays a marginal role (if any). Some notable exceptions in the process discipline are the so-called business artifact-centric approaches [4, 13, 14], and in the structural conceptual modeling discipline, the event reification approach [15]. In this talk, I argue that there are many complex domains (e.g., economics and finance, life sciences, defense, advanced engineering) and application areas (e.g., early warning systems, context-aware computing) that require a fuller modeling approach able to capture subtle aspects of objects and events, as of well as the multiple relations involving them [6, 9]. Moreover, I argue that such an approach should be based on an in-depth ontological analysis of the nature of these entities. In particular, a notion that deserves the conceptual clarification afforded by such an ontological analysis is that of *context*.

From an object perspective, we seldom interact with these entities *qua-themselves*, but we frequently conceive objects *qua-playing-certain-roles* in given "contexts" [7]. For example, most of our interactions with other human beings and, hence, our conceptualizations of these interactions are thought in terms of *roles* such as parent, employee, student, president, citizen, customer, etc. Analogously, when thinking about, for instance, cars, we think about them as means of transportation, insurable items, work-related resources, product offerings, etc. Moreover, we often conceive these "contexts" as *relational* ones [5, 7]: marriages, employments, enrollments, and presidential mandates are themselves concrete "object-like" entities that define a scope in which ordinary objects play complementary roles interacting with each other. Furthermore, these relational entities are constituted by other dependent "object-like" entities (qualities and dispositions) [5] that delimit the properties (e.g., commitments, claims, capacities, powers) that ordinary objects can exhibit in the scope of a given role.

From a behavioral perspective events themselves can also be framed in certain "contexts". In the most obvious way, this refers to complex events of which more basic events can be part (e.g., "that talk happened in the context of that conference") [11]. Moreover, "event contexts" can also refer to certain *scenes* [6] (e.g., a lunch meeting in the presence of a number of other happenings in a restaurant), and *situations* [1] (e.g., "Martin Luther King marching while Lyndon Johnson was the president of the USA"). Finally, there are entities that, while not mereologically related a particular event, do directly influence its manifestation (e.g., "the rain falling during a football match", "the turbulence during a flight", "the headache during a meeting"), thus, in a sense, "contextualizing" that event.

In this talk, I also discuss the ontological nature of a number of these entities including substantials, qualities, dispositions, relationships, events, roles, and scenes. This is done in light of the Unified Foundational Ontology (UFO) [8, 10, 11]. In doing that, I propose some (non-exhaustive) interpretations for the overloaded term "context" when applied to ordinary objects and events dealt with by conceptual modeling. I then discuss the impact of the behavioral vs. structural divide in that field. Finally, I demonstrate how an ontological analysis and conceptual clarification of the nature of these entities can provide the foundations for a fuller conceptual modeling approach, needed for modeling complex domains [2, 9].

Acknowledgment. A significant part of the work reported has been jointly conducted with Nicola Guarino, and with João Paulo Almeida. I am indebted to them for many years of fruitful collaboration.

References

1. Almeida, J.P.A., Costa, P.D., Guizzardi, G.: Towards an ontology of scenes and situations. In: 2018 IEEE Conference on Cognitive and Computational Aspects of Situation Management (CogSIMA), pp. 29–35. IEEE (2018)
2. Almeida, J., Guizzardi, G., Almeida, J., Falbo, R.A.: Events in ontology-driven conceptual modeling. In: Proceedings of 38th ER (2019)
3. Casati, R., Varzi, A.: Events. In: Zalta, E.N. (ed.) The Stanford Encyclopedia of Philosophy, winter 2016 edn. Metaphysics Research Lab, Stanford University (2015)
3. Bock, C., Odell, J.: Ontological behavior modeling. J. Object Technol. **10**(3), 1–36 (2011)
4. Cohn, D., Hull, R.: Business artifacts: a data-centric approach to modeling business operations and processes. IEEE Data Eng. Bull. **32**(3), 3–9 (2009)
5. Guarino, N., Guizzardi, G.: We need to discuss the relationship: revisiting relationships as modeling constructs. In: Zdravkovic,J., Kirikova, M., Johannesson, P. (eds.) CAiSE 2015. LNCS, vol. 9097, pp. 279–294. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-19069-3_18
6. Guarino, N., Guizzardi, G.: Relationships and events: towards a general theory of reification and truthmaking. In: Adorni, G., Cagnoni, S., Gori, M., Maratea, M. (eds.) AI*IA 2016. LNCS, vol. 10037, pp. 237–249. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-49130-1_18
7. Guizzardi, G.: Agent roles, qua individuals and the counting problem. In: Garcia, A., Choren, R., Lucena, C., Giorgini, P., Holvoet, T., Romanovsky, A. (eds.) SELMAS 2005. LNCS, vol. 3914, pp. 143–160. Springer, Heidelberg (2006). https://doi.org/10.1007/11738817_9/11738817_9
8. Guizzardi, G.: Ontological foundations for structural conceptual models. Telematica Institute/CTIT (2005)
9. Guizzardi, G., Guarino, N., Almeida, J.P.A.: Ontological considerations about the representation of events and endurants in business models. In: La Rosa, M., Loos, P., Pastor, O. (eds.) BPM 2016. LNCS, vol. 9850, pp. 20–36. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-45348-4_2
10. Guizzardi, G., Wagner, G., Almeida, J.P.A., Guizzardi, R.S.S.: Towards ontological foundations for conceptual modeling: the Unified Foundational Ontol. (UFO) story. Appl. Ontology **10**(3–4), 259–271 (2015)
11. Guizzardi, G., Wagner, G., de Almeida Falbo, R., Guizzardi, R.S.S., Almeida, J.P.A.: Towards ontological foundations for the conceptual modeling of events. In: Ng, W., Storey, V.C., Trujillo, J.C. (eds.) ER 2013. LNCS, vol. 8217, pp. 327–341. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-41924-9_27
12. Lankhorst, M.M. (ed.): Enterprise Architecture at Work - Modelling, Communication and Analysis, 4th. Springer (2017)
13. Meyer, A., Weske, M.: Activity-centric and artifact-centric process model roundtrip. In: Business Process Management Workshops - BPM 2013 International Workshops, Beijing, China, 26 August 2013, Revised Papers, pp. 167–181 (2013)
14. Nigam, A., Caswell, N.S.: Business artifacts: an approach to operational specification. IBM Syst. J. **42**(3), 428–445 (2003)
15. Olivé, A., Raventós, R.: Modeling events as entities in object-oriented conceptual modeling languages. Data Knowl. Eng. **58**(3), 243–262 (2006)
16. Zachman, J.A.: A framework for information systems architecture. IBM Syst. J. **26**(3), 276–292 (1987)

Contents

Nonlinguistic Aspects of Linguistic Contexts

Margherita Benzi[1](✉) and Carlo Penco[2]

[1] DIGSPES and Center for Artificial Intelligence AI@UPO, University of Eastern Piedmont, Alessandria, Italy
margherita.benzi@uniupo.it

[2] DISPO and Genoa Cognitive Science, University of Genoa, Genoa, Italy
penco@unige.it

Abstract. Our paper works on a proposal recently put forward by Hunter, Asher and Lascarides (2018) on the use of events in discourse context. We basically accept their view and their proposal of using events as explanation in discourse context. However we think that a stricter connection with demonstrations and causal reasoning in everyday conversation is a necessary step in a coherent view of discourse context. We will not deal with any particular formalism, but only with the general problem of taking into account some elements that may simplify or explain what is taken for granted in some steps of our inferences. A central concept used in these setting is the concept of "explanation" as a way to give coherence to the discourse context. This kind of explanation is also based, besides elements of a general encyclopedic knowledge, on default assumptions derived by the ontology present in the lexicon as Asher (2011) has abundantly shown. However, the steps to recover such coherence would gain clarity with a better specification of causal explanation and with a more precise account of the relation between demonstrative and demonstrations in discourse context. On these two aspects we give some suggestions.

Keywords: Discourse context · Demonstratives · Gestures · Causation

1 Nonlinguistic Aspects in Linguistic Context

Since Davidson, many authors tried to insert "events" in semantics[1] and Hunter et al. (2018) (from now on: HAL) made a recent attempt to insert events in Discourse Context[2]. Events apparently are non-linguistic elements. This brings about a general and a specific problem:

---

[1] The most recent attempt has been given by Perry (2019).

[2] Discourse context may be linked to cognitive context as discussed by Stalnaker within the framework of possible worlds semantics, but may use different formalisms. We may also consider discourse context as an organized sequel of utterances. In this way we may theoretically connect the notion of discourse context to the notion of context of utterance in the strict sense (parameters of speaker, time, location and possible worlds) and in the wide sense (including cognitive aspects). The framework actually used by HAL is Segmented Discourse Representation Theory (Lascarides and Asher 2008) where they apply what they call "glue logic", which is a kind of commonsense reasoning using rhetorical relations. Rhetorical relations are what connect different stages in a discourse, and different stages may also be identified with different

G. Bella and P. Bouquet (Eds.): CONTEXT 2019, LNAI 11939, pp. 1–13, 2019.
https://doi.org/10.1007/978-3-030-34974-5_1

(i) How do we use nonlinguistic elements in the context of utterance?
(ii) How may we treat events in discourse context?

As (i) is concerned, we may find an ancient source of the problem in the foundational period of semantics, with Frege's remarks concerning nonlinguistic elements. Discussing complex demonstratives Frege (1918) claimed that in saying "this man" we form a singular term that is composed of a linguistic part (the demonstrative "this" and the common name "man") and a nonlinguistic part, which is the gesture accompanying the expression. After Frege many authors discussing demonstratives made similar claims (Künne 1992; Burge 2005; Kripke 2008; Penco 2013; Textor 2015), which can be summarised – *ceteris paribus* – as treating the use of a demonstrative as analogous of an ordered pair with a linguistic part (the demonstrative expression) and a nonlinguistic part (the demonstration or gesture). However other hints by Frege have been often disregarded, and they concern not only the pointing gesture often accompanying the use of a demonstrative, but other nonlinguistic aspects of an utterance. Concerning the difficulties of interpreting natural language, discussing sentences containing context dependent expressions, like the indexicals "I", "today", "here", Frege (1918: 64) said:

> "In all such cases the mere wording, as it can be presented in writing, is not the complete expression of the thought; the knowledge of certain conditions accompanying the utterance, which are used as means of expressing the thought, is needed for us to grasp the thought correctly. Pointing the finger, hand gestures, glances may belong here too."

Following a previous remark:

> "A sentence like 'I am cold' may seem to be a counterexample to our thesis that a thought is independent of the person thinking it, insofar as it can be true for one person and false for another, and thus not true in itself. The reason for this is that the sentence expresses a different thought in the mouth of one person and false in the mouth of another. In this case the mere words do not contain the entire sense: we have in addition to take into account who utters it. There are many cases like this in which the spoken word has to be supplemented by the speaker's gesture and expression, and the accompanying circumstances" Frege (1979: 124).

Frege gives a relevant suggestion for taking into account different features that enter linguistic communication: co-speech gestures such as ponting, gaze and the position of the body in space are essential for expressing and understanding deixis. We aim to give a wider interpretation of co-speech gestures, typically used together with demonstratives (and indexicals), where gestures are not restricted to pointing, and can be collected under

---

utterance contexts. The dynamic of discourse takes care of anaphoric relations among elements of the discourse and is often represented with updates that restrict the set of worlds. Meaning in dynamic semantic is therefore typically given as "context change potential", but this new aspect of meaning does not abolish the idea of what is said as truth conditional content, which is the typical approach of the classical presentation of the context of utterance. Discourse context is a linguistic context where elements of the lexicon contain information plugged in the ontology of language (as widely analysed by Asher 2011).

the general term of "demonstrations", including different kinds of attention guiding movements or postures.

As (ii) is concerned, we just said that there is a long tradition of inserting events in semantics, which is developed in different ways that give events fundamental role in semantics, first of all because, as John Perry recently suggested, events are the truth makers of propositions or thoughts:

> "Thoughts are made true by events in the physical and mental realms that occur at various times" (Perry 2019).

But there is more: as HAL suggest, events have a role in discourse interpretation as something that role is not always explicitly stated and needs to be derived by the context of the conversation. In order for events to be the truth makers of our assertions we need to find a path to insert into discourse context events that are not explicitly expressed.

HAL aim at studying the interaction between nonlinguistic events and discourse interpretation, in two directions of contextual influence: "the effects of nonlinguistic events on discourse interpretation and the effects of discourse structure on the typing or conceptualization of nonlinguistic events". Although they mainly discuss the first direction, it is apparent that the two directions are strictly connected and the effect of events on discourse interpretation requires acknowledging the type of event in question.

In this paper we basically accept HAL proposal and their discussion of the role of events in understanding and interpreting discourse. However we think that they partly overlook two aspects connected to the general questions posed above and in particular: (i) the role of demonstrations as nonlinguistic *essential* parts of expressions and (ii) the role of causal reasoning as *essential* part of individuation of events to give coherence to discourse context. We think that a stricter connection with demonstrations and causal reasoning in everyday conversation is a useful and perhaps necessary step to ascertain the coherence of discourse context.

We will not deal with any particular formalism, but only on the general problem of taking into account new elements that may sexplain what is given for granted in some steps of HAL's presentation of their assessment of the insertion of events in discourse context. What HAL call "explanation" refers to how people infer what appear to be causal relations that build up the coherence of a discourse.[3] This kind of explanation seems also based on default assumptions derived by the ontology present in the lexicon (see Asher 2011) and from general encyclopedic knowledge. Notwithstanding our basic agreement, we think that the paper suffers of an incomplete characterization of the semantic and explanatory role of gestures, and that it suffers of some vagueness about the mechanism that triggers the search of (causal) explanations for recovering the discourse coherence. In what follows we will discuss, first, the role of nonlinguistic gestures in treating the

[3] We cannot summarize HAL's paper, but it is relevant to consider that explanation is a semantic relation such that "if the content *p* of a discourse move *m* stands in an Explanation relation to the content *q* of a discourse move *n* such that *p* provides the explanans, then the raison d'être of *m* is to provide an explanation of *q*." This gives a less prominent role to intention: an Explanation can be inferred on the basis of features of *p* and *q* and from there an interpreter "can defeasibly infer that the speaker had the intention of using *p* to explain *q*" (p. 18).

different targets of demonstratives (be they a scene, an object or an event) and, later, the role of gestures and events in building explanations in discourse context.

2 Inserting Demonstrations in Context

Let us use as a test case one of HAL's examples. We start with a situation whose elements are two parents, Ann and Peter, their daughter and some scratches on the wall. In this context HAL make us imagine two different cases, with two different utterances made by Anne.

Anne utters

(1) "Our daughter was sent to her room for *this*. It happened while I was cooking dinner."
[While pointing at the scratches on the wall]
(2) "Our daughter was sent to her room"
[and she nods suggestively over her shoulders so that Peter may see the scratches on the wall]

In case (1) "this" is a demonstrative linked to the pointing gesture, and refers both to the scratches, and to the *event* that caused the scratches. Here "this" may mean "the scratches and the event that the scratches was produced by the daughter". It explicitly explains what provoked as consequence the sending of the daughter to her room and punishment of having done damage to the house (but how much is embedded in language to make this kind of reasoning!). Besides, the second sentence in (1) gives a precise time for the event.

In case (2) there is no expression that we can associate with the nonlinguistic event of the daughter scratching the wall. "Nor is there an expression that triggers a search for a nonlinguistic entity." (HAL: 4). Is that true?

Saying "[she] nods suggestively over her shoulder" implies a position in space and a direction of gaze that triggers the attention of the hearer towards the scratches on the wall. Actually we may include in conversation not only conventional pointing gestures, but also – following the lead of Frege's suggestions – gaze and positions in space. We suggest to take into account different ways of referring to objects, scenes and events: typically we use demonstratives completed by demonstrations, but also demonstratives alone sometimes also perform the role of pointing gestures in context, and sometimes demonstrations alone (including postures or placements in space) may take the place of the use of demonstratives.

Utterances may be about situations, as a long tradition from Austin to Barwise and Perry confirms, and the role of demonstratives is highly context dependent because they take their reference in the context of a situation. Stojnic et al. (2013) discuss the concept the "situated utterance",[4] which may help clarifying the use of context-dependent demonstratives. They propose the example of a situation in which a master chef is preparing a wonderful omelet. In this situation, the utterance:

[4] Stojnic et al. (2013) seem to treat "situated utterance" as something special, but every utterance is connected to a situation, unless we think of utterances concerning logical or mathematical formula (but even there we need to refer, at least, of the kind of theory we are using: in elementary

"that's an omelet"

makes it relevant the reference of "that" even without pointing.

This aspect is worth special consideration, given that Kaplan himself, the founder of a proper semantics of demonstratives, abandoned the idea that gestures are a constitutive part of demonstratives to adopt a radical intentional perspective, followed by most philosophers (last, but not least Neale 2016 or Speaks 2016). There were different reasons why Kaplan abandoned the idea of demonstrations as fundamental for the use of demonstratives:

(a) The possibility of demonstratives without demonstration
(b) The possibility of sortal confusion
(c) The lack of syntax of gestures

Stojinc et al. (2013: 505–518) make a strong criticism to the intentionalist stance, dismantling the argument based on (a) put forward by Kaplan. They agree with Kaplan that there may be demonstratives without demonstrations. However, against Kaplan, they claim that it is not the intention to referring to an omelet that disambiguates the reference; on the contrary it is the coherence of the discourse in the real world situation that makes the reference mandatory. They insist on the insertion of updating discourse on the ground of events and situations so that "such updates can capture the interpretation of demonstratives when there's no explicit pointing or demonstration in the utterance".

Against the idea that the only means to explain the source of reference in the use of indexicals is the intention of the speakers, Stojnic et al. (2013, p. 505) insist that reference of demonstratives is given by a linguistic, grammatically encoded mechanisms that "update and access an appropriately structured context". Besides, they argue that:

> "deictic utterances of demonstratives accompanied by the act of demonstration—the pointing gesture and its analogs—is itself a grammaticized constituent of the speaker's utterance that, together with the linguistic meaning (the character) of the demonstrative pronoun, determines the referent on an occasion of use."

---

arithmetics we may use natural numbers or integers or rational or real numbers and everytime we change the rules of the game). There are many representations of situations. Gauker (2012) suggests distinguishing "context" as a formal representation of a domain and "situation" as the actual state of affair pertaining to an utterance. McCarthy used the term "context" to define a situation as a triple containing a domain, axioms (describing the relation of the elements of the domain) and rules of inferences. Kaplan intended context of utterance in narrow sense as a set of parameters: speaker, time, location and possible world. Context of utterance in a wide sense includes other relevant information (often including presuppositions) necessary to disambiguate the reference of demonstratives and other non-automatic indexicals. Barwise-Perry with others followed Austin's idea, used by Stojnic et al. (2013) for speaking of "situated utterance". Let us say that the "context of utterance" is in principle always situated, and may be defined with an increasing and indeterminate degree of precision. Interpretation of what is said in a context should amount to defining the truth conditional content of sentences in context or utterances (and we may have different levels of truth conditions, from reflexive truth conditions to subject matter or referential truth conditions – Korta and Perry (2011), Da Ponte et al. (2019)).

Shortly, intentionalist arguments based on (a) do not work because what counts for the determination of the reference of a demonstrative is not just the intention, but a grammatical mechanism grounded in context.

But what about other arguments in favour of Kaplan's intentionalist stance? Regarding (b), the possibility of sortal confusion (when I point to a person it is not clear what I want to refer to: the person, her clothes, the colour of her clothes, her hair, and so on) is a standard problem since Wittgenstein (when he reminds us that we need to know a lot – for instance classifying things – before giving an ostensive definition pointing towards something).

Indeed, the possibility of sortal confusion is always present, but why should it be suggesting an intentionalist stance? It appears an indication that any use of demonstratives requires a sortal concept together with the demonstration. And the sortal concept is indeed something that derives by the discourse context and by the situation at hand. The intention of the speaker should be such that the audience can grasp it in context, given the requirement of a (possibly implicit) sortal that situation and discourse put into focus.

About (c), what was missing at the time of Kaplan's paper was the idea of a "syntax of gestures", the lack of which was a motive for which Kaplan abandoned the idea that gestures are essential for the use of demonstratives. However, while "syntax of gestures" was almost completely missing at the time of Kaplan's first work on demonstratives (published in 1989, but written in 1977), now we are in a very different situation. In recent times much has been done on this aspect. Among many attempts, also Lascarides and Stone (2009) show a very sophisticated logic of gestures, starting from a basic distinction between "identifying gestures" and "visualizing gestures", the former intended to demonstrate object and the latter to depict some aspects of the world. They acknowledge that "speakers can use facial expressions, eye gaze, hand and arm movements and body posture intentionally to convey meaning" (p. 393), but in their attempt of formalising gestures, they ignore body postures and facial expressions (p. 444). Probably this is due to the fact that facial expressions are typically studied as expression of emotions. We think this is a limitation of their work, which would be enriched and made more coherent with the analysis of demonstratives with the insertion of a more general view of what typically accompanies demonstratives in conversation: demonstrations as a wide category that includes not only pointing gestures, but also different kinds of bodily movements that perform the function of gaining joint attention: placements in space, facial expressions, including the direction of gaze as joint attention devices. An aspect of facial expression that is fundamental as "identifying gesture" is the direction of gaze. On this we have now many studies, among which a recent work on gaze recognition in robotics that has given surprising results on the relevance of interpreting actions and intentions of humans only checking the direction of the gaze (Palinko et al. 2016). This seems a confirmation of the communicative role of gaze direction not only as a completion of demonstratives, but as a possible proxy of the uses of demosntratives.

Lascarides-Stone (2009) distinguishes identifying gestures and visualising gestures; it seems that they consider bodily postures only strictly connected with visualising gestures and expressing different emotive aspects. On the contrary, it is reasonable to claim that the main cognitive role of body posture in shared space is a form of identifying gestures, as most studies on deixis show. We may call all identifying gestures or deictic gestures as "demonstrations", including the different ways in which we use our body to disambiguate the referent of demonstratives (and often pronouns), and sometimes we may use just demonstrations as a proxy for bare or complex demonstratives. As studies in linguistics remind us, demonstratives are a fundamental linguistic stratum in all indoeuropean languages, they form original linguistic stratum which is connected with actions in situations, where gestures and recognition of movements and body postures are central to help people to converge to joint attention (see Borghi-Penco 2017).

The point we want to make is as follows: once we enter a situation, positions in space and gaze, together with other kinds of demonstrations, may be considered, as Frege did, as a proper part of a linguistic exchange, *a nonlinguistic part of a linguistic practice*. Pointing gestures and their analogs are not only a nonlinguistic device used to *supplement* demonstratives lile "this" or "that"; they may be a proper fundamental part of linguistic communication. Following Stojnic et al. (2013), we claim that the reference of demonstratives derives from a grammatically encoded mechanism. This mechanism, however, can be triggered by demonstrations alone, as the case (2) presented above seems to show. To get over the criticism of intentionalists we may take what they conceive as a weakness of demonstratives (sortal ambiguity) as a requirement that for each demonstrative action we need an ordered pair of a demonstration and a predicate (or some descriptive content). Only sometimes demonstrations and predicates are explicit, and most often they are recovered by the context, when a particular movement or position of a speaker triggers the direction towards the intended scene, as in example (2) from HAL, and a predicate is derived by the setting of the particular situation (the scratching on the wall). Considering bare demonstratives as intrinsically connected with a demonstration-predicate pair, we might think of them not so differently from the use of referential descriptions (see Benzi-Penco 2018).

We suggest therefore that what Lascarides and Stone (2009) call "identifying gestures" be intended in the wide meaning of "demonstrations" given above. Demonstrations may either pick an individual or a state of affair or event. In the first case the pair of a demonstration and a predicate in context will suffice to determine the reference of the indexical (either expressed linguistically or hidden in the actual demonstration). In the second case directing the attention to a state of affairs or event (in our case the scratches on the wall) points to a possible role of the nonlinguistic feature (the gaze towards the scratches on the wall) in discourse context: this role may be either directing the attention to an event as a cause of some other event under discussion, or to an event or state that invites a causal explanation (which is the case proposed by HAL in the example in their paper). Even when we do not explicitly refer to actions or events that cause certain states of affairs, pointing to an event or state often compels us to bring about an explanation about which other event provoked that state or event itself.

3 Causal Default Reasoning in Context

This brings us to the second step: HAL claim that "coherence-based reasoning about the connection between Anne's signals and the wall must account for the inference that there is a salient event that caused the (visible) scratch on the wall." This is a wonderful suggestion, but we miss the chain of reasoning that brings to the salient event. Where from does the recognition of the event come? HAL suggest the coherence relation of *explanation*. What kind of explanation? We need an explanation of the derivation from the state of affairs or simple event (the scratch in the wall)[5] and the event that provoked it, which is the main point of their paper. However, it is not clear from the paper how we get the salient event: from intuition? From notions embedded in the lexicon? From general encyclopedic knowledge? From probabilistic reasoning?

Let us go back to our two original examples of Ann's utterances (forget the second sentence of the first case):

(1) "Our daughter was sent to her room for *this*." [pointing to the wall]
(2) "Our daughter was sent to her room." [nodding towards the wall]

We claimed that there is no such a big difference between (1) and (2) because the role of a demonstrative pronoun may be taken by a demonstration alone, where a nonlinguistic element – a demonstration – is part of the expression of a thought, and sometimes is a proxy for the linguistic part of the expression too (Anne's nodding is a proxy for a demonstrative with a pointing gesture towards the scratches in the wall). In both cases it is apparent that the attention of the hearer is brought towards something visible and clearly abnormal. In the representation of the situation we need to have the state of affair "scratches in the wall" even if these are not part of any expressions used by Anne, but belong to the general description of the situation. Granted that, we need to analyse the steps in commonsense reasoning (or, for HAL in "glue logic") that lead to the explanation of Anne's statement. Let us see how HAL put the matter:

Assuming a standard first order dynamic semantics enriched with:

$\pi_1, \pi_2, \ldots$ as elementary discourse units (EDU)
$\varepsilon_1, \varepsilon_2, \ldots$ as elementary event units (EEU)
$i_1, i_2, \ldots$ as agents (speakers of EDU or actors of EEU)
d standing for Ann's daughter
CDUs as complex descriptions units made by more than one EDU or EEU.

We have, from the point of view of Peter, two segments of discourse, the first that the daughter has been sent to her room and the second that Ann made a gesture towards something (or pointed to something saying "for this"). The first relevant step is to define the reference of Ann's gesture towards the wall. We may assume that the salient visual pattern is what is unexpected or abnormal, that is scratches on the wall and the elementary event unit (what we called state of affairs) can be represented approximately as:

$$\varepsilon_1 \; \exists x \; \exists y. \; (on \; (\varepsilon_1, x, y) \; \& \; wall \; (y) \; \& \; scratches \; (x))$$

In this way we may represent the referent of Anne's demonstration, but we are still left without an explanation of

(i) Why Ann referred to ε_1 and
(ii) Why she sent her daughter in her room.

[5] We use "state of affairs" or "singular event" or "event" interchangeably, for the sake of simplicity.

We need a conceptualization and an explanation of (ii) and HAL comment as follows:

> "By hypothesizing an Explanation, the interpreter accords a high probability to Anne's committing to a particular conceptualization of the scratches as the outcome of a nonlinguistic event in which Peter and Anne's daughter caused the scratches. This inference leads to the construction [....] of a CDU containing both the causing event e2 and the scratched state of the wall and a Result relation between them expressing their causal dependency."

The resulting complex event unit, connecting different EEUs, that is nonlinguistic aspects of the discourse, would take a form of the following kind:

$$\varepsilon_3 \colon (\varepsilon_1 \colon \Phi\varepsilon_1 \;\&\; \varepsilon_2 \colon (activity\ (\varepsilon_2) \;\&\; agent\ (\varepsilon_2, d)) \;\&\; Result\ (\varepsilon_2, \varepsilon_1))$$

The event unit represent the activity of the daughter as the event that produced the scratches on the wall. This presentation of a nonlinguistic event would be the complete reference to the demonstrative gesture of Ann towards the wall and at the same time an explanation of why she sent her daughter to her room.

Certainly there is nothing wrong with these passages. They seem to be very intuitive. However, the background presuppositions at stake are really very complex. How much background information should we put in our ontology in cases so specific as the one under consideration? How to answer the question: why a daughter is sent to her room? A daughter is somebody on which parents have an authority. Scratching the wall is an action that is judged as bad, because it causes damage on something which has a value. If somebody makes a bad action must be punished (see Asher and Lascarides 2003). Being sent to her room is a kind of punishing of the daughter because it restricts her freedom of movement. If the daughter has been punished she must have performed a bad action. Scratching the wall is a bad action. Therefore, the daughter was the author of scratching the wall. The daughter scratching the wall is an event that caused what was referred to through a visual pattern (scratches in the wall) to which Anne's demonstration referred. All this reasoning is partly circular and it is not clear what explains what.

Does all this chain of reasoning belong to a needed common-sense reasoning? How can be implemented in a formalization of discourse context? Should we really simulate all the steps of a possible chain of inference of the interlocutor? But why not other arguments, like sending the daughter to her room in order to let her avoiding the sight of some strange scratches on the wall? Therefore sending the daughter to her room would not be a punishment, but prevention for safety? How much cultural encyclopedic knowledge is required?

Can we find a "regular" shortcut that makes the inferential steps simpler? Certainly, there must be a causal reasoning that gives an explanation why a certain state of affairs was brought about and by whom. The inference must be a causal inference. If we need to make events enter semantics, we need a link between states of affairs and events that cause those states. The coherence of the discourse should rely on causal inference, together with default reasoning given the few elements of the discourse at hand. Events should be causally connected with elements of the proposition expressed by Anne's sentence (and we may put inside those elements also nonlinguistic part of (possibly implicit) demonstrative pronouns).

In the case presented by HAL, the speaker refers to a state of affairs, and therefore *indirectly* to what or who caused it. The speaker also refers to another human, in this case the daughter, and therefore she is the most probable actor or cause of the state of affairs.

The standard counterfactual analysis of causation claims that *a* causes *b* when, if *not a*, then *not b*. Here the situation is very basic: we have

b = "scratches in the wall"
a = "x scratching the wall".

The only individual *x* of whom the interlocutors speak is the daughter. Which probability has the daughter to be the subject of "*x* scratching the wall"? To make this step of inference perspicuous we should put it in some normal form. Our very simple suggestion is to insert in the generation of the cause a basic inferential step of *inference to the best explanation* (what is a logical fallacy in classical logic). Given the discourse context and the elements we have we should have some simple steps, on the assumption that scratching the wall is an action worth a punishment:

(x) (*Scratching the wall* (x) → *Punish* (x))
Punish (d)

Scratching the wall (d)

The conclusion is the best explanation of the fact that *d* has been punished. Our only worry is how can we build an explanation of the relations among gestures, parts of discourse and events without the need to put forward all the background information. We may rely on an implicit presupposition trigger: every time we refer to an event we implicitly refer to its cause. In this case referring to scratches in the wall (ε_1) triggers the event of scratching the wall by the daughter (ε_2), which – at the same time – is the event that caused "daughter sent to her room" or "daughter punished". We might say that the event unit may be considered among the weak truth makers of the proposition "daughter sent to her room" for their connection with the demonstration towards the scratches on the wall.[6]

However, in this case, following HAL, the simplest line of explanation is the direction from the event to the cause: scratches on the wall presuppose that somebody made them, like a broken vase presupposes somebody broke the vase. It is like diagnostic reasoning, which aims at causal explanation. Where a causal explanation is concerned, the reasoning goes "from effects to causes", and the notion of causality is a notion of singular or actual causation, on which there is an abundant literature[7]. And here a presuppositional analysis of causality might open new directions of work.

What is needed is some causal machinery that should be activated when there is a missing link between an event and an unknown cause. The steps that HAL take for granted on the ground of intuition and knowledge of human habits in a family (a very

[6] This is only a hint to be developed; we refer here to an interesting suggestion by Guarino et al. (2019), on the requirement of causal connectness for understanding the meaning of a sentence.
[7] See for instance the references in Benzi (2016).

specific local information) may also be triggered by a more general causal model that takes as input a state of affairs (or an elementary event unit) and as output the most probable cause in the context of discourse. The suggestion is just a proposal to connect together different fields of research that have already solid results. This step would help avoiding complicated interconnected inferences with a wide background not everybody may share, and positing a simple presuppositional trigger of causal relations may help generalising to many different situations.

4 Conclusions

We propose two main integrations of Hunter-Asher-Lascarides proposal of the treatment of understanding dialogue in context. Both concern nonlinguistic aspects in linguistic contexts (dialogue contexts), but with two very different concerns. In the first case we deal with non-linguistic aspect of the expression of a thought; in the second case we deal with the ability of inferring non-linguistic components that permit understanding a dialogue context.

In Sect. 2 we suggested that nonlinguistic demonstrations be considered a proper part of the *expression* of thoughts, as a quasi-linguistic part of linguistic expressions. The correct intuition given in Hunter-Asher-Lascarides (2018) about the rendering of a bodily posture of the speaker should be supported more strongly by an analysis of the role of demonstrations in identifying objects or states of affairs. The wide proposal presented by Lascarides and Stone (2009), although very well refined, should be implemented by an analysis a special set of co-speech gestures: demonstrations occurring with demonstratives (and, although rarely, sometimes – for short – as proxy for demonstratives), among which bodily postures and direction of gaze. An analysis of this specific kinds of demonstrations would help improving the interaction between formal representation of discourse contexts, utterance context and robotics, where the interpretation of gaze in interaction with humans is highly developed (Palinko et al. 2016). In Sect. 3 we suggested that an intuitive version of coherence relations in discourse context should be enriched by a mechanism of causal reasoning working like a presupposition trigger: when we refer to a state of affair we indirectly refer to the cause of that state or event. This solution would help in different contexts of interaction (for instance in interaction with medical problems). Our contribution does no amount to propose a new formalism or a new setting, but it is a suggestion of making works used in different fields compatible one another. We have a lot of results in many fields that often do not interact each other and a new interaction, although not easy at first, may be find fruitful for connecting different enterprises towards common goals.

The two parts of our analysis seems somehow unrelated. However there is a common core supporting both of them: the ways we connect linguistic expressions with object and events are basically given by the demonstrative stratum of language. Referential uses of demonstratives are strictly connected with a nonlinguistic part of the expression of our thought: demonstrations that often accompany demonstratives and sometimes also perform the role of demonstratives by being used alone, leaving the linguistic feature (like “this” or “that”) implicit. However, as we have discussed in the last part of the paper, those demonstrations are not only a device for joint attention towards a scene or an event, but in some context requiring explanation, they point both to an event and to its cause. This is really a new perspective on which to work and it is the merit of HAL to let it come out. The nodding of the case (2) we discussed above is not a simple nonlinguistic gesture. It is a kind of demonstration that requires a linguistic background to be properly performed and understood with its causal connections. In the first dialogue Anne says "for *this*" explicitly using a demonstrative as explanation (where "for this" can be translated with "*this* is the cause of …"). In the second there appears to be no linguistic means that may help deriving an explanation, and we are left with a mere gesture, a nodding. But this nodding may be interpreted in contexts as a highly sophisticated form of expression. The point is that, only once we have acquired language, we may use gestures for describing and hinting at connections between events and causes. In these cases, gestures are a shortcut or a proxy for more complex expressions and arguments we have to derive from the context.

References

Asher, N., Lascarides, A.: Logics of Conversation. Cambridge University Press, Cambridge (2003)
Asher, N.: Lexical Meaning in Context. Cambridge University Press, Cambridge (2011)
Benzi, M.: Thought experiment and actual causation. Topoi (2016). https://doi.org/10.1007/s11245-016-9427-7
Benzi, M., Penco, C.: Defeasible arguments and context dependence. Paradigmi **36**(3), 561–578 (2018)
Borghi, G., Penco, C.: Kaplan's sloppy thinker and the demonstrative orginis of indexicals. Quad. Semant. **3–4**, 117–137 (2017)
Burge, T.: Truth, Thought and Reason. Oxford University Press, Oxford (2005)
Da Ponte, M., Korta, K., Perry, J.: Utterance and context. In: Ciecierski, T., Grabarczyk, P. (eds.) The Architecture of Context and Context Sensitivity. Springer, Dordrecht (2019)
Frege, G.: Der Gedanke. Beiträge Zur Philosophie Des Deutschen Idealismus I, 58–77 (1918)
Frege, G.: Postumous Writings. Blackwell, Oxford (1979)
Gauker, C.: Semantics and pragmatics. In: Russell, G., Graff Fara, D. (eds.) The Routledge Companion to the Philosophy of Language. Routledge, London (2012)
Guarino, G., Porello, D., Guizzardi, D.: On weak truthmaking. In: Proceedings of the Joint Ontology Workshops (2019, in press)
Hunter, J., Asher, N., Lascarides, A.: A formal semantics for situated conversation. Semant. Pragmat. **11** (2018). https://doi.org/10.3765/sp.11.10
Korta, K., Perry, J.: Critical Pragmatics. Cambridge University Press, Cambridge (2011)
Kripke, S.A.: Frege's theory of sense and reference: some exegetical notes. Theoria (2008). https://doi.org/10.1111/j.1755-2567.2008.00018.x
Künne, W.: Hybrid proper names. Mind **101**(404), 721–731 (1992)
Lascarides, A., Asher, N.: Segmented discourse representation theory: dynamic semantics with discourse structure. In: Bunt, H., Muskens, R. (eds.) Computing Meaning, vol. 83, pp. 87–124. Springer, Dordrecht (2008). https://doi.org/10.1007/978-1-4020-5958-2_5
Lascarides, A., Stone, M.: A formal semantic analysis of gesture. J. Semant. **26**, 393–449 (2009). https://doi.org/10.1093/jos/ffp004
Neale, S.: Silent reference. In: Ostertag, G. (ed.) Meanings and Other Things - Essays in Honour of Stephen Schiffer. Oxford University Press, Oxford (2016)

Palinko, O., Rea, G., Sandini, F., Sciutti, A.: A robot reading human gaze: why eye tracking is better than head tracking for human-robot collaboration. In: IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 5048–5054 (2016). https://doi.org/10.1109/IROS.2016.7759741

Penco, C.: Indexicals as demonstratives: on the debate between Kripke and Künne. Grazer Philos. Stud. **88**, 1–11 (2013)

Penco, C.: Donnellan's misdescriptions and loose talk. In: Korta, K., De Ponte, M. (eds.) Reference and Representation in Language and Thought, pp. 104–125. Oxford University Press, Oxford (2017)

Perry, J.: Frege's Detour. An Essay on Meaning, Reference, and Truth. Oxford University Press, Oxford (2019)

Speaks, J.: The role of speaker and hearer in the character of demonstratives. Mind **125**(498), 301–339 (2016)

Stojnic, U., Stone, M., Lepore, E.: Deixis (even without pointing). Philos. Perspect. **27**, 502–523 (2013)

Textor, M.: Frege's theory of hybrid proper names extended. Mind **124**, 823–847 (2015)

User-Aware Comfort in Retail Environments

Nicola Bicocchi[1], Stephan Boese[2], and Giacomo Cabri[1](✉)

[1] Università di Modena e Reggio Emilia, Modena, Italy
{nicola.bicocchi,giacomo.cabri}@unimore.it
[2] GK Software, Schöneck, Germany
sboese@gksoftware.it

Abstract. A retail environment can be thought as an environment where customers can buy products, goods or services. The user-experience in physical retail environments is important not only for facilitating the selling of goods and services, but also for providing satisfaction and appealing to retain customers over the long term. The user-experience can be enhanced by adapting aspects of the physical environment such as music, colour, fragrance to the tastes of the customers.

In this paper we propose a user-aware approach to adapt physical aspects of a retail environment in order to improve the perceived comfort level. We introduce a model of the user context, which can be used both for representing information about the customers and for driving the adaptation of the environment. The proposed decision system is based on a microservice architecture providing both modularity and flexibility. Real-world examples are also used to show applications of the approach.

Keywords: User-awareness · Customer comfort · Software services

1 Introduction

When people enter in a retail environment, their decision about *whether* to buy and *what* to buy is influenced by several factors. Their needs and the appeal of the products on sale are the most important factors. Nevertheless, the environment can play a key role as well. For instance, it is possible to manipulate factors such as music, colour, fragrance, what is displayed on monitors, the advertisements, and so on; we call them *comfort aspects*.

We remark that the above-mentioned aspects can be configured for providing comfort in order to encourage customers to buy products or services in the environment. More importantly, the appropriate configuration of those aspects can provide customers with a good impression of the environment. The latter has several advantages:

– the customers feel satisfied of the time used for the visit;
– satisfied customers are likely to visit again;
– satisfied customers are the best advertisement to other potential customers;

G. Bella and P. Bouquet (Eds.): CONTEXT 2019, LNAI 11939, pp. 14–25, 2019.
https://doi.org/10.1007/978-3-030-34974-5_2

– the overall reputation of the environment improves.

In this paper we propose an approach for the automatic configurations of comfort aspects in retail environments. The contribution of this paper is twofold. On the one hand, we propose a model of user *context* based on *external*, *statistical* and *personal* information, which will be exploited by the decision system to set the comfort aspects. On the other hand, we propose a *microservice* architecture boosting modularity and flexibility, useful to adapt our approach to different scenarios. It is worth noting that we will not address neither *psychological* aspects, such as which music is the most appropriate in a given situation, nor *decision algorithms*, which can be chosen among those available in literature.

The rest of the paper is organised as follows. First, in Sect. 2 we describe related work in the field. Then, in Sect. 3 we show how we model the context of customers, i.e. how we organize the information that can be exploited to take decisions about comfort aspects. In Sect. 4 we describe the proposed architecture and provide a general description of the microservices used. In Sect. 5 we detail three case studies for showing real-world applications of the approach. Finally, Sect. 6 concludes the paper and draws final remarks.

2 Related Work

The issue of increasing the comfort level inside retail stores by means of digital technologies can be considered as an aspect of the digitalisation of our urban environments. Zambonelli et al. discussed this process in several works arguing the relevance of systems capable of taking both reactive and goal-oriented actions based on the user experience [5,14].

Betzing et al. [2] recently argued that customer experience theory insufficiently accounts for the transformative power of mobile technologies enabling digital and contextual services. Thus, they propose eight propositions to frame enriched customer experience enabled by mobile technologies. In line with these propositions, they also propose several design principles for shaping digital customer experience.

Meyer et al. [11] examine if there is a possibility to digitalise the retailers' advantage of personal advice, supported by contextual data, to achieve a closer and more personal digital relationship. Specifically, they explore the impact of various combinations of technology-enabled advanced services in customer-retail-relationships. Authors compare and evaluate a combination of both context-aware and emotional approaches/services discussing both online and offline advantages.

Bauer et al. [1] recently argued that the presence of digital signage showing emotional content creates favourable shopping experiences and positively influences consumer behaviour. Authors review recent findings related to digital signage in retailing and conclude that, leaving aside technical skills, knowledge from a number of fields (e.g., psychology, design) is needed for effectively deploy customer-ready systems.

Lamche et al. [10] investigated the customers' experience and context in e-commerce, in light of providing suggestions about what to buy. Authors enhanced a baseline recommender system by the integration of context conditions like weather, time, temperature and the user's company. These context conditions are embedded into the recommendation algorithm via pre- and post-filtering. However, the "comfort" issues are limited to the case of online or mobile shopping while our work is more focused on *physical* environments.

Finally, an important aspect to be considered for extracting and using customer data is *privacy*. We do not address privacy issues in this paper because we focus on the architectural aspects. Nevertheless, set aside further investigations, Schaub, Könings and Weber [12] have tackled the issue in the broader field of ubiquitous computing and their results can be applied to the specific case of retail environments as well.

3 Context Model

Taking decisions about which configuration of the comfort aspects could be more enjoyable for a specific set of costumers requires the modelling and understanding of costumers' preferences. Since there might be more than one customer inside the environment, with possibly different preferences, the decision system must also weight individual preferences to find optimal solutions.

We build the user context from different sources of information from which we derive three contexts: *External* context, *Statistical* context, and *Personal* context, as depicted in Fig. 1. Furthermore, Table 1 summarises the different kinds of contexts and reports their key features.

Table 1. Summary of the features of the three different context types: external, statistical, and personal.

Type of context	Depend on customers	Easy to achieve	Influence to comfort
External context	No	Yes	Low
Statistical context	Weakly	Yes	Somehow
Personal context	Strongly	No	High

The *external* context accounts for information unrelated with the customers such as weather, time of day, day of the week, temperature, location, and type of store. Of course, additional information can be taken into consideration.

These pieces of information can be useful to frame the general situation; for instance, in a spring day lights with warm colours inside the environment are suggested instead of cold colours; as other examples, a hard-rock music hardly convinces people to buy sweets and biscuits, as well as a classical music is likely to not provide comfort in a weapon store. This context is *easy* to gather, it could be retrieved even without the use of specific services. On the other hand, this

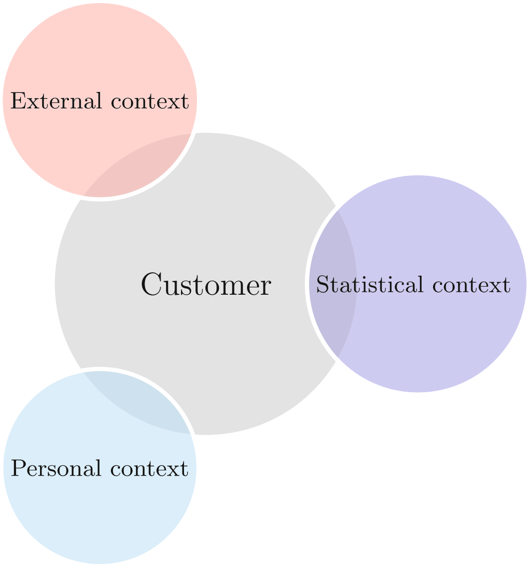

Fig. 1. Different kinds of contexts of customers.

context is the *least informative* one, since it does not concern the preferences of individual customers.

The *statistical* context concerns the distribution of the preferences in time (i.e., which situations are more comfortable in a given hour or in a given day), in space (i.e., which situations are more comfortable in a given place), and depending on other parameters (e.g., statistical distribution of age, gender, presence of families). For instance, people are likely to prefer lively music in the morning and calm music in the evening. In this case, the gathered information is *weakly* dependent on the people in the environment: the statistical approach tells us that customers *are likely* to have some preferences, but nothing grants they *actually* have them. This context is *easy* to gather, but requires specific services. Given it reports statistical preferences, it is *somehow influencing* the customers in the environment.

The *personal* context models information about specific customers such as age, gender, personal interests, habits, shopping history. Also in this case additional information can be considered depending on the store. This is the most dependent on the customers actually present in the environment, thus it can be *very influencing* on their own behaviour. On the other hand, it is the *most complex* to gather. The easiest way is to rely on a smartphone app, which can gather customer's preferences and information and can notify the environment about the customer's presence. Otherwise, a generic identification can be carried out by cameras or other devices, trying to estimate the age, the gender, the sentiment and other information about the customers in the environment. Needless to say, this kind of context might lead to significant privacy concerns.

Summarising, the proposed system is conceived to gather information and take decisions about the configuration of the environment accordingly. The information used is of *different kinds* (e.g., registry information, external information, contingent information) and is gathered from *different sources* (e.g., from services, from smartphone apps, statistical). The targets of the decision can also be different comfort aspects (e.g., music, colour, fragrance, advertisements).

As already mentioned, we do not adopt any specific decision algorithm because we focus on the general architecture. Thus, any decision algorithm can be actually implemented within the proposed system.

4 Architecture

The architecture of the proposed decision system is based on microservices[1]. Newman defines microservices as "*small, autonomous service that work together*" [12]. The microservice architecture enables the rapid, frequent and reliable delivery of large, complex applications. It also enables an organization to evolve its technology stack. More specifically:

– it enables a focused development, since each developer is devoted to one (autonomous) component;
– it allows for heterogeneous technologies to be adopted;
– it simplifies testing in the small;
– it simplifies the deployment, removal and maintenance of services.

Of course, they also have drawbacks, which must be carefully considered before and during development:

– the development of the whole system could be more complex, since they present the typical issues of distributed systems and there is no specific IDE available;
– a good and stable agreement about interfaces is needed;
– the testing in the large is more difficult;
– more resources needed, in particular, memory.

One of the main issues in adopting microservices is the definition of *which* services must be available and with which *granularity*. In our case, the microservices are divided into the following families, depending on their use:

– *decision* services, the core of the architecture, mainly the `Decision service`;
– *general purpose* services, which provide basic information, such as `Time service`, which provides the time of the day, `Location service`, which provides the location of the environment, and so on;
– *context* services, which provide information about the context of the customer and can be in turn divided into:
  • *external* services, such as `Weather service`, which provides information about the current weather given the location of the environment;

[1] https://microservices.io/.

  - *statistical* services, such as `Demographic service`, which provides statistical information about customers' age, gender, origin, ...;
  - *personal* services, which provide personal information about a user, identified in some way;
- *elaboration* services, which can elaborate the provided information, for instance to calculate averages;
- *decision support* services, which provide information useful to take decision, such as the `Weight service` that provides a table of weights;
- *actuator* services, which enact the decision, such as the `Music service`, the `Color service`, and so on.

All the available services can be orchestrated and coordinated to take decision about the *comfort aspects*. In the next section we will show some examples of microservice composition.

5 Case Study

In this section we present a case study to provide examples of the proposed approach. We consider a retail store in which a music is played in background; for simplicity's sake, we focus on music only, but the examples can be easily extended to other comfort aspects and their combinations. The aim of this case study is not about the kind of music to choose in a *specific* store, but to enable the choice of a given music depending on the context.

5.1 The Decision System

For concreteness's sake, we propose a decision approach taken from a previous work [6] relying on the following three features:

- *classification of customers*
- *definition of weights*
- *definition of a composition graph*

The idea is that, after the profile of the customers in the retail environment has been identified, a weight-based map is used for selecting the most appropriate value of the comfort aspect. If more than one comfort aspect have to be decided, a composition graph can be exploited to take a single decision considering not only the suitability of the single aspects, but also the suitability of couples of them (for instance, if a given music is suitable with a given fragrance).

We now consider the use case of a decision system used for configuring the music and the fragrance in a retail store.

The *classification of customers* can be made in different ways. For instance, customers can be classified based on their age, on their gender, on their social status, on their degree of loyalty to a brand, and so on. The system relies on a predefined classification of the customers to classify the actual customers in the

Table 2. Example of coefficients for the different kinds of music on the base of customers' age.

	Rock	Classical	Kids
Elder	0.1	0.7	0.3
Young	0.8	0.1	0.5
Family	0.4	0.2	0.8

environment in a given moment. How to recognise the customers in a store is discussed in Sect. 5.2.

The *definition of weights* allows to create connections between classes of customers and values of the comfort aspects. For instance, Table 2 reports an example of weights between customers classified by their age and group (family) and kinds of music. Similar tables can be defined for the other customers' classes and values of comfort aspects.

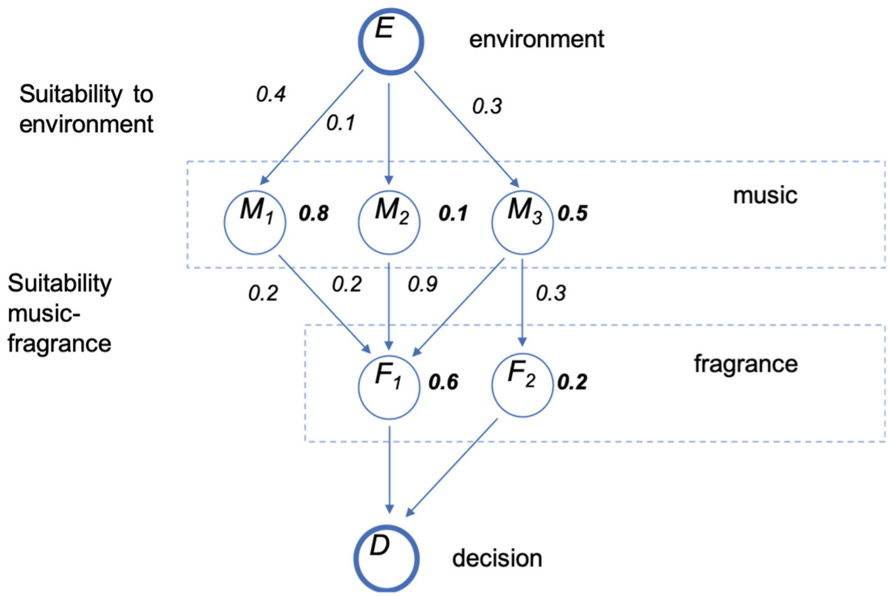

Fig. 2. Decision graph for music and fragrance.

Finally, *a composition graph* enables to define different compositions of comfort aspects. Figure 2 shows a decision graph regarding the use case summarised in Table 2. Let us suppose that the system identifies that the most of the customer are *young*, so the weights in the second raw of Table 2 is considered; the corresponding weights are reported in the nodes M_1, M_2 and M_3. The weight of the fragrances (nodes F_1 and F_2) can be retrieved in a similar way. Then, we define the weights of the edges, which represent the suitability for *couples* of nodes. For instance, the upper edges represent the suitability of a given music to the environment, while the middle edges represent the suitability of a music to a fragrance. Note that not all combination could be possible (for instance, there is no edge between the node M_1 and the node F_2). Once we have normalised all weights to the range [0,1], the weight of each path from the environment E to the decision D can be computed as the product of the node and edge weights of the path. The path with the highest weight product represents the best combination of music and fragrance for the environment E when the most of the customers are young.

Interested readers can find more details about the decision graph in [6].

Service Composition Based on Statistical Information
The first example concerns the decision about what music to play in a retail store to provide comfort to the customers. Of course, the "best" music depends on the tastes of the persons present in the store. As a first approach, we can exploit statistical information to know information about potential people in the store, such as age, gender, and other. If the music preferences are not a piece of the statistical information, the system can rely on a specific service that provides the weight to give to the pieces of information. The decision service can put all information together and decide the music to play.

5.2 Examples of Service Composition

Figure 3 reports a simple example in which the age of the potential customers is exploited to decide the music to play. Following the microservice architecture, in this example we exploit a service to get the hour of the day (`Time service`) which gives the input to a service (`Demographic service`) that provides statistical information about the mean ages of the customers in that hour to another service that ranges the ages (`Age range service`). The `Decision service` takes the range of ages as input along with the weights to be considered for each age range from the `Weight service` and decide which music to play. The decided music is passed to the `Music service` that finally play the music in the store.

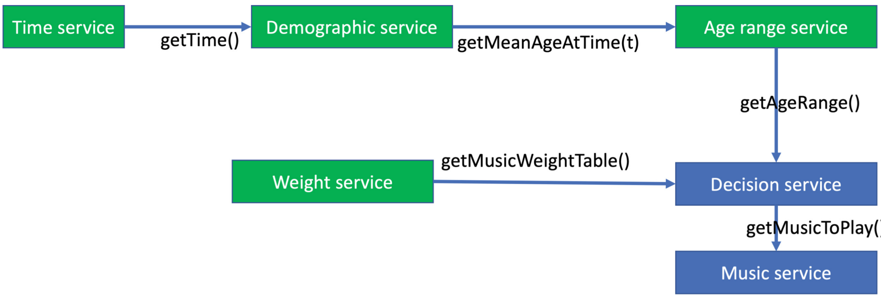

Fig. 3. Example of service composition to decide the most appropriate music, using statistical information.

Service Composition Based on Identification Information

In connection with the first example, we can think at a more refined approach that identifies the customer inside the retail store. For example, by providing customers with an app on their smartphones capable of communicating their presence in the store.

More sophisticated approaches can be enacted as well, in particular biometric recognition [9] or face recognition [7] or context-aware edge computing [8]. These approaches imply privacy issues, because people are identified and tracked possibly without explicit consent. If the identification is carried out by an app, we can assume that the customer has agreed in being sensed inside a store; nevertheless, in this way less people can be identified (only those with the store app and only those who granted the permission to the app to access the position). Otherwise, recognising people by means of face or biometric recognition, allows the identification of more people, but is more complex to be deployed and it might be forbidden by national laws.

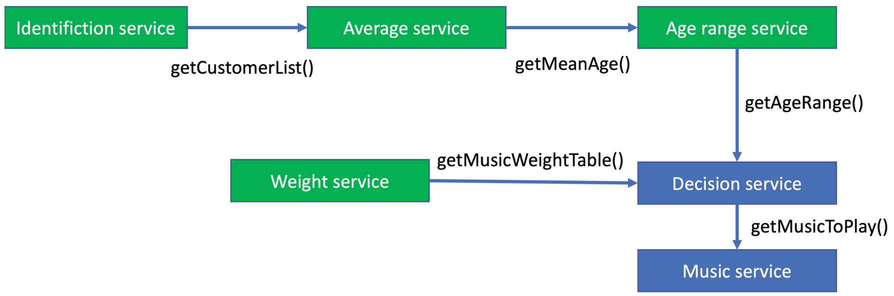

Fig. 4. Example of service composition to decide the most appropriate music using identification of customers in the store.

Independently of implementation details, microservices can be composed as depicted in Fig. 4. Some services have been reused from the previous case; this is one benefit of exploiting microservices. Furthermore, we introduced a statistical service (`Demographic service` in Fig. 3) for identifying customers (`Identification service`) coupled with a service that calculate the average age distribution of the customers in the store (`Average service`).

Service Composition with Different Sources

As mentioned above, the decision system can rely on different and heterogeneous information sources [13]. For instance, the decision system can get information from both a statistical and an identification services (see Fig. 5). Having more sources can improve the precision of the context, but it is also more difficult to manage.

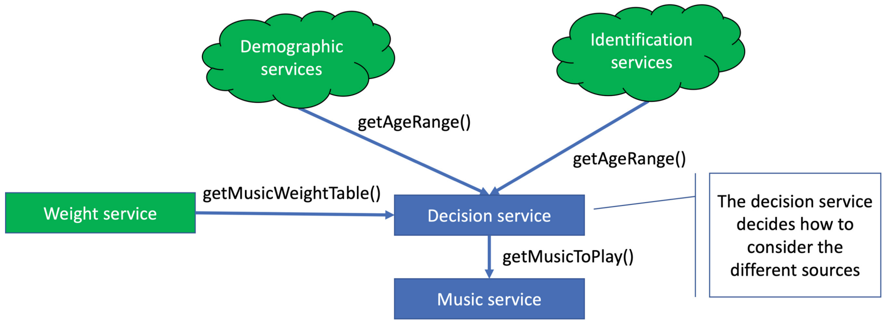

Fig. 5. Example of decision system for *music*, weighting different sources.

The clouds represent the fact that the information is likely to be provided by a set of coordinated services. In this case, the `Decision service` must apply weights to the different pieces of information in order to take a decision.

Service Composition for Colours

To assess the modularity of the proposed architecture, we also consider the case in which the system can control the lights inside a store.

As depicted in Fig. 6, using the microservice based architecture we can compose the set of needed services by reusing the statistical and the identification clouds of service, and exploiting the `Weight service` for the colours and the `Color service` to enact the decision. The modularity of the microservices enables us to adapt the architecture to other decisions, with only small changes. The differences with the previous case are highlighted by red circles in Fig. 6.

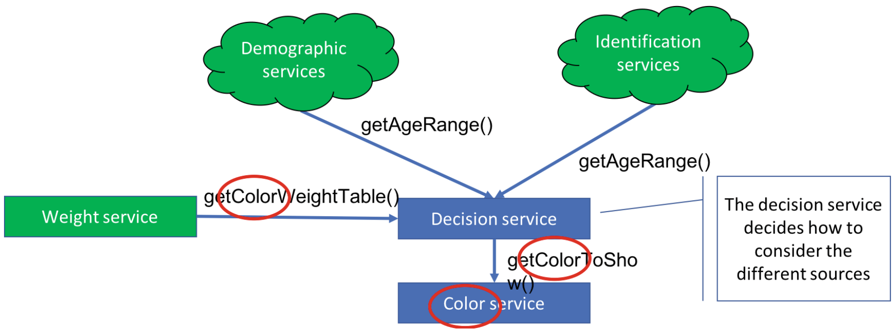

Fig. 6. Example of decision system for *colours*, weighting different sources; the differences between Fig. 5 are highlighted. (Color figure online)

6 Conclusion and Future Work

In this paper we have presented a user-aware architecture to enhance comfort in retail environments. The comfort is achieved by choosing the most appropriate comfort aspects (e.g., music, colours, fragrance, ...) inside the retail store.

We have proposed a model of user context relying on *external*, *statistical* and *personal* information. This context is then used for taking decisions to maximise the comfort level of the customers inside the store.

We have chosen to use an architecture based on *microservices*, granting modularity, flexibility and ease of deployment if real-world cases. Three case studies show concrete applications of the architecture.

Regarding future work, we sketch some main directions. First, privacy issues should be taken into account, in particular when people inside the store are identified and tracked. Second, we will evaluate interoperability issues possibly related to service discovery and composition [3,4]. Third, it would be interesting to explore how the change of the environment (e.g., a change in the store colours) can affect the customers' context, triggering a sort of "loop". Last, we are developing a prototype of the architecture that will be tested in actual environments.

Acknowledgment. This work was supported by the EU H2020 program under Grant No. 734599 - FIRST project.

References

1. Bauer, C., Garaus, M., Strauss, C., Wagner, U.: Research directions for digital signage systems in retail. Proc. Comput. Sci. **141**, 503–506 (2018)
2. Betzing, J.H., Beverungen, D., Becker, J.: Design principles for co-creating digital customer experience in high street retail. In: Proceedings of the Multikonferenz Wirtschaftsinformatik, MKWI 2018 (2018)
3. Bicocchi, N., Cabri, G., Mandreoli, F., Mecella, M.: Dealing with data and software interoperability issues in digital factories. In: Transdisciplinary Engineering Methods for Social Innovation of Industry 4.0: Proceedings of the 25th ISPE Inc. International Conference on Transdisciplinary Engineering, 3–6 July 2018, vol. 7, p. 13. IOS Press (2018)
4. Bicocchi, N., Cabri, G., Mandreoli, F., Mecella, M.: Dynamic digital factories for agile supply chains: an architectural approach. J. Ind. Inf. Integr. **15**, 111–121 (2019)
5. Bicocchi, N., Leonardi, L., Zambonelii, F.: Software-intensive systems for smart cities: from ensembles to superorganisms. In: De Nicola, R., Hennicker, R. (eds.) Software, Services, and Systems. LNCS, vol. 8950, pp. 538–551. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-15545-6_31
6. Cabri, G., Leoncini, M., Martoglia, R., Zambonelli, F.: Towards user-aware service composition. In: Vinh, P., Barolli, L. (eds.) ICTCC 2016. LNICST, vol. 168, pp. 11–21. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46909-6_2
7. Chiang, H.H., et al.: Development of smart shopping carts with customer-oriented service. In: 2016 International Conference on System Science and Engineering (ICSSE), pp. 1–2. IEEE (2016)

8. Cho, J., Sundaresan, K., Mahindra, R., Van der Merwe, J., Rangarajan, S.: ACACIA: context-aware edge computing for continuous interactive applications over mobile networks. In: Proceedings of the 12th International on Conference on Emerging Networking EXperiments and Technologies, pp. 375–389. ACM (2016)
9. Karpey, D., Pender, M.: Customer identification through voice biometrics. US Patent 9,396,730, 19 July 2016
10. Lamche, B., Rödl, Y., Hauptmann, C., Wörndl, W.: Context-aware recommendations for mobile shopping. In: LocalRec@ RecSys, pp. 21–27 (2015)
11. Meyer, M., Helmholz, P., Robra-Bissantz, S.: Digital transformation in retail: can customer value services enhance the experience? (2018)
12. Schaub, F., Könings, B., Weber, M.: Context-adaptive privacy: leveraging context awareness to support privacy decision making. IEEE Pervasive Comput. **14**(1), 34–43 (2015)
13. Uddin, I., Rakib, A., Haque, H.M.U., Vinh, P.C.: Modeling and reasoning about preference-based context-aware agents over heterogeneous knowledge sources. Mobile Netw. Appl. **23**(1), 13–26 (2018)
14. Zambonelli, F.: Toward sociotechnical urban superorganisms. Computer **45**(8), 76–78 (2012)

Justifiable Exceptions in General Contextual Hierarchies

Loris Bozzato[1(✉)], Thomas Eiter[2], and Luciano Serafini[1]

[1] Fondazione Bruno Kessler, Via Sommarive 18, 38123 Trento, Italy
{bozzato,serafini}@fbk.eu
[2] Institute of Logic and Computation, Technische Universität Wien,
Favoritenstraße 9-11, 1040 Vienna, Austria
eiter@kr.tuwien.ac.at

Abstract. The problem of representing and reasoning with context dependent knowledge has recently gained interest in the area of description logics: among the several proposals, we consider the Contextualized Knowledge Repository (CKR) framework. In CKR applications it is often useful to reason over a hierarchical organization of contexts: for this reason, in our recent work we extended the CKR model to allow for the representation of exception handling in the inheritance of knowledge across contexts. However, to simplify the definition of reasoning procedures, we limited our approach to a particular kind of context organization, i.e. ranked hierarchies. In this paper, we further develop the proposal to extend the reasoning on exception handling for CKRs with general contextual hierarchies. We adapt the semantics (on a core version of CKR) to cope with contextual defeasible axioms in general hierarchies; on the base of this, we define an ASP based reasoning procedure that is complete w.r.t. instance checking under the proposed semantics for general contextual hierarchies.

1 Introduction

Representing and reasoning with contexts has recently gained increasing interest in the Semantic Web area, due to the need for interpreting knowledge resources with respect to contextual information given in their metadata. This led to a number of (description) logic based approaches e.g. [13,14,17,18]. In this line of works, we consider the recent proposal of the *Contextualized Knowledge Repository (CKR) framework* [6,17], with its latest formulation in [4].

A CKR knowledge base is a two-layer structure where the higher level consists of a *global context* and the lower level consists of a set of *local contexts*. The global context contains context-independent knowledge about the domain of discourse (*global object knowledge*) and the structure and properties of the local contexts (*meta-knowledge*). Local contexts contain knowledge that holds under specific situations (e.g. during a certain period of time, region in space). The global object knowledge is propagated to the local contexts and it is used to constrain local knowledge in different contexts. In [4] an extension to CKR was proposed by introducing a notion of *justifiable exceptions*. Axioms in the global context may be specified as *defeasible*, meaning that in general they are "inherited" in local instances, but these can be "overridden" on some

G. Bella and P. Bouquet (Eds.): CONTEXT 2019, LNAI 11939, pp. 26–39, 2019.
https://doi.org/10.1007/978-3-030-34974-5_3

exceptional instance if they would cause a local contradiction. A limitation of the proposal in [4] is that inheritance of defeasible axioms is restricted to the direction from the global to the local contexts: in general, one may want to specify more complex structures of contexts and control the knowledge propagation across such structures e.g. in the case of hierarchies of contexts specified by a context *coverage* relation [17].

Thus, in [7] we generalize this approach by allowing for local defeasible axioms in coverage contextual hierarchies. For the interpretation of overridings, we prefer models that prioritize the validity of defeasible axioms at the most specific contexts. In [7] we concentrate on *ranked contextual hierarchies*, namely hierarchies that can be divided in a linear order of levels: this restriction allows us to define a simple "global" preference on models based on the level of the overridden defeasible axioms. This also permits to easily adapt the translation to ASP programs from [4] by computing preference across answer sets by means of *weak constraints* [16] on the level of overridings.

In this paper, we continue the work in [7] by considering the case of CKRs with *general* contextual coverage structures. In order to cope with the interpretation of overriding in generic hierarchies, we need to adopt a "local" preference on models. Intuitively, a (non-strict) preference on local defeasible axioms is derived from their position in the coverage hierarchy: we prefer models which override the axioms in the higher contexts in the hierarchy, in order to prefer the most specific axioms in the lower contexts. However, while in [7] preference was mapped on the linear approximation provided by levels, with general hierarchies such preference has to be defined by considering the local coverage relations of the contexts of the overridden axioms. This provides a more accurate definition of the preference, but the comparison on the models is more complex. This aspect reflects on the reasoning method we provide for instance checking: we provide an algorithm, based on the preference semantic definition, that is able to derive the "preferred" answer sets which encode the expected interpretation of inheritance.

The contributions of this paper can be summarized as follows:

- We describe the extension of the CKR semantics with defeasible axioms in local contexts, as provided in [7]. Inheritance and overriding of defeasible axioms is defined over a hierarchical coverage relation across local contexts: in order to concentrate on the contextual structure, we work on a restriction of CKR that we call *simple CKR (sCKR)*. In this paper we consider general contextual hierarchies: in the definition of the semantics, we refine the definition of model preference to consider the "local" ordering of overridings in the contextual hierarchy.
- We summarize the computational complexity of major reasoning tasks, in particular axiom entailment and conjunctive query (CQ) answering in the case of reasoning on general hierarchies. Under the new definition of model preference, we can show that axiom entailment is Π_2^p-complete and CQ-answering is Π_2^p-complete: as in the case of level-based preference, reasoning with preference increases the complexity of entailment, but it does not for CQ answering.
- We extend to general hierarchies the reasoning method by a translation to datalog (with negation under answer set semantics) for simple CKRs in $\mathcal{SROIQ}$-RL proposed in [7]. In order to restrict reasoning on preferred models, we provide an algorithm for comparing answer sets based on the semantic definition of the local

preference of overridings. The resulting reasoning procedure provides a sound and complete method for instance checking and conjunctive query answering on sCKR.

For space limitations, we refer the reader to [4,7] for preliminaries on the definitions of $\mathcal{SROIQ}$-RL language and datalog programs under answer set semantics that we assume in the following sections.

2 Simple CKR with Justifiable Exceptions

We provide in this section the definition of simple CKR (sCKR) introduced in [7] adapted to the case of general hierarchies. With respect to the original formulation of CKR presented in [3,4,6], a simple CKR is still a two layered structure: however, in order to emphasize the role of the coverage relation, we simplify the upper layer to be a poset based on such relation.

Syntax. Consider a non empty-set of *context names* $\mathbf{N} \subseteq \mathrm{NI}$. We define a *coverage* relation $\prec \subseteq \mathbf{N} \times \mathbf{N}$. Given context names $\mathsf{c}_1, \mathsf{c}_2 \in \mathbf{N}$, we say that c_2 *covers* c_1 if $\mathsf{c}_1 \prec \mathsf{c}_2$. The coverage relation $\prec$ is a strict partial order relation on $\mathbf{N}$, i.e. it is irreflexive and transitive. Intuitively, $\mathsf{c}_1 \prec \mathsf{c}_2$ means that c_2 is more general than c_1, in the sense that c_2 refers to a portion of the world that covers the one described by c_1 [17]. We may use the non-strict relation $\mathsf{c}_1 \preceq \mathsf{c}_2$ to indicate that c_1 can be covered by c_2 or is the same context.
We can now define the language used in the local contexts to express their knowledge.

Definition 1 (contextual language). *Given a set of context names* $\mathbf{N}$*, for every description language* $\mathcal{L}_\Sigma$ *we define* $\mathcal{L}_{\Sigma,\mathbf{N}}$ *as the description language* $\mathcal{L}$ *with the following additional rule for concept and role formation:* $\mathsf{eval}(X, \mathsf{c})$ *is a concept (resp. role) of* $\mathcal{L}_{\Sigma,\mathbf{N}}$ *if* X *is a concept (resp. role) of* $\mathcal{L}_\Sigma$ *and* $\mathsf{c} \in \mathbf{N}$.

Definition 2 (defeasible axiom). *A* defeasible axiom *is any expression of the form* $\mathrm{D}(\alpha)$*, where* α *is an axiom of* $\mathcal{L}_\Sigma$*. The DL language* $\mathcal{L}^{\mathrm{D}}_{\Sigma,\mathbf{N}}$ *extends* $\mathcal{L}_{\Sigma,\mathbf{N}}$ *with the set of defeasible axioms in* $\mathcal{L}_\Sigma$*.*

Using these definitions, simple CKRs are defined as follows:

Definition 3 (simple Contextualized Knowledge Repository, sCKR). *A* Simple Contextualized Knowledge Repository (sCKR) *over* Σ *and* $\mathbf{N}$ *is a structure* $\mathfrak{K} = \langle \mathfrak{C}, \mathrm{K}_\mathbf{N} \rangle$ *where:*

- $\mathfrak{C}$ *is a poset* $(\mathbf{N}, \prec)$ *and*
- $\mathrm{K}_\mathbf{N} = \{\mathrm{K}_\mathsf{c}\}_{\mathsf{c} \in \mathbf{N}}$ *for each context name* $\mathsf{c} \in \mathbf{N}$*,* K_c *is a DL knowledge base over* $\mathcal{L}^{\mathrm{D}}_{\Sigma,\mathbf{N}}$.

Example 1. We adapt the corporation example from [7] to the case of a simple non-ranked hierarchy. Let us consider a sCKR $\mathfrak{K}_{org} = \langle \mathfrak{C}, \mathrm{K}_\mathbf{N} \rangle$ describing the organization of a corporation. The corporation wants to define different policies with respect to its local branches. The hierarchy of contexts $\mathfrak{C}$, representing the corporation organization, is shown in Fig. 1.

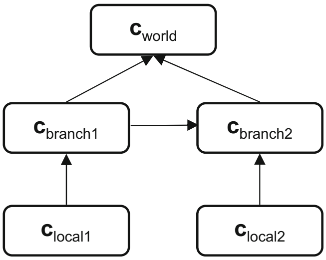

Fig. 1. Context hierarchy of example sCKR.

The company can enforce policies depending on the branch and its influence on the local sites. The corporation is active in the fields of Musical instruments (M), Electronics (E) and Robotics (R). A supervisor (S) can be assigned to manage only one of these fields. By putting defeasible axioms in the correct contexts in K_{N}, we can assign local supervisors to their field:

$$\begin{array}{ll} \mathsf{c}_{world} & : \{M \sqcap E \sqsubseteq \bot, M \sqcap R \sqsubseteq \bot, E \sqcap R \sqsubseteq \bot, \mathrm{D}(S \sqsubseteq E)\} \\ \mathsf{c}_{branch1} & : \{\mathrm{D}(S \sqsubseteq M)\} \qquad \mathsf{c}_{branch2} : \{\mathrm{D}(S \sqsubseteq R)\} \\ \mathsf{c}_{local1} & : \{S(i)\} \end{array}$$

In c_{world} we say that supervisors are assigned to Electronics, while in the subcontext for $\mathsf{c}_{branch2}$ we contradict this by assigning all local supervisors to the Robotics area and in $\mathsf{c}_{branch1}$ we further specialize this by assigning supervisors to the Musical instruments area. In the contexts for local sites we have information about the instances (here we consider only c_{local1} for simplicity). Note that different assignments of areas for i are possible by instantiating the defeasible axioms: intuitively, we want to prefer the interpretations that override the higher defeasible axioms in c_{world} and $\mathsf{c}_{branch2}$. $\diamondsuit$

sCKRs restrict the original CKR definition in [3,4,6] in order to concentrate on aspects of defeasibility across contexts. Any sCKR can be easily translated in a CKR (by suitably adapting the interpretation of coverage context relation like in the following definition of semantics). Differently from [7], we do not restrict the form of the poset $\mathfrak{C}$. Thus, the definition of preference across different models can not be based on syntactic properties of the hierarchy: it will be defined over a local ordering on elements of the semantics (clashing assumptions).

Semantics. An interpretation for a sCKR is a set containing an interpretation for each local context.

Definition 4 (sCKR interpretation). *An interpretation for $\mathcal{L}_{\Sigma,\mathsf{N}}$ is a family $\mathfrak{I} = \{\mathcal{I}(\mathsf{c})\}_{\mathsf{c}\in\mathsf{N}}$ of $\mathcal{L}_{\Sigma}$ interpretations, such that $\Delta^{\mathcal{I}(\mathsf{c})} = \Delta^{\mathcal{I}(\mathsf{c}')}$ and $a^{\mathcal{I}(\mathsf{c})} = a^{\mathcal{I}(\mathsf{c}')}$, for every $a \in \mathrm{NI}_{\Sigma}$ and $\mathsf{c}, \mathsf{c}' \in \mathsf{N}$.*

The interpretation of concepts and role expressions in $\mathcal{L}_{\Sigma,\mathsf{N}}$ is obtained by extending the standard interpretation to *eval* expressions: for every $\mathsf{c} \in \mathsf{N}$, $\mathit{eval}(X, \mathsf{c}')^{\mathcal{I}(\mathsf{c})} = X^{\mathcal{I}(\mathsf{c}')}$. We consider the notion of axiom instantiation and clashing assumption as defined in [4].

Definition 5 (axiom instantiation). *Given an axiom $\alpha \in \mathcal{L}_\Sigma$ with FO-translation $\forall \mathbf{x}.\phi_\alpha(\mathbf{x})$, the instantiation of α with a tuple $\mathbf{e}$ of individuals in* NI*, written $\alpha(\mathbf{e})$, is the specialization of α to $\mathbf{e}$, i.e., $\phi_\alpha(\mathbf{e})$, depending on the type of α.*

Definition 6 (clashing assumptions and sets). *A* clashing assumption *for a context* c *is a pair $\langle \alpha, \mathbf{e} \rangle$ such that $\alpha(\mathbf{e})$ is an axiom instantiation of α, and $\mathrm{D}(\alpha) \in \mathrm{K}_{\mathsf{c}'}$ is a defeasible axiom of some $\mathsf{c}' \succ \mathsf{c}$. A* clashing set *for $\langle \alpha, \mathbf{e} \rangle$ is a satisfiable set S of ABox assertions such that $S \cup \{\alpha(\mathbf{e})\}$ is unsatisfiable.*

A clashing assumption $\langle \alpha, \mathbf{e} \rangle$ represents that $\alpha(\mathbf{e})$ is not (DL-)satisfiable, and a clashing set S provides an assertional "justification" for the assumption of local overriding of α on $\mathbf{e}$. We extend our interpretations to account for such notions.

Definition 7 (CAS-interpretation). *A* CAS-interpretation *is a pair $\mathfrak{I}_{CAS} = \langle \mathfrak{I}, \chi \rangle$ where $\mathfrak{I}$ is an interpretation and χ maps every $\mathsf{c} \in \mathbf{N}$ to a set $\chi(\mathsf{c})$ of clashing assumptions for context* c.

Definition 8 (CAS-model). *Given a sCKR $\mathfrak{K}$, a CAS-interpretation $\mathfrak{I}_{CAS} = \langle \mathfrak{I}, \chi \rangle$ is a CAS-model for $\mathfrak{K}$ (denoted $\mathfrak{I}_{CAS} \models \mathfrak{K}$), if the following holds:*

(i) for every $\alpha \in \mathrm{K}_{\mathsf{c}}$ (strict axiom), and $\mathsf{c}' \preceq \mathsf{c}$, $\mathcal{I}(\mathsf{c}') \models \alpha$;
(ii) for every $\mathrm{D}(\alpha) \in \mathrm{K}_{\mathsf{c}}$, $\mathcal{I}(\mathsf{c}) \models \alpha$;
(iii) for every $\mathrm{D}(\alpha) \in \mathrm{K}_{\mathsf{c}}$ and $\mathsf{c}' \prec \mathsf{c}$, if $\mathbf{d} \notin \{\mathbf{e} \mid \langle \alpha, \mathbf{e} \rangle \in \chi(\mathsf{c}')\}$, then $\mathcal{I}(\mathsf{c}') \models \phi_\alpha(\mathbf{d})$.

In order to generalize the model preference to general hierarchies, in this paper we consider a *local* preference on clashing assumption sets (cf. discussion in [7]). The preference is defined directly along the coverage relation:

> $\chi_1(\mathsf{c}) > \chi_2(\mathsf{c})$, if for every $\eta = \langle \alpha_1, \mathbf{e} \rangle \in \chi_1(\mathsf{c}) \setminus \chi_2(\mathsf{c})$ with $\mathrm{D}(\alpha_1)$ at a context $\mathsf{c}_1 \succ \mathsf{c}$, there exists an $\eta' = \langle \alpha_2, \mathbf{f} \rangle \in \chi_2(\mathsf{c}) \setminus \chi_1(\mathsf{c})$ with $\mathrm{D}(\alpha_2)$ at context $\mathsf{c}_2 \succ \mathsf{c}$ such that $\mathsf{c}_1 \succ \mathsf{c}_2$.

This definition reflects the intuition that if we make in $\chi_1(\mathsf{c})$ an exception at c_1, then a "more costly" exception should be made at a context c_2 below c_1 by $\chi_2(\mathsf{c})$ that is not made by $\chi_1(\mathsf{c})$.

Two DL interpretations $\mathcal{I}_1$ and $\mathcal{I}_2$ are NI-*congruent*, if $c^{\mathcal{I}_1} = c^{\mathcal{I}_2}$ holds for every $c \in$ NI. This naturally extends to a (CAS) interpretation $\mathfrak{I}_{CAS} = \langle \mathfrak{I}, \chi \rangle$ by considering all context interpretations $\mathcal{I}(\mathsf{c})$ in $\mathfrak{I}$. We say that a clashing assumption $\langle \alpha, \mathbf{e} \rangle \in \chi(\mathsf{c})$ is *justified* for a CAS model $\mathfrak{I}_{CAS}$, if some clashing set $S = S_{\langle \alpha, \mathbf{e} \rangle, \mathsf{c}}$ exists such that, for every CAS-model $\mathfrak{I}'_{CAS} = \langle \mathfrak{I}', \chi \rangle$ of $\mathfrak{K}$ that is NI-congruent with $\mathfrak{I}_{CAS}$, it holds that $\mathcal{I}'(\mathsf{c}) \models S_{\langle \alpha, \mathbf{e} \rangle, \mathsf{c}}$.

Definition 9 (justified CAS model). *A CAS model $\mathfrak{I}_{CAS}$ of a sCKR $\mathfrak{K}$ is* justified*, if every $\langle \alpha, \mathbf{e} \rangle \in CAS$ is justified in CKR.*

To interpret the intended preference on defeasible axioms, sCKR models not only need to require the existence of a CAS model justifying the exceptions, but also require that such CAS model "minimizes" the position in the hierarchy of the overridden defeasible axioms: in this way, in case of alternative solutions, axioms at the lower parts of the coverage hierarchy (i.e. more specific) are preferred to axioms at the higher contexts (i.e. more general). In the current setting, model preference is defined from local priority by the following condition:

$\mathfrak{I}^1_{CAS} = \langle \mathfrak{I}^1, \chi_1 \rangle$ is preferred to $\mathfrak{I}^2_{CAS} = \langle \mathfrak{I}^2, \chi_2 \rangle$ iff there exists some $\mathsf{c} \in \mathbf{N}$ s.t. $\chi_1(\mathsf{c}) > \chi_2(\mathsf{c})$ and not $\chi_2(\mathsf{c}) > \chi_1(\mathsf{c})$, and for no context $\mathsf{c}' \neq \mathsf{c} \in \mathbf{N}$ it holds that $\chi_1(\mathsf{c}') < \chi_2(\mathsf{c}')$ and not $\chi_2(\mathsf{c}') < \chi_1(\mathsf{c}')$.

We note that this definition of model preference (together with the ordering on clashing assumption sets) provides a *non-symmetric* and *non-transitive* relation over models.

Definition 10 (CKR model). *An interpretation* $\mathfrak{I}$ *is a* CKR model *of a sCKR* $\mathfrak{K}$ *(in symbols,* $\mathfrak{I} \models \mathfrak{K}$*) if:*

- $\mathfrak{K}$ *has some justified CAS model* $\mathfrak{I}_{CAS} = \langle \mathfrak{I}, \chi \rangle$*;*
- *there exists no justified CAS model* $\mathfrak{I}'_{CAS} = \langle \mathfrak{I}, \chi' \rangle$ *that is preferred to* $\mathfrak{I}_{CAS}$*.*

Example 2. Considering the sCKR $\mathfrak{K}_{org}$ of previous example, we note that different justified CAS models are possible, corresponding to the different assignments of the supervisor individual i in the c_{local1} context to the alternative product areas denoted by the defeasible axioms in the upper contexts. We have three possible assignments, corresponding to three different clashing assumptions maps for the local context:

$$\chi_1(\mathsf{c}_{local1}) = \{\langle S \sqsubseteq E, i \rangle, \langle S \sqsubseteq R, i \rangle\}$$
$$\chi_2(\mathsf{c}_{local1}) = \{\langle S \sqsubseteq M, i \rangle, \langle S \sqsubseteq R, i \rangle\}$$
$$\chi_3(\mathsf{c}_{local1}) = \{\langle S \sqsubseteq M, i \rangle, \langle S \sqsubseteq E, i \rangle\}$$

By the definition of ordering on clashing assumption sets, we have in particular that:

$$\chi_1(\mathsf{c}_{local1}) > \chi_2(\mathsf{c}_{local1}) \qquad \chi_1(\mathsf{c}_{local1}) > \chi_3(\mathsf{c}_{local1}) \qquad \chi_3(\mathsf{c}_{local1}) > \chi_2(\mathsf{c}_{local1})$$

Thus, following the definition of model preference, there is one preferred model for our sCKR which corresponds to χ_1: note that it corresponds to the intended interpretation in which the defeasible axiom $\mathrm{D}(S \sqsubseteq M)$ associated to the most specific context wins over the more general rules asserted for the higher contexts. $\diamondsuit$

We remark that the presented local model preference relation is only one of the possible solutions for an ordering condition that encodes our intended reading for the priority of overridings. Further ordering conditions can be devised e.g. by considering instances and axioms in comparisons or different properties of the ordering relation. We leave the formulation and study of properties for such alternative orderings as a direction for future work.

Reasoning and Complexity. We summarize in the following the reasoning tasks and the main complexity results as in [7]. We consider these reasoning tasks:

- c-*entailment*, where $\mathfrak{K} \models \mathsf{c} : \alpha$ denotes for an axiom α that α is entailed in every CKR-model of $\mathfrak{K}$ at context c (i.e., $\mathcal{I}(\mathsf{c}) \models \alpha$);
- *(Boolean) conjunctive query (CQ) answering* $\mathfrak{K} \models \exists \mathbf{y} \gamma(\mathbf{y})$, where $\gamma(\mathbf{y}) = \gamma_1 \wedge \cdots \wedge \gamma_m$ is a conjunction of atoms $\gamma_i = \mathsf{c}_i : \alpha_i(\mathbf{t}_i)$, where each c_i is a context name and $\alpha_i(\mathbf{t}_i)$ is an assertion in which variables occur, which is existentially closed.

It has been shown in [4] that justified CAS-model checking, i.e. deciding whether a given CAS-interpretation is a justified CAS-model of a given CKR $\mathfrak{K}$ is feasible in polynomial time, and that satisfiability (existence of a CKR-model) is NP-complete. Furthermore, c-*entailment* testing and (Boolean) CQ-answering were shown to be coNP- and Π^p_2-complete problems, respectively.

In the case of reasoning with contextual hierarchies, while the complexity of satisfiability remains unchanged, model checking is intractable already for the ranked hierarchies of [7]. As a consequence, the complexity of c-entailment increases, while CQ answering remains unchanged. In what follows, we assume the setting of [4] for the complexity analysis.

Proposition 1. *Deciding whether a CAS-interpretation $\mathfrak{I}_{CAS}$ of a sCKR $\mathfrak{K}$ is a CKR-model is coNP-complete.*

Informally, $\mathfrak{I}_{CAS}$ can be refuted if it is not a justified CAS-model of $\mathfrak{K}$, which can be checked in polynomial time using the techniques in [4], or some preferred model $\mathfrak{I}'_{CAS}$ exists; the latter can be guessed and checked in polynomial time. The coNP-hardness can be shown by a reduction from a variant of UNSAT.

Theorem 1. *Suppose $\mathfrak{K}$ is a sCKR with global preference induced by a local preference $>$ that is polynomial-time decidable. Then deciding c-entailment $\mathfrak{K} \models \mathsf{c} : \alpha$ is Π^p_2-complete.*

In particular, we note that the model ordering we propose in this paper satisfies the condition (CP) considered in [7] to motivate the Π^p_2-hardness.

Theorem 2. *Deciding whether an sCKR $\mathfrak{K}$ entails a Boolean CQ γ is Π^p_2-complete for profile-based preference.*

These results can be motivated similarly: intuitively, a CKR-model $\mathfrak{I}_{CAS}$ that does not entail γ can be guessed and checked with the help of an NP oracle (ask whether no preferred $\mathfrak{I}'_{CAS}$ exists and whether γ is entailed), and similarly for local preference. The Π^p_2-hardness is inherited from ordinary CKR.

For data complexity (i.e. the CKR $\mathfrak{K}$ is fixed and only the assertions in the knowledge modules vary), while CKR model checking remains coNP-complete, the complexity of c-entailment drops to $\Delta^p_2[O(\log n)] = \mathrm{P}^{\mathrm{NP}}|_{\|[k]}$ (cf. [10]): the problem can be decided with a constant number k of rounds of parallel NP oracle queries and it is complete for this class. On the other hand, CQ entailment remains Π^p_2-complete.

3 Reasoning Procedure for General Hierarchies

In this section we adapt the reasoning method for sCKR in $\mathcal{SROIQ}$-RL presented in [7] to the case of general hierarchies. Basically, the datalog translation presented in [7] (and based on the one in [4]) can be adopted to the current setting for reasoning with simple CKRs and local defeasible axioms: while in the translation in [4] inference is obtained from *all* answer sets of the resulting program (i.e. cautious reasoning), in order to reason on the inheritance of local defeasible axioms we need to select the *preferred* answer sets accordingly to the model ordering defined on sCKR models. However, in this case we can not take advantage of a specific form of the contextual hierarchies: thus, we provide a general algorithm (based on the semantic definition of model preference) to compute the preferred models w.r.t. the ordering of their clashing assumptions.

Language and Normal Form. As in the original version of the translation, we limit the defeasible axioms to the language of $\mathcal{SROIQ}$-RLD: i.e. in defeasible axioms, $D \sqcap D$ can not appear as a right-side concept and each right-side concept $\forall R.D$ has $D \in \mathrm{NC}$. Also, we consider the normal form and normal form transformation proposed in [4] for the formulation of the rules (considering the formulas that can appear in the simple CKRs) and we assume again the Unique Name Assumption.

Translation Overview. The translation to datalog extends the one presented in [4]: the non-trivial use of non-monotonic negation and inference on negative literals by contradiction for the interpretation of exceptions is here extended to reason on local defeasible inheritance. However, since we do not consider the computation of preference in the translation, differently from rules in [7] we do not consider weak constraints on the level of overriding.

Formally, we consider the datalog translation composed by the rules in [8] by leaving out the rules in Table 8 and the global input rule for the interpretation of levels (igl-level) in Table 3.

As in the original formulation (inspired by the materialization calculus in [15]), the translation includes sets of *input rules* (which encode DL axioms and signature in datalog), *deduction rules* (datalog rules providing instance level inference) and *output rules* (that encode in terms of a datalog fact the ABox assertion to be proved). The translation is composed by the following sets of rules:

$\mathcal{SROIQ}$-RL *Input and Deduction Rules:* rules in $I_{rl}(S, c)$ translate to datalog facts (in a given context c) $\mathcal{SROIQ}$-RL axioms and signature. For example, atomic concept inclusions are translated with the rule $A \sqsubseteq B \mapsto \{\mathtt{subClass}(A, B, c)\}$. Deduction rules in P_{rl} encode the inference rules for $\mathcal{SROIQ}$-RL axioms: e.g., for atomic concept inclusions:

$$\mathtt{instd}(x, z, c, t) \leftarrow \mathtt{subClass}(y, z, c), \mathtt{instd}(x, y, c, t).$$

Global and Local Translations: global input rules $I_{glob}(\mathfrak{C})$ encode the interpretation of the contextual structure. E.g., $\mathtt{c}_1 \prec \mathtt{c}_2 \mapsto \{\mathtt{prec}(\mathtt{c}_1, \mathtt{c}_2)\}$ translates coverage across contexts as instances of the $\mathtt{prec}$ predicate. Local input and deduction roles implement the interpretation of *eval*: note that differently from [4], *eval* can be only defined over single contexts instead of context classes.

Defeasible Axioms Input Translations: defeasible input rules $I_{\mathrm{D}}(S, c)$ declare that a local axiom is defeasible: differently from the translation in [4], the resulting atoms also contain the context in which the axiom has been introduced. For example, $\mathrm{D}(A \sqsubseteq B)$ in context $\mathtt{c}$ translates to $\mathtt{def\_subclass}(A, B, \mathtt{c})$.

Overriding Rules: overriding rules in P_{D} provide rules defining when an axiom of a certain form is locally overridden. With respect to rules in [4], this version of overriding rules has to consider the context in which the defeasible axiom has been declared: this is needed in order to reason on the local preference of the overridden axiom in the algorithm for the computation of preferred models. For example, for axioms of the form $\mathrm{D}(A \sqsubseteq B)$ in context c:

$$\begin{aligned}\texttt{ovr}(\texttt{subClass}, x, y, z, c_1, c) \leftarrow\ & \texttt{def\_subclass}(y, z, c_1),\ \texttt{prec}(c, c_1),\\ & \texttt{instd}(x, y, c, \mathsf{main}),\ \texttt{not}\ \texttt{test\_fails}(\texttt{nlit}(x, z, c)).\end{aligned}$$

Inheritance Rules: P_{D} provides the rules for defeasible inheritance of axioms from the higher to the lower local contexts in the coverage structure. The definition of rules in the case of (general) hierarchies also considers the `prec` relation across contexts, which defines the direction of inheritance. E.g., the following rule propagates a (possibly defeasible) atomic concept inclusion axiom:

$$\begin{aligned}\texttt{instd}(x, z, c, t) \leftarrow\ & \texttt{subClass}(y, z, c_1), \texttt{instd}(x, y, c, t),\\ & \texttt{prec}(c, c_1), \texttt{not}\ \texttt{ovr}(\texttt{subClass}, x, y, z, c_1, c).\end{aligned}$$

Note that this rule propagates also to instances of strict axioms, since their overriding is never verified.

Test Rules: the test rules in P_{D} are used to instantiate and define the "environments" for the tests for negative literals in overriding rules. The mechanism is analogous to the one in [4], but rules need to be adapted to the new definition of `ovr` atoms and `prec` (i.e., they need to consider the context in which the axiom has been declared).

```
PrefModels(P){
    S = get all answer sets of P;
    topM = ∅;
    for( A ∈ S ){
        isTop = true;
        for( B ≠ A ∈ S ){
            if( lessThanM(A, B) ){
                isTop = false;
                break; } }
        if( isTop )  topM.add(A); }
    return topM; }
```

Fig. 2. Procedure `PrefModels` for computation of answer set preference

Model Preference Algorithm. The presented translation computes all justified models for the input sCKR: in order to recognize the preferred models, we need to apply to the computed answer sets the conditions of the model ordering definition. In the case of general hierarchies, we provide the algorithm `PrefModels` that, given the sCKR translation $PK(\mathfrak{K})$ as input, produces the set of preferred answer sets as output.

A pseudo-code for `PrefModels` is presented in Fig. 2: for every answer set A of the input program, the procedure tests whether there does not exist a different answer set

```
lessThanM(A, B){
    lessThan = false;
    for( c ∈ N ){
        if( lessThanCAS(A, B, c) ∧ ¬lessThanCAS(B, A, c) )
            lessThan = true;
        else if( lessThanCAS(B, A, c) ∧ ¬lessThanCAS(A, B, c) ){
            lessThan = false;
            break; } }
    return lessThan; }
```

Fig. 3. Procedure `lessThanM` for comparison of answer sets

B for which $\texttt{lessThanM}(A, B)$ holds (i.e. A is minimal w.r.t. the preference). If that is the case, A is added to the set of results (i.e. preferred models).

The procedure for `lessThanM` is shown in Fig. 3. Intuitively, it provides a comparison between two input answer sets A and B by implementing the condition for model comparison: A is recognized as preceding B if there exists a context $\mathsf{c} \in \mathbf{N}$ such that the preference on clashing assumptions in context c (tested by the procedure `lessThanCAS`) is verified and in no other context this order is disregarded.

The procedure for `lessThanCAS` is presented in Fig. 4. It provides a comparison across clashing assumptions sets for the specific context c from the input answer sets A and B, based on the definition of local (clashing assumption sets) preference in the semantics: with respect to c, B is preferred to A if, for every clashing assumption (i.e. `ovr` atom) from B (and not in A) on a defeasible axiom in $\mathsf{c}_2 \succ \mathsf{c}$, there exists a clashing assumption from A (and not in B) on a defeasible axiom in $\mathsf{c}_1 \succ \mathsf{c}$ such that $\mathsf{c}_1 \prec \mathsf{c}_2$.

```
lessThanCAS(A, B, c){
    ovrA_c = { ovr(p(e), c1, c) ∈ A };
    ovrB_c = { ovr(p(e), c1, c) ∈ B };
    diffA_c = ovrA_c \ ovrB_c;
    diffB_c = ovrB_c \ ovrA_c;
    for( ovr(p(e), c2, c) ∈ diffB_c ∧ c2 ≻ c ){
        lessThan = false;
        for( ovr(p(e), c1, c) ∈ diffA_c ∧ c1 ≻ c ){
            if( c1 ≺ c2 ){
                lessThan = true;
                break; } } }
    return lessThan; }
```

Fig. 4. Procedure `lessThanCAS` for comparison on clashing assumptions sets

We note that, in case of a transitive definition of the model preference, we can modify `PrefModels` so that the pre-computation and storage of all answer sets is not needed: we can compute an answer set A of P at a time and verify if the `lessThanM` condition holds w.r.t. a set of candidate models $Cand$ (initially empty); in case A is recognized as a candidate for preferred model, it is added to $Cand$ and any other model B in $Cand$ s.t. $\texttt{lessThanM}(B, A)$ is removed from $Cand$.

Translation Process. Given a sCKR $\mathfrak{K} = \langle \mathfrak{C}, \mathrm{K}_{\mathbf{N}} \rangle$ in $\mathcal{SROIQ}$-RLD normal form, a program $PK(\mathfrak{K})$ that encodes $\mathfrak{K}$ is obtained as follows:

1. the *global program* for $\mathfrak{C}$ is built as: $PG(\mathfrak{C}) = I_{glob}(\mathfrak{C})$
2. for each $\mathsf{c} \in \mathbf{N}$, we define each local program $PC(\mathsf{c}, \mathfrak{K})$ as:

$$PC(\mathsf{c}, \mathfrak{K}) = P_{rl} \cup P_{loc} \cup P_{\mathrm{D}} \cup I_{loc}(\mathrm{K}_{\mathsf{c}}, \mathsf{c}) \cup I_{rl}(\mathrm{K}_{\mathsf{c}} \cup \mathrm{K}^{\mathrm{D}}_{\mathsf{c}}, \mathsf{c}) \cup I_{\mathrm{D}}(\mathrm{K}_{\mathsf{c}}, \mathsf{c})$$

   where $\mathrm{K}^{\mathrm{D}}_{\mathsf{c}} = \{\alpha \in \mathcal{L}_{\Sigma} \mid \mathrm{D}(\alpha) \in \mathrm{K}_{\mathsf{c}}\}$.
3. The *CKR program* $PK(\mathfrak{K})$ is defined as: $PK(\mathfrak{K}) = PG(\mathfrak{C}) \cup \bigcup_{\mathsf{c} \in \mathbf{N}} PC(\mathsf{c}, \mathfrak{K})$

Query answering $\mathfrak{K} \models \mathsf{c} : \alpha$ is then obtained by testing whether the (instance) query, translated to datalog by $O(\alpha, \mathsf{c})$, is a consequence of the *preferred* models of $PK(\mathfrak{K})$, i.e., whether $\texttt{PrefModels}(PK(\mathfrak{K})) \models O(\alpha, \mathsf{c})$ holds. Analogously, this can be extended to conjunctive queries as in [4].

Correctness. We can show that the presented process provides a sound and complete reasoning method for instance checking (with respect to c-entailment) and conjunctive query answering for normal form $\mathcal{SROIQ}$-RLD simple CKRs with general context hierarchies. The result is shown by extending the correspondence between minimal justified CKR-models of $\mathfrak{K}$ and answer sets of $PK(\mathfrak{K})$ from [4] to the "preferred" answer sets computed by the `PrefModels` algorithm. As in the original formulation, the adoption of UNA and *named models* (i.e. restricting to models having a $N \subseteq \mathrm{NI} \setminus \mathrm{NI}_S$ s.t. the interpretation of atomic concepts and roles belongs to $N^{\mathcal{I}}$) allows to concentrate on Herbrand models for $\mathfrak{K}$, denoted as $\hat{\mathfrak{I}}(\chi)$.

Let $\mathfrak{I}_{CAS} = \langle \mathfrak{I}, \chi \rangle$ be a justified named CAS-model. We can build from its components a corresponding Herbrand interpretation $I(\mathfrak{I}_{CAS})$ of the program $PK(\mathfrak{K})$: the construction is similar to the one in [4] (and its adaptation to hierarchies detailed in [8]).

It is then possible to show that the answer sets of the final program $PK(\mathfrak{K})$ correspond with the least justified models of $\mathfrak{K}$ by the following result:

Lemma 1. *Let $\mathfrak{K}$ be a sCKR in $\mathcal{SROIQ}$-RLD normal form, then:*

(i). for every (named) justified clashing assumption χ, the interpretation $S = I(\hat{\mathfrak{I}}(\chi))$ is an answer set of $PK(\mathfrak{K})$;
(ii). every answer set S of $PK(\mathfrak{K})$ is of the form $S = I(\hat{\mathfrak{I}}(\chi))$ with χ a (named) justified clashing assumption for $\mathfrak{K}$.

As in the case of [7], the result can be proved along the lines of Lemma 6 in [4] by showing that the answer sets of $PK(\mathfrak{K})$ coincide with the sets $S = I(\hat{\mathfrak{I}}(\chi))$ where χ is a justified clashing assumption of $\mathfrak{K}$.

In the case of general hierarchies, the correspondence with sCKR models is obtained by considering the set of answer sets returned by the `PrefModels` algorithm and the notion of model preference on justified CAS-models in the semantics.

Lemma 2. *Let $\mathfrak{K}$ be a sCKR in $\mathcal{SROIQ}$-RLD normal form. Then, $\hat{\mathfrak{I}}$ is a CKR model of $\mathfrak{K}$ iff there exists a (named) justified clashing assumption χ s.t. $I(\hat{\mathfrak{I}}(\chi))$ is a preferred answer set of* `PrefModels`$(PK(\mathfrak{K}))$.

Proof (Sketch). The existence of a justified χ corresponds to the first condition of Definition 10 and is derived from Lemma 1. Then, we have to show that $I(\hat{\mathfrak{I}}(\chi))$ is an answer set returned by `PrefModels`$(PK(\mathfrak{K}))$ if it corresponds to a preferred (justified) model of $\mathfrak{K}$ as in the second condition of Definition 10. This can be verified by the formulation of the algorithm: the procedure `lessThanCAS` applies the definition of preference on clashing assumption sets to the corresponding `ovr` atoms of the answer sets; the preference is then lifted to preference on answer sets by `lessThanM` using the definition of model preference; finally, `PrefModels` applies this preference test on all answer sets, in order to verify that there does not exist other answer sets that are preferred to the ones that are returned. □

The correctness for instance checking is obtained as consequence of previous results:

Theorem 3. *Let $\mathfrak{K}$ be a sCKR in $\mathcal{SROIQ}$-RLD normal form. α and $\mathtt{c}$ such that $O(\alpha, \mathtt{c})$ is defined. Then $\mathfrak{K} \models \mathtt{c} : \alpha$ iff* `PrefModels`$(PK(\mathfrak{K})) \models O(\alpha, \mathtt{c})$.

Similarly to [4], this result can be extended to answering of a conjunctive query Q, by constructing its translation $O(Q)$ by applying output rules to its atoms.

Theorem 4. *Let $\mathfrak{K}$ be a sCKR in $\mathcal{SROIQ}$-RLD normal form. and let $Q = \exists \mathbf{y}\gamma(\mathbf{y})$ be a Boolean CQ on $\mathfrak{K}$. Then $\mathfrak{K} \models Q$ iff* `PrefModels`$(PK(\mathfrak{K})) \models O(Q)$.

4 Related Work

We briefly recall some relevant works related to CKR that include notions of defeasibility in contextual systems and in DLs (we refer to [4,7] for an extended comparison).

We first notice an analogy with non-monotonic multi-context systems (MCS) [9]. The idea of MCS is to align knowledge from different contexts (locally based on possibly different logics) in a single system using non-monotonic *bridge rules*. CKRs with defeasible inheritance may be realized in the MCS framework by controlling knowledge propagation by bridge rules: on the other hand, in sCKR the knowledge propagation is implicitly defined by the coverage semantics. A different non-monotonic semantics for MCS was proposed in [1], based on *argumentation semantics* of Defeasible Logic extended with distribution of knowledge and preferences across contexts. Conflicts over external literals are resolved using a local context preference, where clashes across arguments are considered. In CKR, preference is defined by the interpretation of the coverage structure. Our notion of overriding compares to a "conflict" among two arguments for conflicting literals.

Different proposals have been made in DLs to incorporate notions of "normality" in concepts and subsumptions. For example, [12] formalize in their logic $\mathcal{ALC}+\mathbf{T}_{min}$ the intuition that a prototypical element of a concept C is a "typical element" of C. The typicality operator $\mathbf{T}$ on concepts is interpreted by extending DL interpretations with a preference relation on the domain: each element in $\mathbf{T}(C)$ is a member of C minimal w.r.t. such preference. The models of $\mathcal{ALC}+\mathbf{T}_{min}$ are restricted to the ones which minimize the set of exceptional instances. Similarly to our approach, in this work membership of an element in a concept must be blocked: however, instead of using model minimization, in CKR exceptions have to be justified in terms of a semantic consequence.

Another approach to represent overriding in DLs is [2]: it proposes a family $\mathcal{DL}^N$ of non monotonic DLs defined by extending DLs with an operator NC for *normality concepts* and with *defeasible inclusions (DIs)* $C \sqsubseteq_n D$, interpreted as "normally, instances of C are instances of D, unless stated otherwise". The semantics of a defeasible inclusion $C \sqsubseteq_n D$ w.r.t. normal individuals NE is defined to manage the conflicts of inclusions on NE: to decide which DIs should be overridden, a priority relation $\prec$ is defined on DIs. The idea of axiom overriding is similar in spirit to our approach. A difference stands in the definition of precedence between defeasible axioms: in CKR, precedence is defined by the coverage hierarchy. As shown in [4], similarly to this approach we can deal with property inheritance at the instance level: however, in case of clashing inheritances that can not be resolved using preference, our semantics allows to reason by cases on all alternative models.

5 Conclusions

In this paper we extended the work introduced in [7] on CKR contextual framework with defeasible axioms in local contexts and knowledge propagation along a context hierarchy. We considered the case of general coverage hierarchies: we defined a CKR model preference relation by lifting a local preference on overridings (i.e. clashing assumptions). The ordering preserves the intuition of prioritizing the validity of defeasible axioms at more specific contexts. Then, we extended the ASP based reasoning method proposed in our previous works with an algorithm for the computation of preferred models: we have shown that this leads to a complete reasoning method for instance checking and CQ answering with respect to the proposed semantics.

There are different directions for future work. As discussed in previous sections, we are interested in defining different notion of preference on defeasible axioms: we aim to study their different semantic properties, their behavior with respect to complexity of reasoning and their different effects of knowledge propagation. We can also consider to introduce different contextual relations other than coverage (with different rules for knowledge propagation) and study their interaction. We are also interested in applying the current work on contextual hierarchies to CKRs in different DL languages (see e.g. [5]) and study different reasoning approaches and their implementation. For example, the datalog translation and the computation of preferred models by the `PrefModels` algorithm can be implemented under a common formalism in nested HEX-programs [11].

References

1. Bikakis, A., Antoniou, G.: Defeasible contextual reasoning with arguments in ambient intelligence. IEEE TKDE **22**(11), 1492–1506 (2010)
2. Bonatti, P.A., Faella, M., Petrova, I., Sauro, L.: A new semantics for overriding in description logics. Artif. Intell. **222**, 1–48 (2015)
3. Bozzato, L., Eiter, T., Serafini, L.: Contextualized knowledge repositories with justifiable exceptions. In: DL2014. CEUR-WP, vol. 1193, pp. 112–123. CEUR-WS.org (2014)
4. Bozzato, L., Eiter, T., Serafini, L.: Enhancing context knowledge repositories with justifiable exceptions. Artif. Intell. **257**, 72–126 (2018)
5. Bozzato, L., Eiter, T., Serafini, L.: Reasoning with justifiable exceptions in $\mathcal{EL}_\bot$ contextualized knowledge repositories. In: Lutz, C., Sattler, U., Tinelli, C., Turhan, A.-Y., Wolter, F. (eds.) Description Logic, Theory Combination, and All That. LNCS, vol. 11560, pp. 110–134. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-22102-7_5
6. Bozzato, L., Serafini, L.: Materialization calculus for contexts in the semantic web. In: DL2013. CEUR-WP, vol. 1014, pp. 552–572. CEUR-WS.org (2013)
7. Bozzato, L., Serafini, L., Eiter, T.: Reasoning with justifiable exceptions in contextual hierarchies. In: KR 2018, pp. 329–338. AAAI Press (2018)
8. Bozzato, L., Serafini, L., Eiter, T.: Reasoning with justifiable exceptions in contextual hierarchies (appendix). CoRR abs/1808.01874 (2018), http://arxiv.org/abs/1808.01874
9. Brewka, G., Eiter, T.: Equilibria in heterogeneous nonmonotonic multi-context systems. In: AAAI-07. pp. 385–390. AAAI Press (2007)
10. Eiter, T., Gottlob, G.: The complexity class Θ_2^p: recent results and applications in AI and modal logic. In: Chlebus, B.S., Czaja, L. (eds.) FCT 1997. LNCS, vol. 1279, pp. 1–18. Springer, Heidelberg (1997). https://doi.org/10.1007/BFb0036168
11. Eiter, T., Krennwallner, T., Redl, C.: HEX-programs with nested program calls. In: Tompits, H., et al. (eds.) INAP/WLP -2011. LNCS (LNAI), vol. 7773, pp. 269–278. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-41524-1_15
12. Giordano, L., Gliozzi, V., Olivetti, N., Pozzato, G.L.: A non-monotonic description logic for reasoning about typicality. Artif. Intell. **195**, 165–202 (2013)
13. Khriyenko, O., Terziyan, V.: A framework for context sensitive metadata description. Int. J. Metadata Semant. Ontol. **1**(2), 154–164 (2006)
14. Klarman, S.: Reasoning with Contexts in Description Logics. Ph.D. thesis, Free University of Amsterdam (2013)
15. Krötzsch, M.: Efficient inferencing for OWL EL. In: Janhunen, T., Niemelä, I. (eds.) JELIA 2010. LNCS (LNAI), vol. 6341, pp. 234–246. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-642-15675-5_21
16. Leone, N., et al.: The DLV system for knowledge representation and reasoning. CoRR cs.AI/0211004 (2002). http://arxiv.org/abs/cs.AI/0211004
17. Serafini, L., Homola, M.: Contextualized knowledge repositories for the semantic web. J. Web Semant. **12**, 64–87 (2012)
18. Straccia, U., Lopes, N., Lukácsy, G., Polleres, A.: A general framework for representing and reasoning with annotated semantic web data. In: AAAI-10, Special Track on Artificial Intelligence and the Web. AAAI Press, July 2010

Service-Microservice Architecture for Context-Aware Content Delivery in National Geoinformation Center of Bulgaria

Todor Branzov(✉), Krassimira Ivanova and Mladen Georgiev

Institute of Mathematics and Informatics at the Bulgarian Academy of Sciences, Block 8, Acad. B. Bonchev Str., Sofia 1113, Bulgaria
todor.branzov@gmail.com, kivanova@math.bas.bg, mladen.georgiev@live.com

Abstract. National Geoinformation Center of Bulgaria is a national scientific infrastructure – a consortium with a mission to produce and provide value-added products of Earth observation data gathered by various agencies governed by the state or received through international cooperation. The architecture of the information system of the center is presented – a layered structure that implements concepts of service and microservice, building an organization that reflects the federated nature of the consortium. The paper reviews various implemented design decisions within the perspective of context-aware systems. Four sources of data exploited for the determination of the context of the user are described. The paper concludes with three scenarios prepared as patterns for building context-aware services within the center.

Keywords: Service-oriented architecture · Context-aware services · Microservice · Federated systems · Earth observation centers

1 Introduction

Earth observation centers manage an ever-increasing amount of data and products with a variety of data formats. One traditional approach to data presentation is to supply all available information for a specific domain (i.e., atmosphere) that the end-user may need and to provide the tools that can filter and sort the information by some criteria. Some popular providers of Earth observation data such as NOAA [1] and Meteoblue [2] still use this approach in various products intended for public and professional use. However, that puts requirements towards users to have knowledge of system functionalities, over the topic domain, and to be the active side in the information obtainment process. In some cases, when there is a need for rapid response, activities that have to be completed during that process may prove to be too time-consuming; in instances where the user is a state authority or agency that needs a set of information to make a decision in a critical situation it may even endanger human lives. That raises a natural necessity for geoinformation centers to be designed as an environment that is suitable for the development of context-aware software systems. The invasion of mobile communication and smart devices at the

G. Bella and P. Bouquet (Eds.): CONTEXT 2019, LNAI 11939, pp. 40–50, 2019.
https://doi.org/10.1007/978-3-030-34974-5_4

beginning of the century have facilitated the access to physical context and have burned the idea that the physical and logical context can vigorously affect the behavior of the services queried by the user [3]. Especially in the last decade, it is becoming standard for the systems that provide Earth observation information to use context-sensitive data presentation that is pervasive for various users and allow them to do their job most efficiently according to their location, needs, and even condition. Both the providers mentioned above (and many others) offer products that even in its basic version, provide information according to the geolocation of the user. At least since 2005, researchers report [4] stated that NOAA also provides services that can seamlessly be integrated into context-aware services.

In our recent publication [5], we reviewed the architecture of the National Geoinformation Center of Bulgaria (NGIC) and proposed a model that uses concepts of service and microservice to solve some significant issues related to data access in federated organizations with a high level of independence. We developed the model as a result of the analysis of the type of organizational structure of 12 contemporary national and international GIC. We assumed that the model may be used as a reference input during the elaboration phase of a system design of any GIC or any system with a similar federated organization structure. One of the main goals we tried to achieve in NGIC design was to prepare infrastructure that may be used in the development of context-aware software products.

The article is structured as follows: Sect. 2 presents the main goals, purposes, and partners that form the National Geoinformation Center of Bulgaria (NGIC). Section 3 presents the conceptual model of NGIC and a sample partner organization. In Sect. 4, we use those models as a supplement of the description of implemented approaches and designed solutions that are related to the development of context-aware services in NGIC.

2 National Geoinformation Center (NGIC) of Bulgaria

The National Geoinformation Center is a newly founded Bulgarian national scientific infrastructure with the main purpose to integrate the primary national Earth observation sources and to link them into a dynamic ICT-based network. The network shall provide resources for the development of multidisciplinary, integrated data products (IDP) which can be of use to a wide range of users such as government structures, local authorities, businesses, and the public. NGIC is intended to provide infrastructure for solving important national and international tasks related to the prediction and prevention of natural and anthropogenic risks and disasters.

2.1 Towards the Foundation of NGIC

NGIC unites virtually all state-recognized national monitoring infrastructure networks (seismic, GPS/GNSS geodetic, maritime, meteorological and hydrological) with national HPC infrastructure and Bulgarian research and educational network (BREN).

The strategic purpose of NGIC is to improve the coordination and to integrate the efforts of the existing scientific infrastructure networks in Bulgaria. The long-term goal is the unification of the results and the efforts for their analysis for assessment, forecasting, and prevention of natural and anthropogenic risks and disasters. Another primary goal stated is the inclusion in relevant European networks, programs, and projects in the domain of Earth Sciences.

The specific goals of the NGIC are:

- Creating a new science-based ICT infrastructure to develop integrated products to reduce damage from natural disasters and industrial accidents based on primary geo-information products from available monitoring networks.
- Improvement of the technical infrastructure (available and in-development monitoring networks) to obtain up-to-date and reliable data on the geo-environment in the country.
- Ensuring continued access of governmental institutions and local authorities to the new ICT infrastructure and staff training to develop effective and efficient plans to prevent and protect the population from significant damage caused by natural disasters and industrial accidents.
- Expanding the territorial range of the ICT geo-database and developing new integrated products by inclusion in leading European geographic networks and participating in European infrastructure projects and interacting with infrastructural networks on the Balkan Peninsula, intensifying the partnership for data exchange and results.
- Raising the awareness of the population about natural disasters and industrial accidents through new interactive products – the creation of an information system for the promotion of the NGIC and the obtained results.
- Interaction with business through technology and commercialization of the scientific results of the NGIC's activities.

2.2 Cooperation Network

The NGIC cooperation network consists of 6 partners. They may be classified into two groups, according to the intentions and their contributions:

- DDSS (Data, Data Products, Services and Software) providers: National Institute of Geophysics, Geodesy and Geography – BAS (NIGGG); National Institute of Meteorology and Hydrology (NIMH); Institute of Oceanology – BAS (IO); Geological Institute – BAS (GI);
- ICT support providers: Institute of Mathematics and Informatics – BAS (IMI); Institute of Information and Communication Technologies – BAS (IICT).

3 NGIC as a Case Study of Service-Microservice Basic System Architecture Model

In [5] we presented a conceptual model of system architecture for GIC based on the service-microservice approach. The model allowed easier reuse of good practices when NGIC was planned and elaborated; it was also intended to improve interoperability between various centers. The architecture of the Bulgarian NGIC is provided as an illustrative case study, with some of the key concepts and decisions behind them presented with details.

We have analyzed several national and quazi-national GIC and reviewed the advantages of the federated organization model [6] when various independent entities own the sources of data, and a common goal for integration exists. We found that integration causes the necessity of building some of the subsystems with vertical governance. We also found that microservices as a concept are at least planned for implementation, if not already used, in GIC architectures, usually for giving flexibility to the system.

3.1 Conceptual Model of NGIC

The purpose of the model is to be useful in the planning or optimization phases of the NGIC information systems (IS) design cycle. We have proposed a conceptual model of the system architecture that uses both service and microservice concepts and may be altered according to the specifics of the organization environment, and development goals of particular IS.

The conceptual model consists of three layers (Fig. 1). We use the architectural concept for a service, defined as a system component (service provider) that acts to achieve desired results under a request by another component (service consumer) [7] and microservice as an independent, single purposed and loosely coupled component that supports interoperability through message-based communication [8].

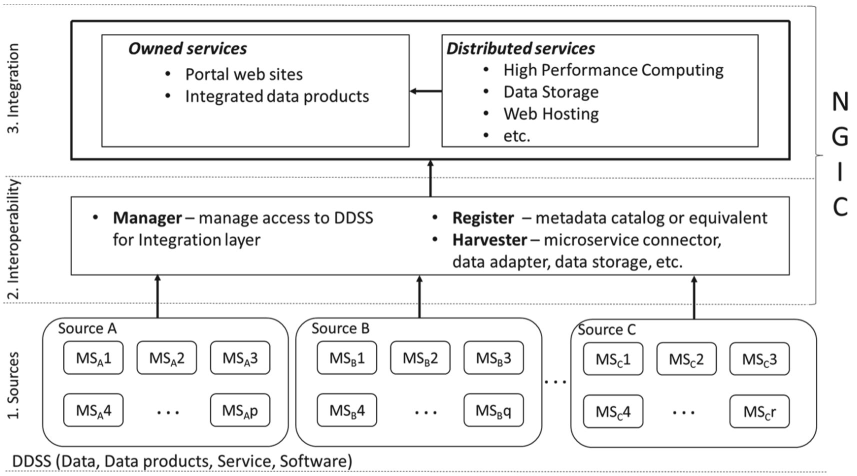

Fig. 1. Conceptual service-microservice model of NGIC

"Sources" layer contains the providers of Data, Data Products, Services and Software (DDSS) that are used by the system to produce advanced integrated products. Providers are presented as sources – collections of microservices. Microservice concept by definition is designed with a purpose for providing maximum agility to the development of the system, so, its usage gives optimal environment to scaling the systems in the scenarios as – expansion through adding new DDSS by existing source, expansion through adding a new source, shrinking through excluding either DDSS or entire source.

The components in "Interoperability" and "Integration" layers may use either monolithic service or microservice concept. Since they are centrally managed it is possible and convenient to implement the service management framework of choice.

"Interoperability" layer includes components that are mentioned in three abstract categories: "Manager", which regulate the access to the DDSS in "Sources" layer; "Register", which provide automation of discovery and selection of DDSS; and "Harvester", which includes advanced automated subsystems for data collection from the sources (like data harvesters, data adapters, storages for data buffering, etc.). From these categories, only "Manager" is mandatory, but practically in all contemporary GIC we reviewed, the two others also exist.

"Integration" layer produces integrated data products (IDP). That layer is where the value of the GIC is created and delivered. The service architecture is built around two columns – owned services, which provide access to the IDP. The IDP itself is prepared by owned and distributed services, assured by the ICT support providers or providers that are not partners in NGIC.

Taking a global perspective and in some service development scenarios, the model may be seen as an implementation of a variation of the "broker" architectural pattern [9]. Although one of the primary characteristics of the Sources is that they are always passive (i.e., they are designed only to respond to requests), the layered nature of the model along with the complete authority of NGIC over Interoperability and Integration layers allow implementation of virtually all other major patterns in manufacturing process of value-added products.

For example, even patterns like "observer" and "publisher-subscriber" that require active side that rises events could be implemented in the top two layer – one approach is to implement a service that regularly pulls DDSS from Sources and builds repository that is under authority of NGIC; on the next step that service (or another) may push events on a bus or inform observers that are either service of NGIC or some outside consumers.

3.2 DDSS Provider Approach

One perspective to the mission of NGIC is that it provides value-added products (content) that are produced using sources of Data, Data Products, Services and Software (DDSS). That sources are independent entities (partners in NGIC) whose internal organizational structure varies. The principal challenge is to provide an architecture model that is both acceptable to all of the partners, and the system may utilize it. We assumed that the model has to be (1) capable to be applied to all of the three primary organizational models – centralized, decentralized and federated [6] and (2) to be technology independent, reflecting various data exchange protocols, internal procedures, etc. that may exist in each source.

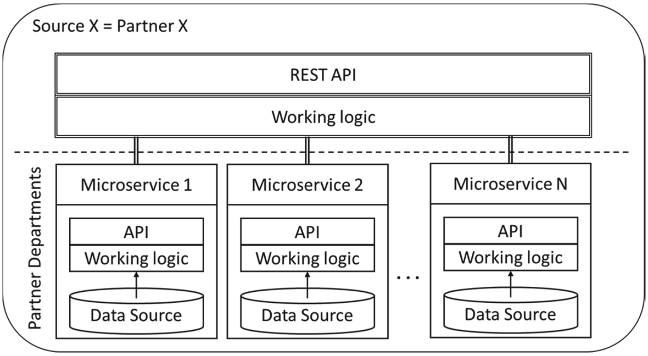

Fig. 2. Microservice approach in DDSS Provider

We provide a graphical representation of the model in Fig. 2. Again, the microservice concept is applied, this time on department level (department here is a generic functional building unit of the partner). Any department has complete technological and technical independence in the process of production of the DDSS. The design and technologies used for the development of microservice are also under the management of the department. The characteristics of the output of the microservice, however, are decided on the partner level. A department delivers the output to the business logic layer through a commonly agreed interface (API). The business logic layer is administered centrally by the partner and may alter some of the characteristics of the original outputs. The products are delivered to NGIC using commonly agreed API that is the only contact surface of the particular partner with the NGIC governed layers. Note that although in NGIC, we are using REST API, any RPC mechanism that is commonly agreed between sources and GIC may be applied.

This model is commonly agreed between the partners and NGIC. Thus, the NGIC is relieved from internal organizational issues and also has available, clearly specified, and even standardized interface to DDSS that any given source agrees to provide.

4 NGIC Like Platform for Context-Aware Services (CAS)

The basic steps of the process of consumer (we are using *consumer* intentionally interchangeably with a *user*, to denote service architecture of the NGIC) context gathering and content delivery are extracted from overview works of [10–12]. We denote four steps that are executed in sequence (Fig. 3):

- consumer request – some user (human or nonhuman) is requesting information from the NGIC;
- the context of the consumer is determined;
- the content that is most relevant to the context is identified and prepared for delivery;
- the content is delivered to the consumer.

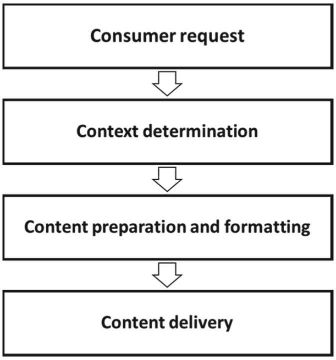

Fig. 3. Context determination and content delivery process

In general, the content provided as a result of a consumer request may be one of the following:

1. Raw data set extracted by sensor networks operated by partners in NGIC.
2. Data product, software, or service prepared and managed by one of the partners.
3. Integrated data product, which is prepared by a production process managed in NGIC, using as input DDSS from one or several partners.

4.1 Context Determination and Sources of Context in NGIC

Two major categories of context are studied and attempted to be determined by CAS in NGIC – physical context and logical context. The scope of physical context may include the current location of the device, from which the request is invoked, but also other aspects of the physical environment such as screen size, communication capabilities, network identity, GPS sensors, and others. The logical context consists of information about identity, privileges, preferences, and others [3].

The process of determination of the context is using as input a combination of data sets that describe the physical and logical context of the consumer. That data is extracted either from the consumer itself or from NGIC (see Fig. 4).

The physical context can be gathered directly from sensors in a mobile device or data in the header of an HTML request by the consumer. The logical context comes directly from the consumer (via user authentication procedure) or is deduced from interactions the user has made with NGIC owned services over time (historical context). We also assume that another source of logical context can be extrapolated (predicted context) based on clustering the users by some criteria and proposing a possible scenario according to the common behavior of the group where the user is clustered [13].

Partners of NGIC provide the last type of data that is used to determine the consumer context. That type of data may be a historical context that is stored in partner or context extrapolated on the basis of physical conditions of the environment according to data received by partners (i.e. atmospheric temperature, storm conditions, alarm codes by MeteoAlarm early warning system [14], etc. in the area around the consumer).

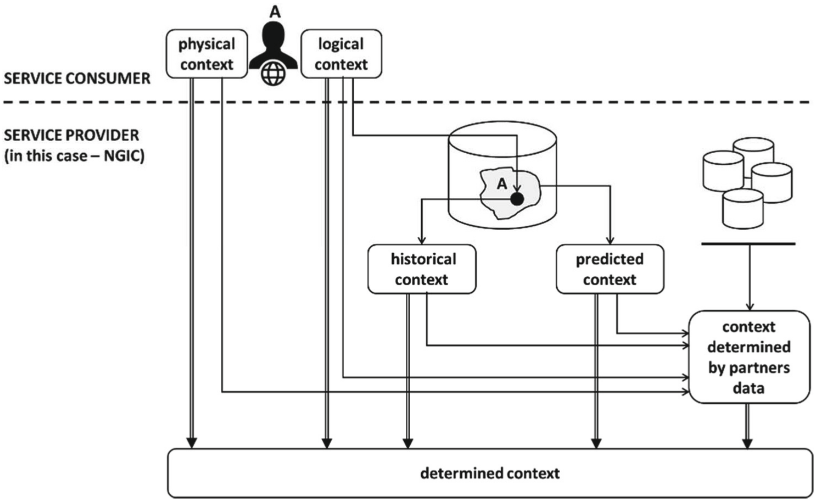

Fig. 4. Sources of context in NGIC

The determined context is used to prepare and deliver the most relevant data presentation according to the situation to the consumer.

4.2 Content Preparation and Delivery

After the determination of the current context of the consumer exist various use cases (scenarios) that affect the preparation of the content. During the early stages of the NGIC design phase, 12 scenarios were determined, with three of them assumed as basic and the other nine as variations of them, regarding manufacturing and delivery process (see Sect. 4.3. *Elaboration and Validation of the architecture*). We have to stress that the meaning of the *scenario* here is as an attempt to predict and describe the paths that data took when a user requests service from the system, with the mechanisms that eventually process that data before providing it to the user. The graphical description of the three basic scenarios is presented in Fig. 5.

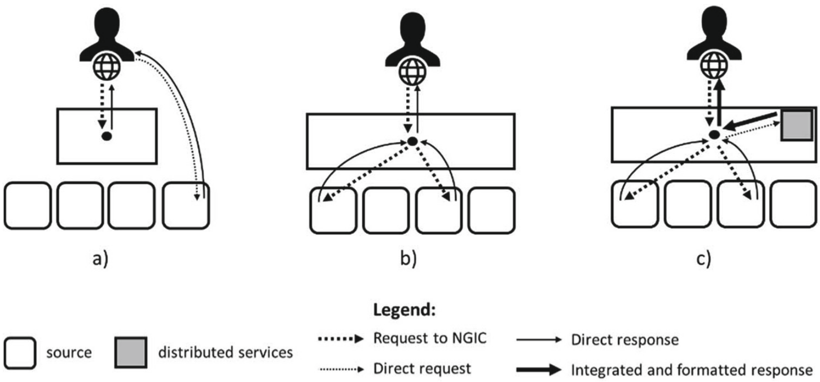

Fig. 5. Scenarios of content delivery

Scenario (a) is the simplest and depicts a situation in which after an initial user request its context is determined and using static business rule the most relevant content is specified to be a DDSS that is available at a partner and is suitable to be consumed as-is. In that scenario, the relevant NGIC service provides the user with information that allows direct contact with the partner and retrieval of needed DDSS. In that scenario, NGIC works as a catalog or meta-data catalog. A simple implementation example of that scenario is a service – web site that maps various sources of data products that are available for public access describes its content and provides a user with a URL that leads it directly to the source; that URL may be modified according to the determined context of the consumer. One of the first services that started development in NGIC was "A catalog of information sources" and was an implementation of that scenario.

Scenario (b) depicts a situation in which after an initial user request its context is determined and by using static business rules, the most relevant content is retrieved, formatted and delivered to the consumer. There is no other processing then formatting, so information again is delivered as-is provided by the partner. A sample of implementation of that service is a case in which NGIC works as a billing system, i.e., a subscriber customer of NGIC is provided with service and billed according to the relevant service contract; in that case, the scenario avails NGIC to control the flow of information and to calculate the bill.

Scenario (c) is the most sophisticated between the trio and is where the real value of NGIC and its purpose is seen. It depicts a situation in which after an initial user request its context is determined and using static business rules or/and some AI mechanism the most relevant content is specified to be delivered. AI mechanisms are either internal or distributed to partner services; they also may be external for NGIC, i.e., developed over platforms like Google's TensorFlow or Yahoo's CaffeOnSpark. The content is produced using DDSS from partners, and an integrated product is prepared. The production process may again use a combination of owned and distributed services. The final product is delivered to the customer within the parameters of the service contract.

4.3 Elaboration and Validation of the Architecture

We presented a high-level conceptual model of the software system architecture of the NGIC. We also focused on the components and scenarios in regards to the context-aware content production and delivery. Both are developed during system requirements elicitation, using three major groups of inputs.

The first group are the findings during studies of documentation and analysis of existing national and quasi-national systems with similar use (11 studied, for the full list see [5]). Some of the most significant findings are almost universal usage of federated system architecture and service-based architecture, and various techniques used for decoupling and outsourcing.

The second group includes the results of several interviews with the main stakeholders – the four partners in NGIC that provide DDSS and representatives of three of the state agencies that are most probably going to use the products of NGIC. The interviews with partners were structured around a standardized questionnaire developed in collaboration with the experts of the leading partner – NIGGG. The questionnaire was split into three parts - a set of questions regarding existing technology process for manufacturing data products, a set of questions regarding the clients and the methods, that are used to deliver the products; and a set of questions that is aimed to identify the issues related with manufacturing and delivery, especially potential problems with manufacturing of IDP. The interviews with state agencies representatives included a questionnaire, with a focus on gathering information on some of the most significant procedures in the agencies and data products used and needed in them. The interviews continued with a series of open-format discussions to find approaches and delivery methods that would optimize those procedures. Along with some product characteristics, probably the most significant finding was that the frequency of usage of data products must not be used mechanically for determination of its importance. The output of the interviews was used for building a set of use cases and content delivery scenarios.

The third group inputs are applicable and universal good practices and patterns in system design and expert knowledge.

Validation of the system architecture based on measured characteristics is not deemed feasible at this stage of its development. Thus qualitative approaches were discussed and accepted. The proposed architecture was presented and discussed in a series of three workshops – two with a panel of IT experts and one with a mixed panel of domain (earth sciences) experts and IT experts. As a result of workshops, the architecture was optimized, and proposed for formal acceptance by the NGIC governing body. Also during the workshops was planned the establishment of continuous feedback loop with consumers, implementation of scenario-based software architecture evaluation methods of Software Architecture Analysis Method (SAAM) and Architecture-Level Modifiability Analysis (ALMA) [15] and possible implementation of quantitative methods and metrics.

5 Conclusion and Discussion

The main goal at this early stage was to develop a conceptual model of the software system architecture that we may use as a foundation for more elaborated technical models and documentation. It was not deemed feasible to develop technical models or to establish models of context and implement methods for context since most of the planned products lack complete specification.

The concept of context-awareness is promoted to consumers that are engaged with civil protection with positive feedback. The actual possibilities for improvement of procedures of such agencies seem very positive. The service architecture that is used for delivery of NGIC products would allow non-sophisticated integration with various existing early warning systems and may feed them with most suitable content or may allow them to alter their systems to "outsource" context-awareness to NGIC.

Although the results presented in this paper are based on the early phase of system development, we presume they may be used as a case study of emerging GIC that is planned and designed to be context-aware from the beginning. Assumptions and suggestions are available for consideration and may provide value for the design of any federated systems in general.

The governing body of NGIC accepted the presented conceptual architecture. Various IT solutions and prototypes are in a process of implementation, and the basic infrastructure is planned to become operational in 2020. One prototype service is available to the public for test purposes and five are planned to be launched in the period 2019–2021. Three of these services are planned to be converted to context-aware and prototypes that are tested in a laboratory environment are being built. PAAS like already mentioned Google TensorFlow are considered; however, the results of tests are still not conclusive towards a decision of implementation.

Acknowledgments. This work is funded by Contract DO1-161/28.08.2018 "NGIC – National Geoinformation Center for monitoring, assessment and prediction natural and anthropogenic risks and disasters" under the Program "National Roadmap for Scientific Infrastructure 2017–2023", financed by Bulgarian Ministry of Education and Science.

References

1. NOAA. https://www.noaa.gov/. Accessed 08 Aug 2019
2. Meteoblue – about us. https://content.meteoblue.com/en/about-us. Accessed 08 Aug 2019
3. Jones, K.: Building a context-aware service architecture, 6 p., 12 December 2008. https://ibm.com/developerWorks
4. Sheng, Q.Z., Benatallah, B.: ContextUML – a UML-based modeling language for model-driven development of context-aware web services. In: International Conference on ICMB 2005, pp. 206–212. IEEE (2005)
5. Branzov, T., Ivanova, Kr., Georgiev, M.: Service-microservice basic system architecture model for geoinformation centers. In: Proceedings of XIX International Multidisciplinary Scientific GeoConference SGEM 2019, vol. 19, no. 2.1, pp. 587–594 (2019)
6. Rychkova, I., Zdravkovic, J., Speckert, T.: Challenges of EA methodologies facing progressive decentralization in modern organizations. In: CEUR Workshop Proceedings, vol. 1023, pp. 18–28 (2013)
7. He, H.: What is Service-Oriented Architecture? O'Reilly Media, Inc. (2003). https://www.xml.com/pub/a/ws/2003/09/30/soa.html
8. Nadareishvili, I., Mitra, R., McLarty, M., Amundsen, M.: Microservice Architecture Aligning Principles, Practices, and Culture, 126 p. O'Reilly Media Inc. (2016)
9. Avgeriou, P., Zdun, U.: Architectural patterns revisited – a pattern language. In: Proceedings of 10th European Conference on Pattern Languages of Programs (EuroPlop 2005), pp. 431–470 (2005)
10. Brézillon, P.: Context in human-machine problem solving: a survey. Knowl. Eng. Rev. **14**, 1–37 (1996)
11. Baldauf, M., Dustdar, M., Rosenberg, S.: A survey on context-aware systems. Int. J. Ad Hoc Ubiquit. Comput. **2**(4), 263–277 (2007)
12. Snidaro, L., García, J., Llinas, J.: Context-based information fusion: a survey and discussion. Inf. Fusion **25**(C), 16–31 (2015)
13. Shishkov, B., Larsen, J.B., Warnier, M., Janssen, M.: Three categories of context-aware systems. In: Shishkov, B. (ed.) BMSD 2018. LNBIP, vol. 319, pp. 185–202. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-94214-8_12
14. Alfieri, L., Salamon, P., Pappenberger, F., Wetterhall, F., Thielen, J.: Operational early warning systems for water-related hazards in Europe. Environ. Sci. Policy **21**, 35–49 (2012)
15. Ionita, M.T., Hammer, D.K., Obbink, H.: Scenario-based software architecture evaluation methods: an overview. In: Workshop on Methods and Techniques for Software Architecture Review and Assessment at the International Conference on Software Engineering, pp. 1–12 (2002)

Hybrid Expressions

Tadeusz Ciecierski (✉)

University of Warsaw, Warsaw, Poland
taci@uw.edu.pl

Abstract. The paper discusses the idea of hybrid names. First, the three theories of hybrid names (according to Künne, Kripke and Textor) are briefly discussed and compared. Second, the paper briefly discusses some problems that the theories face. Third, an alternative hybrid view is outlined. According to that view, utterances are contextually perduring objects. It is argued that in order to determine the contextual distribution of a particular utterance, one has to take into account the admissible distributions of contextual parameters, their potential referents and the speaker's referential intentions. Finally, the merits of the view are briefly discussed. Two of the most important are: the analysis of cases of multiple occurrences of indexicals and demonstratives, and the solution to the so-called problem of missing demonstrations.

Keywords: Hybrid names · Indexicals · Demonstratives

1 Hybrid Names

According to Frege, '(...) circumstances accompanying the utterance (...) are used as means of expressing the thought'. In recent decades this and other cryptic remarks by Frege have inspired several philosophers (cf, Künne 1992, 2010; Textor 2007, 2015; Kripke 2011) to formulate various versions of the *Hybrid Name Hypothesis (HNH).* According to the *HNH*, indexicals and demonstratives are special kinds of singular expressions that contain as proper parts *both* linguistic components and broadly conceived extra-linguistic circumstances. Only such hybrid expressions can properly be said to be the bearers of sense as well as arguments of the reference relation. Below, I shall propose a certain interpretation of the Hybrid Name Hypothesis that departs from those of Künne, Textor and Kripke.

Unsurprisingly, the Hybrid Name Hypothesis might be developed in several ways, depending on the answers to two crucial issues:

Q1. What are the circumstantial components of hybrid names?
Q2. What is the role of circumstances in ascribing semantic properties (content/sense and reference) to hybrid names?

The work on this paper was funded by National Science Center, Poland, grant under award number 2018/29/B/HS1/01868.

G. Bella and P. Bouquet (Eds.): CONTEXT 2019, LNAI 11939, pp. 51–61, 2019.
https://doi.org/10.1007/978-3-030-34974-5_5

Answers to the first question are usually given by enumeration, that is, by providing a list of appropriate circumstances that correspond to particular types of indexicals and demonstratives. Künne, Textor and Kripke generally agree[1] that in the case of demonstratives it is a demonstration that plays the role of the circumstantial component of the hybrid expression. At the same time, they differ in taking the demonstration to be either a linguistic item (Kripke) or a non-linguistic sign (Textor and Künne). In the case of proper indexicals, the main difference is between the object-theorists (Kripke and Künne) and the use-theorists (Textor). The former take times, locations and speakers to be parts of the hybrid names. The latter take the *uses* of circumstances to be such parts.

Answers to the second question differ when it comes to indexicals and demonstratives. Since demonstrations are interpreted as special kinds of signs, the answer to Q2 must say something about the relation between the demonstration, demonstratum and the demonstrative expression. Mark Textor (Textor 2007: 954, who follows Berckmans 1990 in this) suggests, for instance, that when a demonstrative occurs in the utterance *u*, then one intends (we might call this intention the *attention directing intention*)—by making a demonstration *d*—to draw the addressee's attention to some particular object *x* and to communicate a proposition concerning it (we might call this intention the *communicative intention*). Assuming that the sameness of attention-directing intention and the sameness of communicative intention is necessary for two demonstrations to be of the same type, we arrive at the sufficient (but not necessary) condition for sameness of sense for hybrid names: two hybrid names have the same sense if they contain synonymous expressions as linguistic components and the same demonstrations as non-linguistic components. Depending on the existence of the object of the *attention-directing intention*, the sense might have a corresponding referent (or not) and the corresponding proposition concerning an object might be expressed (or not). If additional situational constraints are met, the expressed proposition is not only expressed but also successfully communicated. Künne, on the other hand, argues that the identity of both components of hybrid names is neither necessary nor sufficient for the identity of senses of hybrid names (cf. Künne 1992: 728).

For Kripke, whatever is part of a thought-expression is a piece of language. Hence, demonstrations qua components of hybrid names are pieces of language (the same applies, of course, to all other circumstances). What is more, Kripke takes the linguistic component of the hybrid name to be a functional sign while interpreting the circumstantial component as the symbol for the complement of the functional sign. He applies to this theory his idea of autonymous designation (extending the direct quotation cases to other ones) according to which the circumstantial component of a hybrid name refers to itself, just like the expression 'cat' refers to itself in ≪'cat'≫. A necessary condition for grasping the sense of the hybrid name is, according to Kripke, the fact of being acquainted with the circumstantial component of the hybrid name. The reference of the entire hybrid name is, therefore, an object that is appropriately related[2] to the known

[1] Künne's (1992) original view was different, as he took demonstratum to be part of the hybrid name. He changed his mind, however, and in Künne (2010) he parts company with Textor and Kripke.

[2] What the nature of the relation in question is remains unexplained, although in non-demonstrative cases like the ones of 'I', 'here' and 'now' the relation is that of identity.

demonstration. The same can be said of non-demonstrative cases, the only difference being that the role of demonstration is now played by aspects of the circumstances, like speakers, times and locations.

When it comes to proper indexicals, Künne sides with Kripke as object theorist but differs from Kripke in *not* treating circumstances as pieces of language and *not* assuming the theory of autonymous designation. However, at least in cases of basic uses of ‘I’, ‘here’ and ‘now’, they agree that the semantics of hybrid names involves rules along the following lines (cf. Textor 2015: 828):

∀*s* ([‘I’ ⊕ *s*] refers to *s*)
∀*t* ([‘now’ ⊕ *t*] refers to *t*)
∀*p* ([‘here’ ⊕ *p*] refers to *p*)

where ‘[(…) ⊕ (---)]’ represents the mereological fusion of (…) and (---)[3]. Künne also believes that the difference between proper indexicals and demonstratives is that hybrid names that contain the former (but not the latter) ‘are just as immune against the risk of *Bedeutungslosigkeit* as are quotational designators’ (Künne 2010: 545).

Textor (2015), as a use-theorist, opposes both Kripke and Künne here. For him, *utterances* and their *properties* are natural signs of objects, like locations, times and speakers. For instance, a particular articulation and writing style might be said to naturally signify a particular person, while a verbal utterance might indicate the place qua the source of the verbal message; it might also indicate the time as the time simultaneous with the utterance event, etc. As such, they (utterances and their properties) are intentionally used to signify particular objects. Now, indexicals also have a clear linguistic meaning that encodes information that a certain kind of object is the referent of the indexical token. Linguistic meaning together with knowledge of the use of the circumstances enables one to grasp the thought expressed, that is to capture that the thought concerns a particular object given in this particular manner.

Let us consider the three theories in the context of the following scenario[4]:

Three persons: A, B and C are working at a quarry. All of them observe a detonation. A sees what is going on at the quarry but cannot hear what is going on. B, on the other hand, does not see what is going on but she can hear it. C can neither see the quarry nor hear what is going on, but observes the seismograph that responds to ground motions in the quarry. A, B and C all use text messages to communicate. Now, consider the three parallel ways in which the scenario may evolve:

In Case 1, A, after seeing a huge explosion writes to B and C:

A: That was spectacular!

In Case 2, it is B who starts the conversation, with:

B: That was spectacular!

While in Case 3, it is C who writes:

C: That was spectacular!

---

[3] Kripke uses the ordered pair notation here, but stresses that nothing is dependent on that.
[4] This scenario is a slightly more complex variant of the case described by Künne (1992): 728.

Now, all three described theories agree that the hybrid forms of A's, B's and C's utterances are:

(Case 1) ['That' $\oplus$ *A's demonstration*] was spectacular!
(Case 2) ['That' $\oplus$ *B's demonstration*] was spectacular!
(Case 3) ['That' $\oplus$ *C's demonstration*] was spectacular!

However, they each predict different things regarding the relation between the *thought* (t_A) *expressed by A*, the *thought* (t_B) *expressed by B* and the *thought* (t_C) *expressed by C*[5]. First, Künne's theory *by itself* does not make any predictions regarding the sameness and difference of t_A, t_B *and* t_C. One has to invoke additional considerations in order to arrive at some decisive answer here. The important point is that such considerations, no matter what their rationale happens to be, are not in any way implied by Künne's version of *HNH*. Textor, on the other hand, carefully states the sufficient conditions for the sense identity of hybrid names (two hybrid names have the same sense if they contain synonymous expressions as linguistic components and the same demonstrations as non-linguistic components) and, prima facie, the fact that the condition is not necessary opens up a Fregean possibility that t_A, t_B *and* t_C might be the same (such a situation is analogous to Frege's well-known example with 'today' and 'yesterday'). However, a closer look at Textor's considerations and the stress he puts on conveying the right way of thinking about an object as being decisive for the sameness of senses suggests that he would rather either vote for a difference of t_A, t_B *and* t_C or, more interestingly, interpret cases 1, 2 and 3 as *underspecified* in one important aspect. Let me explain. In the scenario, I have not specified whether A, B and C have the correct belief about the conditions in which they are, respectively. If they all do, then the following hypothesis emerges: A's, B's and C's demonstrations involve a single but *compound* manner of thinking about the explosion as the object presented thusly directly visually from one perspective, thusly auditorily for another perspective and thusly indirectly visually for yet another perspective. This means that t_A, t_B and t_C might in fact be identical thoughts involving compound ways of presenting the explosion. If A's, B's and C's beliefs about the situation differ, then the theory predicts that at least two thoughts in question must be distinct. Kripke's theory, on the other hand, probably denies the sameness of t_A, t_B and t_C—in all three cases only one of the speakers is acquainted with the demonstration involving different modes of perceptual presentation. This seems, at least, to be a natural extension of his approach to indexical cases (Kripke 2011: 277).

[5] The problem might be stated as a question if (in each of the cases 1–3) the thought expressed by A/B/C is the same as the one decoded by the remaining two agents. I am avoiding this way of phrasing the problem because it involves additional issues connected with the speaker/addressee roles in the communication.

2 Problems

The three versions of *HNH* contain several insightful philosophical points[6]. However, I think that neither is free of problems. Although I do not consider the problems to be knockdown arguments against each theory, they might indicate the need for an alternative version of *HNH*.

Let me start with Künne's theory as applied to proper indexicals (bear in mind that in the case of demonstratives, Künne's revised theory is basically the same as Textor's). An important part of the theory is the analogy with quotational designators and their immunity to becoming vacuous terms. There can be no doubt that in regular cases, this is the property we expect indexicals to have. However, there is a sharp contrast between indexicals like 'I', 'here' and 'now' (or at least: basic uses of such indexicals) and other deictic expressions. Consider the cases of 'yesterday' and 'tomorrow'. What if, for instance, one utters 'yesterday' in a Russellian, five-minute hypothesis scenario, or uses 'tomorrow' just before the Earth's destruction in some cosmic catastrophic event? In both cases the indexicals fail to refer, but the hybrid names, namely ['yesterday' $\oplus$ *t*] and ['tomorrow' $\oplus$ *t*], are *not* defective in any sense (in both cases *t* represents the moment of utterance). Similar cases can be constructed for 'you', 'we', 'there' and common non-basic uses of 'here' and 'now', where we refer to areas or periods of time that properly comprise the place or moment of utterance. Although some such examples ('you', for instance) might be interpreted demonstratively, the extension to other cases (pace Penco 2013) remains controversial. Neither Textor's nor Kripke's treatment of proper indexicals, I think, treats indexicals as immune to reference failure. This could be used as an argument in their favour.

Kripke's theory, as pointed out by Künne, has a serious conceptual problem, however. It is difficult to make any sense of the remark that circumstances accompanying the use of an indexical are pieces of language and things then can be understood or misunderstood[7]. Moreover, there is a selection problem in Kripke's theory (cf. Textor 2015: 831). A necessary condition for grasping the sense of the hybrid name is, according to Kripke, the fact of being acquainted with the circumstantial component of the hybrid name. Yet this does not universally provide the acquaintance relation that ensures that we think of the circumstantial component in the right manner.

Textor's theory is free of such problems. But there are some challenges it has to face. The most important concerns Textor's treatment of the so-called cases of missing demonstration. Suppose someone utters the sentence, *That was way louder than that*, after hearing a sequence of two consecutive noises, and does not perform any accompanying gestures. Textor approaches cases of missing demonstration by claiming that the utterance itself *plays the role* of or simply *is* the demonstration (cf. Textor 2007: 957). This approach fails, however, to provide the correct analysis of the relational statements like *That was way louder than that*: first, because there is only one utterance and two noises; and second, because depending on the situation (and speaker's intention) the utterance might link the first demonstrative with the first noise and the second with the

---

6 See Predelli (2006) for additional reasons for endorsing *HNH*.

7 'How could a time possibly designate anything? A time of utterance is something one can neither understand nor misunderstand, so how can it have a Fregean *Bedeutung*?' (Künne 2010: 541).

second noise, or *vice versa*. This means that the order in which the demonstratives occur in the utterance is not helpful here. Hence, appealing solely to the presence and the structure of the utterance cannot by itself solve the problem of missing demonstrations.

3 An Alternative: Contextual Perdurantism

Consider the utterance *u* of the following sentence:

I was wrong yesterday and you are wrong today.

Ignore for a moment the distinction between aspects of contexts qua contextual parameters and aspects of contexts qua components of hybrid names. Whatever the respective circumstances are, one needs *at least* three kinds of contextual parameters to interpret the utterance *u*: the speaker parameter, the addressee parameter and the time of utterance parameter. Roughly speaking, one could provide a list of the respective parameters just by looking at the indexicals and demonstratives that occur in the utterance and by considering what the linguistic meaning of the respective indexicals says. This, obviously, does not exhaust the need for other contextual factors in the interpretation of *u* (for instance, information about the issue(s) about which the speaker and the addressee are wrong are also provided contextually). Nor does it tell us how to pick the speaker and the addressee (cf. Penco 2015). However, for the sake of exposition let us ignore such complications. Now, we might treat each relevant aspect of context as an independent dimension in contextual space. It is, of course, difficult to say what the possible dimensions of such contextual space are exactly, but the relevant point is that for each particular utterance *u* it is easy to enumerate the indexical constituents of *u* and (by looking at the linguistic meaning of such constituents) to identify the contextual dimensions relevant for the interpretation of *u*.

One might think of each such dimension in terms of sets generated from the class of possible objects that are potential values of contextual parameters, that is speakers, times (of utterance), places (of utterance), etc. In order to make the dimensions disjoint one might think of values not in terms of regular objects (this enables one and the same element to be, for instance, the speaker and the addressee) but rather in terms of *qua-objects*, that is objects falling under a particular description (cf. Fine 1982). Now, starting with the classes of possible values of contextual parameters allows to consider all permutations with repetition of a length that is no longer than the number of occurrences of the particular indexical or demonstrative in a given (possible) utterance. The class of all such permutations (of the respective qua-objects) could be equated with a particular contextual dimension and the collection of all such dimensions – with the contextual space. The table below contains step-by-step description of the method of generating contextual dimensions and the contextual space:

Contextual dimensions and contextual space generation	**Example**
Sentence (as used in the utterance *u*) that contains *n* indexicals	'I was wrong yesterday and you are wrong today', used in the utterance *u* by *a* addressing *b* at the time *t*; contains four indexicals
Each indexical is linked to a certain contextual parameter	'I' is linked to the speaker-parameter 'you' is linked to the addressee-parameter 'yesterday' is linked to the time-of-utterance-parameter 'today' is linked to the time-of-utterance-parameter
All indexicals linked to a single contextual parameter are linked to a single contextual dimension	'I' is linked to the speaker-dimension 'you' is linked to the addressee-dimension 'yesterday' and 'today' are linked to the time-of-utterance-dimension
A contextual dimension consists of sequences of potential values of contextual parameters of the length corresponding to the number of occurrences of indexicals linked to the contextual dimension	Speaker-dimension contains *at least*[a] $<a>$ (as well as other possible sequences of possible speakers) Addressee-dimension contains *at least* $<b>$ (as well as other possible sequences of possible addressees) Time of utterance dimension contains *at least* $<t, t>$ (as well as other possible sequences of possible pairs of times of utterance)
Contextual space is the collection of all contextual dimensions	Contextual space consists of at least three contextual dimensions: speaker, addressee and time of utterance dimension

[a] 'At least' because the procedure applies to all possible utterances.

A contextual dimension contains the class of all possible distributions of a particular contextual parameter. However, in a given situation only certain distributions are potentially relevant, or as I prefer to say: *admissible*. Take again our utterance *u*. If it is uttered on a particular day *d* that contains the moment of utterance t_d, then the admissible distribution would be simply (*a* is the speaker, *b* – the addressee):

$$<\text{'I was wrong yesterday and you are wrong today'}, \{<a>\}, \{<t_d, t_d>\}, \{<b>\}>$$

However, more complicated cases are possible. Imagine that *a* utters *u* to *b* just before midnight of the day *d* and finishes it after the midnight on the next day *d*'. This means that the set of admissible distributions within that dimension would contain three elements[8]:

[8] How is the class of relevant distributions determined? Determination involves considerations pertaining to the relevance of particular configurations of contextual factors in the situation. As such, it is pragmatically determined.

$<$‘I was wrong yesterday and you are wrong today’, $\{<a>\}, \{<t_d, t_d>, <t_d, t_{d'}>, <t_{d'}, t_{d'}>\}, \{<b>\}>$

Now, for each such element one can assign the class of *potential* referents (not to be confused with the values of contextual parameters!) of the indexical or demonstrative expressions that occur in the utterance. Potential referents here are established purely semantically, that is just by consulting the linguistic meaning of indexicals and demonstratives (rules like: ‘I’ refers to the speaker of the context, etc.).

The final step in reference assignment involves the speaker’s referential intentions. Normally, the person who is the speaker has in mind particular objects as the intended referents of indexicals and demonstratives. However, in both the former and latter cases the intentions are restricted either by conventional constraints present in the meaning of indexicals (I cannot refer to Socrates when using ‘I’[9]) or by some looser pragmatic constraints (I normally cannot refer to the planet Kepler-186f by pointing to my dog Bobo and uttering ‘This must be habitable’). The required result is achieved here by comparing the class of classes of potential referents (which are conventionally established) and the class of intended referents. If (and only if) the latter class is a member of the former can the speaker successfully refer to particular objects which now become the *actual referents* of indexicals and demonstratives. Having the actual referents enables us to go back to the class of admissible distributions and select the *actual distribution of contenxtual parameters*.

According to our theory, utterances are *aggregates of contextual parts across actual distributions of contextual parameters* or, as one might put it, *contextually perduring objects*[10]. In addition to the traditional spatiotemporal dimension, contextual perdurance enables personal (speaker and addressee) as well as other (social, standard related etc.) dimensions. In this way, we arrive at a potentially novel concept of *hybrid expression* and a novel concept of hybrid name. Circumstances here are parts of the expression of the thought, not as mereological parts of [expression $\oplus$ circumstance] fusions but rather as aspects or parts of contextually perduring utterances[11].

One could, however, attempt to reconstruct both the objectual idea and use-theory idea of the hybrid name within our theory of utterances. Given that we have the actual distribution and unique syntactic decomposition of the utterance, we might enumerate

---

[9] There are problematic cases here, of course (cf. Mount 2008). However, they could be explained away as cases of non-standard uses of indexicals.

[10] For a general discussion on perduring and perduratism see: Lowe (2002): 41–48.

[11] The contextual perdurantism is a thesis solely about utterances (‘Utterances have contextual parts’) and, as such, is independent from the standard perdurantism (‘Physical objects have spatial and temporal parts’). This can be seen, firstly, in cases of utterances that contain only non-temporal and non-spatial indexicals. Such utterances have contextual parts (possibly: speaker, addressee, epistemic standard etc. parts) but our theory is silent about them having temporal or spatial parts at all. Secondly, since the points in each contextual dimensions are not regular objects but sequences of qua-objects even the temporal or spatial parts of some utterances (the ones that contain temporal or spatial indexicals) are in fact instants qua time of utterance and locations qua place of utterance. This contrast with standard perdurantism which does not treat time and space as consisting of qua objects.

all the hybrid names present in the utterance. So, given that $\{<t_d, t_{d'}>\}$ is the actual distribution within the time of utterance dimension:

$<$‘I was wrong yesterday and you are wrong today’, $\{a\}$, $\{<t_d, t_{d'}>\}$, $\{b\}>$

might, if we follow the objectual view, be broken down into a structure with the terminal elements being: [‘I’ $\oplus a$], [‘yesterday’ $\oplus t_d$], [‘you’ $\oplus b$], [‘today’ $\oplus t_{d'}$], while in the use-theoretical view we simply replace persons and times with uses of circumstances. One has to keep in mind, however, that the hybrid names in question are always determined with respect to the actual distribution, which is established both conventionally (by appealing to the linguistic meaning of indexicals/demonstratives) and intentionally (by appealing to pragmatic considerations in establishing the class of admissible distributions as well as to the referential intentions of the speaker).

The theory outlined above has the following merits. First, as illustrated above, it deals well with cases of distributed utterances (cf. McCullagh (forthcoming)), that is cases of utterances where more than one indexical or demonstrative occurring in the sentence is linked to a single aspect of the context which may take different values relevant for the interpretation of the respective indexicals or demonstratives (like in the sentence: ‘It is now 3 o’clock and it is now past three’ uttered by the speaker who intentionally started speaking at 3 o’clock but finished one past three). Moreover, since the cases of multiple occurrences of indexicals and demonstratives are nothing more than a special cases of distributed utterances the theory applies also to utterances with multiple occurrences of indexicals and demonstratives. Second, by clearly separating the distribution of contextual factors and reference it is not committed to the claim that indexicals are immune to reference failure. Interestingly, it also enables cases where, due to a discrepancy in referential intentions and potential referents, the context of utterance embraces not a single distribution but a class of distributions. This is possible if the class of admissible distributions contains at least two elements, and where there is no way back from potential referents to the actual ones. Third, the theory does not presuppose any kind of acquaintance account of circumstantial components of hybrid names and their referents. I agree with Textor that the problem of providing the right way of thinking about referents is one of the main challenges of the theory of hybrid names and hybrid expressions. However, I disagree with him in not making the following important assumption: that the problem is to single out a particular *manner of thinking* about a single object from numerous ways of thinking about it. I prefer instead to think of the problem in terms of a choice of *entities* that might occur in a given distribution. Roughly speaking, it is common to think of the choice of the values of contextual parameters in terms, for instance, of dates and hours, areas in the objective space, etc. However, nothing prohibits us from also allowing as such values the entities that are perspectival, that is entities that are derived from agent’s point of view[12]. Elements of temporal A-series (cf. McTaggart 1908), locations in egocentric time or space (cf. Russell 1948), egos or agents occupying roles (cf. Prior 1968) and other perspectival entities occupy prominent place in the history of philosophy. Once we enable such indexical metaphysics as a background

[12] A sophisticated version of a theory that enables this was defended by Tomis Kapitan (cf. Kapitan 2001; Babb 2019).

to the theory of contextual space, both the referents and values of contextual parameters start to have the agent’s perspective built in from the very beginning[13].

Let me close this paper with some remarks on the problem of missing demonstrations. Here I disagree with Textor in thinking that the very utterance plays the role of demonstration, but I agree with him that some demonstration is in fact present. Demonstrations are, as all participants of the debate agree, *actions*. Among the various kinds of actions an agent can perform are *omissions*, where agents act by omitting to perform certain kinds of actions. Cases of missing demonstration, I would argue, are cases where demonstrations are nothing more but omissions. So, admissible distributions for our problematic utterance of *That was way louder than that* are:

<‘That was way louder that’, {<demonstration qua ommision$_1$, demonstration qua ommision$_2$>, <demonstration qua ommision$_2$, demonstration qua ommision$_1$>}>

Here, the absence of other possible actions of demonstration invites the addressee to formulate a hypothesis pertaining to the referential intentions of the speaker and the reasons she has to avoid making other demonstrating actions. The auxiliary assumptions that help to arrive at such a hypothesis concern facts about the attention-attracting features of the situation. This explanation, I think, fits well with Textor’s general idea that demonstrations are signs. They are indeed signs in the traditional (in fact, Stoic!) sense, as they induce participants of the communicative event to perform inferences that start from the assumption of the presence of some entity (a demonstration that might be an omission) and arrive at a hypothesis about the presence of some other entity or event (a demonstratum).

References

Berckmans, P.: Demonstrative utterances. Philos. Stud. **60**, 281–295 (1990)

Babb, M.: Kapitan on indexicals and indexical thought: a retrospective. South. J. Philos. **57**(2), 279–294 (2019)

Fine, K.: Acts, events and things. In: Leinfellner, W., Kraemer, E., Schank, J. (eds.) Proceedings of the 6th International Wittgenstein Symposium, pp. 97–105. Hölder-Pichler-Tempsky, Wien (1982)

Kapitan, T.: Indexical metaphysics. In: Meixner, U. (ed.) Metaphysics in the Post-metaphysical Age. Proceedings of the 22nd International Wittgenstein Symposium, pp. 81–88. Obvethpt Verlagsgesellschaft mbH & Co. KG, Vienna (2001)

Kripke, S.: Frege’s theory of sense and reference: some exegetical notes. In: Kripke, S. (ed.) Philosophical Troubles, pp. 254–291. Oxford University Press, Oxford (2011)

Künne, W.: Hybrid proper names. Mind **101**, 721–731 (1992)

Künne, W.: Sense, reference and hybridity. Dialectica **64**, 529–551 (2010)

Lowe, J.E.: A Survey of Metaphysics. Oxford University Press, Oxford (2002)

McCullagh, M.: Distributed Utterances (forthcoming)

McTaggart, J.M.E.: The unreality of time. Mind **17**, 457–473 (1908)

[13] Such metaphysics can be combined with the idea of primitive senses (Tichy 1986) to provide an attractive solution to the problem of the nature of higher-order senses.

Mount, A.: The impurity of ‘Pure’ indexicals. Philos. Stud. **138**, 193–209 (2008)
Penco, C.: Indexicals as demonstratives: on the debate between Kripke and Künne. Grazer Philos. Stud. **88**, 55–73 (2013)
Penco, C.: Context dependence, MOPs, WHIMs and procedures. In: Christiansen, H., Stojanovic, I., Papadopoulos, George A. (eds.) CONTEXT 2015. LNCS (LNAI), vol. 9405, pp. 410–422. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-25591-0_30
Predelli, S.: Hybrid indexicals and deixis. Erkenntnis **65**, 385–403 (2006)
Prior, A.N.: Egocentric logic. Noûs **2**(3), 191–207 (1968)
Russell, B.: Human Knowledge: Its Scope and Limits. Simon and Schuster, New York (1948)
Textor, M.: Frege’s theory of hybrid proper names developed and defended. Mind **116**, 947–981 (2007)
Textor, M.: Frege’s theory of hybrid proper names extended. Mind **124**, 823–847 (2015)
Tichy, P.: Frege and the case of the missing sense. Grazer Philos. Stud. **27**(1), 27–47 (1986)

Context-Driven Corpus-Based Model for Automatic Text Segmentation and Part of Speech Tagging in Setswana Using OpenNLP Tool

Mary Ambrossine Dibitso, Pius Adewale Owolawi and Sunday Olusegun Ojo(✉)

Faculty of ICT, Tshwane University of Technology, Pretoria, South Africa
madibitso@gmail.com, {owolawipa,ojoso}@tut.ac.za

Abstract. Setswana is an under-resourced Bantu African language that is morphologically rich with the disjunctive writing system. Developing NLP pipeline tools for such a language could be challenging, due to the need to balance the linguistics semantics robustness of the tool with computational parsimony. A Part-of-Speech (POS) tagger is one such NLP tool for assigning lexical categories like noun, verb, pronoun, and so on, to each word in a text corpus. POS tagging is an important task in Natural Language Processing (NLP) applications such as information extraction, Machine Translation, Word prediction, etc. Developing a POS tagger for a morphologically rich language such as Setswana has computational linguistics challenges that could affect the effectiveness of the entire NLP system. This is due to some contextual semantics features of the language, that demand a fine-grained granularity level for the required POS tagset, with the need to balance tool semantic robustness with computational parsimony. In this paper, a context-driven corpus-based model for text segmentation and POS tagging for the language is presented. The tagger is developed using the Apache OpenNLP tool and returns the accuracy of 96.73%.

Keywords: NLP · Context · Text segmentation · POS tagging · Setswana · OpenNLP

1 Introduction

Setswana is a Southern Bantu language which is closely related to other two Sotho group languages such as Sesotho (also known as Southern Sotho) and Sesotho sa Leboa (also known as Northern Sotho). Setswana is spoken in at least four countries which are South Africa, Botswana, Namibia, and Zimbabwe [1]. It is an under-resourced language that is written disjunctively and has the lexical semantics features that poses some challenges in the development of NLP pipeline tools, including POS tagger.

G. Bella and P. Bouquet (Eds.): CONTEXT 2019, LNAI 11939, pp. 62–73, 2019.
https://doi.org/10.1007/978-3-030-34974-5_6

POS tagging is a crucial task in NLP, it assigns appropriate grammatical categories or word classes to every word in a sentence, depending on the usage context. POS tagger for morphologically rich languages such as Setswana poses computational challenge that could affect the effectiveness of the entire NLP system. This is due to some contextual semantics features of the language, which are morphosyntactic in nature that demands a fine-grained granularity level for the required POS tagset, with the need to balance tool semantic robustness with computational parsimony. Due to morphosyntactic ambiguity a word can have multiple lexical tags, based on its definition and usage context.

There is already work that has been done on African languages which are rich in morphology but it is not sufficient, hence the focus of the study is on Setswana which is one of South African language that is spoken by approximately 8.8% of the population [2]. Gauteng (11%), Northern Cape (33.4%), and North West (over 71.5%) are the three South African provinces with the most Setswana speaking people [2]. In Setswana there are words that can be used to express more than one meaning, as a result it becomes challenging to assign them with appropriate lexical cagetories. Therefore the main aim of the study is to analyse some of the Setswana context within a text and their meaning, its lexical ambiguities, and also create a POS tagging model which is one of the NLP tasks.

Since electronic text includes sequences of characters, numbers, punctuation, whitespace, etc. text segmentation is a very crucial step in NLP that identifies the sentence and token boundaries [3]. Text segmentation is the task of dividing text into sentences and linguistic units. Sentence segmentation is the process of identifying sentence boundaries between characters so that it can be used further for processing. It is also referred to as a sentence detection or sentence boundary disambiguation [4]. Sentence segmentation has to be done first before tokenization n because the two tasks can only be done sequentially.

In this study, we present an integrated context-driven corpus-based model for text segmentation, and POS tagging, as a single model, using the Apache OpenNLP for Setswana African language. The purpose is to examine how contextual semantic issues of the language affect the development of the NLP tools, balancing tool semantic robustness with computational parsimony.

The rest of the paper is organized as follows: Sect. 2 reviews related works, Sect. 3 discusses morphosyntatic contextual features of ambiguity in Setswana, Sect. 4 describes the dataset used, Sect. 5 describes the methodology, Sect. 6 presents evaluation and results, and Sect. 7 presents the conclusion.

2 Previous Related Works

There are several existing POS taggers developed for various languages, using various techniques such as supervised, unsupervised, and semi-supervised. In a supervised method, the annotated data is used to train the model first then input sentence is tagged after the training [5]. In the case of unsupervised method, the labeled data is not required, estimation techniques like clustering can be applied. In a semi-supervised method, both labeled and unlabeled data are used in the dataset [6]. However, studies in the context of some Indian languages have shown that different approaches produce varied results for different languages, due to their varied morphosytactic context [7]. Hence, the need to study how the contextual features of a language affect NLP pipeline tools development for the language.

Paul et al. [8] present a Hidden Markov Model (HMM)-based statistical model for Nepali corpora which was collected from (Technology Development for Indian Languages) TDIL. The BIS tagset was used containing 42 tags and the Natural Language Toolkit (NLTK) was considered for implementation. The Viterbi algorithm was used for POS tagging model, giving 96% accuracy for the known words but failed to attain such a high accuracy while the method was used for the unknown words. However, the contextual issues of the languages in relation the effect on POS tagging was not elicited.

Amri et al. [9] presented an approach that combines different taggers in order to exploit their unique properties and reduce some errors that may occur for Amazigh language. The CombiTagger system was used to combine the three taggers which are based on the Conditional Random Fields (CRF), Support Vector Machines (SVM), and Tree-Tagger (TT), and it is language and tagset independent. For training, Cross-validation technique was used to validate the accuracy of all these taggers. These techniques were applied to the Amazigh corpus, then CombiTagger system was later used to improve the accuracy by using the voting algorithm. The overall accuracy of taggers are as follows: CRF with 86.75%, SVM with 86.42%, TT with 87.26%, and the hybrid CombiTagger that combines all the three taggers with the accuracy size of 89.06%, out-performed the three. However, the issue of using context was not addressed.

Sinha et al. [10] proposed a hybrid based approached model using HMM, a statistical approach, and rule-based method. The tagset that was used consists of 43 tags with the Nepali corpus containing 15430 words that were collected from the newspaper and other sources. The model through a hybrid approach has produced an accuracy of 93.15%. However, the use of context in the study was not emphasized.

In the contribution of Freihat et al. [11], they created a single model that integrates the NLP tasks of word segmentation, POS tagging and named entity recognition in the model. The model was designed using the OpenNLP. The annotation was achieved by adopting the three well known approaches like Base annotation for POS tagging, Word segment annotation, and lastly the Name Entity annotation. The POS tagset consists of 58 tags that were classified into five main categories of Noun, adjective verb, adverb, preposition, and Particle. The model was trained on the annotated corpus using the OpenNLP Maximum Entropy POS tagger with default features and cutoff = 3. The corpus used for evaluation was taken from Aljazeera news portal and the Altibbi medical consultancy web portal containing 9990 tokens. The accuracy of each task are as follows: Segmentation 99.7%, Coarse-grained POS 98.7%, and Fine-grained POS 97.9%. The named entity evaluation was done separately with the corpus that contained 674 named entity tags that denote 297 named entitie, resulting in 89.2% precision, 94.6% recall, and F1-Measure = 91.8%. This work has similarity with the present study in the aspect of integrating some NLP pipeline tools in a single model. However, the issue of context use in the development of the NLP tools is not elicited.

3 Contextual Feature of Ambiguity in Setswana

3.1 Ambiguity in Text Segmentation

The main task of sentence segmentation is to disambiguate the correct sentence end mark in a sentence. Tokenization is important for identifying individual tokens and to handle punctuation such as hyphens and full stops in a manner that would allow the other technologies to handle abbreviations and contracted forms correctly. The common ambiguity problems in the context of Setswana are:

- Period mark in an abbreviation of a person's name
- Period mark used as a decimal point
- Period mark after the title of a person
- Period mark in the abbreviation of a foreign name

3.2 Ambiguity in POS Tagging

POS-tagging a Setswana language which is a disjunctively written Bantu language has major issues caused by their morphology and syntax. Setswana uses a disjunctive orthography, where some of the tokens (words) are written separately but they are meaningful as a unit. Therefore words play various functions in a sentence depending on the words surrounding them. This poses a challenge as a word can have more than one meaning. The key challenge how to balance POS tagger semantic robustness with its computational parsimony. Below are some of the examples of ambiguity contextual feature of Setswana which were addressed using fine-grained granularity level of POS tagging:

3.2.1 The word **ba** can be used as

- Possessive Concord - Bana **ba** gagwe (her children)
- Pronoun - **Ba** itumetse (They are happy)
- Demonstratives - Bana **ba** (These children)

3.2.2 The word **re** can be used as:

- Copulative Subject Concord - **Re** mo toropong (We are in town)
- Object Concord - Ba **re** amogetse (They welcomed us).

3.2.3 The word **ka** can be used as

- Particle – **Ka** kwa morago (At the back)
- Tense Marker - Ba **ka** se re amogele (They will not welcome us)

3.2.4 The word **go** can be used as

- Norminal Prefix - go goroga (To arrive)
- Subject Concord - go a tonya/maruru (It is cold)
- Object Concord - Re a go tlhoka (we need you)

3.2.5 The word **bona** can be used as

- Emphatic Pronun - Ba ja dijo tsa bona (They are eating their food).
- Verb – Re simolola go bona ditlamorago (we are beginning to see the consequences).

3.2.6 Bona word can also be separated into two morphemes which are **bo + na**, and these two morphemes are tagged as follows:

- Auxiliary Verb - Kotsi e e tshwanetse ya bo e diragetse mo mosong (This accident could have happened in the morning).
- Copulative Verb - ke na le kakanyo entle (I have a brilliant idea).

Table 1 shows the foregoing examples that elicit some of the lexical ambiguities in Setswana language, where a word can play different roles depending on its usage context defined by the other words that surround it. Some of the words are polysemous having more than one meaning and others are homonymous such as the word **bona** from the examples above which can refer to the word look or them. In Setswana words such as pronouns, concords, and demonstrative are in almost every sentence. For example the word re, as an object concord in a sentence, is written conjunctively to the verb stem, and has noun class information either as a singular or plural form.

Table 1. POS tagging ambiguity

Ambiguity	Example
Possessive Concord vs Pronoun + Demonstratives	Ba
Copulative Subject Concord vs Object Concord	Re
Particle vs Tense Marker	Ka
Norminal Prefix vs Subject Concord vs Object Concord	Go
Emphatic Pronun vs Verb vs Auxiliary Verb vs Copulative Verb	Bona

When dealing with languages, context has become a very important in a process of disambiguating the meanings as well as in understanding the actual meaning of words in areas such as linguistics, lexical semantics and computer linguistics. In a sentence, the meaning of words is determined through the contextual usage.

Setswana is an agglutinative language, where words may contain different morphemes to determine their meanings but all of these morphemes remain unchanged after their combinations. The new words can be generated by adding appropriate suffixes, prefixes, and infixes. For example a verb reka (buy) can be

- reka + ile = rekile (bought)
- reka + ela = rekela (will buy)
- reka + isa = rekisa (sell)

Morpheme is a smallest meaningful unit in a language. In Setswana not all the morphemes have meaning but these morphemes have some grammatical functions. Morphemes can be classified according to their phonological shape such as additive, reduplicative and zero morphemes [12].

3.3 Types of Morphemes

3.3.1 Additive morphems also known as affixation are morphemes that may be divided into prefixation, suffixation, or inifixation.

  - Suffixation morphemes examples roka + ile = rokile (sew)
  - Prefixation morphemes examples nku + di = dinku (sheep - sheep)
  - Suffixation morphemes examples tsamaya + ile = tsamaile (go - went)

3.3.2 Reduplicative morphemes contains a repetition of a stem or root, for example bogalegale (very brave)

3.3.3 Zero morphemes occurs when plural and singular are the same for example madi (blood) and metsi (water)

Adverbs are a diverse group of items in Setswana which differ from other main word-classes such as nouns, adjectives, verbs, etc. Their main role is to provide more information about an action which is performed. In Setswana the adverb category is vague; its features are not clearly defined. There are several semantic types of adverbs in setswana which includes adverbs of place/location, time, manner, etc. [13]:

- The adverbs for place indicates where the verb is taking place e.g. sekolong (at school)
- The adverbs for time indicates the time e.g. motshegare (afternoon/during the day)
- The adverbs for manner indicates the manner of the verb e.g. fela (only)

Semantically, adverbs can be related to questions like kae? (where?), leng? (when?), le mang? (with whom?). Different adverbs can be formed by

- adding locative suffix –ng to nouns e.g. sekolo + ng = sekolong (at school)
- adding noun class prefix to adjective stem e.g. se + ntle = sentle (beautiful)
- using the basic adverbs such as tota (really)

In this study, we present a context-driven corpus-based model that performs text segmentation, and POS tagging for Setswana Natural Language, integrated into a single model.

4 Dataset Description

In this paper, the annotated Setswana corpus is used which contains 65784 tokens, developed by specialist linguistics annotators, based on government domain documents. Defining a semantically robust tagset is very crucial when developing an automatic POS tagger, with appropriate balance struck with computational parsimony of the NLP tool. Bearing this in mind, for this paper, a tagset consisting of 128 tags was designed and used as described in Table 2.

Table 2. Designed Setswana tagset

No	Description	Tags
1	Adjective	ADJ01 -10, ADJ14, ADJLOC
2	Adverb	ADV
3	Demonstrative	CD01-11, CD14-16, CDLOC
4	Norminal prefix	CN15
5	Conjunctive	CONJ
6	Object concord	COPERS
7	Possessive concord	CPOSS01-10, CPOSS14-15, CPOSS17
8	Subject concord	CS01-11, CS14-15, CSINDEF, CSLOC, CSNEUT, CSPERS
9	Enumerative	ENUM
10	Interjection	INT
11	Negative morpheme	MNEG
12	Noun	N01-10, N01a, N02b, N14, N17-18, NPP, NLOC
13	Particle	PART
14	Emphatic pronoun	PROEMP01-10, PROEMP14, PROEMPLOC, PROEMPPERS
15	Posessive pronoun	PROPOSS02, PROPOSSPERS
16	Quantitative pronoun	PROQUANT01-10, PROQUANT14, PROQUANTLOC
17	Question word	QUE
18	Tense marker	TENSE
19	Verb	V
20	Auxiliary verb	VAUX
21	Copulative verb	VCOP
22	Punctuation	ZE, ZPL, ZM, ZPR

These tagset was designed following the EAGLES guidelines, and also, and the guidelines by Taljard and others in the context of African languages, to meet the morphosyntactic requirements of the language [14]. The complimentary use of the latter guidelines is to deal with the trade-off between what is linguistically desirable in the context of a language, and what is computationally feasible.

As shown in Table 2 the noun class has 17 tags, the ones with the uneven class number (1, 3, 5, 7, 9) present the singular nouns, while plural nouns belong to the evenly numbered classes (2, 4, 6, 8, 10). The Adjectives class has 12 tags. The Adverb class has no further specification in terms of the features time and manner is indicated.

Making use of the result of a semantic analysis of the usage context of a word to improve the accuracy of the POS tagging is crucial [14]. In the tagset design, the level of granularity is balanced with computational feasibility, as dictated by the morphosyntactic context of the Setswana language.

5 Methodology

For this study, OpenNLP tool is used for both Setswana text segmentation and POS tagging. OpenNLP is a Java Library for processing natural language with pre-[built models for other languages such as English [15]. In linguistics, the analysis of natural language is crucial in the NLP system, since the words and sentences identified at this stage are the important elements passed to be used in other NLP tasks, such as morphological analyzers, part-of-speech taggers, Named Entity Recognition, and etc. The tasks are broken down into two integrated phases, which are text segmentation and POS tagging.

In computational linguistics, tokenization is the first step in POS tagging, it chops given sentences into smaller units known as tokens. Driven by the tagset context designed, the methodology is structured into phases as follows, using the OpenNLP tools:

Phase 1: Setswana Text Segmentation

- The input text is saved in a text file with the Unicode Transformation Format (UTF)-8 format.
- Then the text divides into sentences by using the Sentence detector which is the pre-trained model of the OpenNLP tool.
- Sentences are then tokenized using the WhitespaceTokenizer which divides words into tokens by using the white space in the OpenNLP tool.

Phase 2: POS tagging

- Create and train the Setswana POS tagger using the OpenNLP POS tagging tool with the Maximum Entropy Algorithm.
- Then evaluate the created POS tagger to provide accuracy.

A. Maximum Entropy Model

Maximum Entropy (ME) that is the underpinning model for the POS tagger, is a conditional probabilistic model, with the ability to learn from the sample data to predict probability distributions. MaxEnt classifier can be used for sentiment analysis, sentence detector analysis, POS tagging, and parsing [16]. Probabilistic models are currently used mainly in NLP because of their high accuracy performances, they enable the automatic building of language independent components, and they make it easier to include linguistically important features [16].

In the sentence detection process, the use of a boundary token is modeled by the joint probability p(b, c) of its actual role and its context, where b represents token boundary and c is the context of the token. Maximum Entropy Principle is the correct distribution of p(b, c) maximizes the entropy (uncertainty) subject to the constraints that represent the known facts:

$$H(p) = -\sum_{(b,c)\in B\times C} p(b, c)logp(b, c) \tag{1}$$

where b is the set of classes (boundary, not boundary) and c is the set of the contexts [17]. The ME technique builds a model which assumes the constraints by defining feature functions. A feature function is a Boolean function that captures several aspects of the language which is relevant to the sequence classification task. Features represent the co-occurrence relation between the predicted class b and the context c, an example feature function for POS tagging is:

$$f_j(b, c) = \begin{cases} 1 \; if \; token \; is \; alphanumeric \\ 0 \qquad otherwise \end{cases} \tag{2}$$

where b is the possible label and c is the context. The models require features to be defined and also how to be used. The relationship between feature functions and labels as shown in the training corpus is expressed as constraints [18]. The constraints on the features have the form of:

$$E_p[f_j] = E_{\tilde{p}}[f_j] \tag{3}$$

where $E_p[f_j]$ is the model expectation of f_j and $E_{\tilde{p}}[f_j]$ is the observed expectation of f_j:

$$E_{\tilde{p}}[f_j] = \sum_{(b,c)\in B\times C} \tilde{p}\,(b, c) f_j(b, c) \tag{4}$$

where $\tilde{p}\,(b, c)$ is the probability of f_j(b, c) estimated on the training corpus.

B. POS Tagging Features

For the Setswana POS tagging task, we identified and used the following three main feature functions in the context of the language:

1. *Context-based features:*
   This feature focuses on the POS tags, which checks the previous tagged token, the current token, and lastly the next token.

2. *Word features:*
   This feature shows the lexical and morphological properties of the word that need to be tagged. By analyzing the words, the categorization of the words into a special character, digits, or even the suffixes. A suffix is a letter or a collection of letters joint at the end of a word to form a new word which can even change the part of speech of the word.
3. *Corpus-based features*
   These features rely on the information that is acquired from the corpus during training.

6 Evaluation and Results

In this section, the performance of both sentence segmentation and POS tagger is discussed. The sentence detection model is trained with the data that has one sentence per line on the pre-built English model creating the new model. Then the new sentence detection model was evaluated using the Sentence Detector Evaluator OpenNLP tool giving the Precision of 99.07%, Recall of 99.53%, and F-Measure of 99.30%.

For the POS tagger, the model performance is shown in Table 3 below. The POS tagging model was trained with the Setswana annotated corpus using the OpenNLP Maximum Entropy POS tagger with default features and cut-off of 3. The data set was divided into two data types, the training data, and the test data, using varied windows as shown in Table 3. For evaluation, OpenNLP POS Tagger Evaluator tool was used to evaluate the performance of the model.

Table 3. Tagging accuracies for Setswana

Exp (%)	Training data (in tokens)	Accuracy
10	6595 tokens	83.69%
20	13175 tokens	87.92%
30	17595 tokens	89.25%
40	26372 tokens	91.28%
50	32895 tokens	92.71%
60	39571 tokens	94.40%
70	52678 tokens	96.08%
80	59213 tokens	96.73%

Table 3 shows different accuracy performance achieved by the model created using different training and test data sizes. Figure 1 demonstrates different experimental accuracy with the different split of training data that have been conducted and also shows that accuracy of the POS tagger increases simultaneously with the increase of training data from 10% till it reaches 80%.

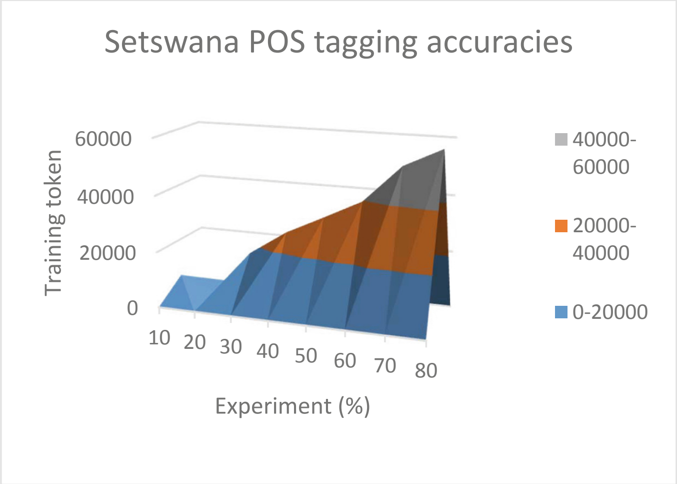

Fig. 1. Setswana POS tagging with different training – test data split

7 Conclusion

In this paper, we have presented a context-driven corpus-based approach for text segmentation and POS Tagging of Setswana language using the OpenNLP tools. To evaluate the accuracy of the proposed POS Tagger, a series of experiments have been conducted using Setswana corpus containing 65 784 tokens, which resulted with the accuracy of about 96.73%. The experiment also shows that the accuracy increases with the increasing of the number of tokens in the training corpus.

The work has allowed us to confirm that defining the adequate dataset and tagset are the core task in building an automatic POS tagger, this however leads us to our future work. In the future work, we intend to develop a tagger that will be able to tag the unlabeled data in order to use it later to develop an annotated corpus with more words and also build a model that shows the similarities and differences between the different taggers. Future works will also further examine the balancing of POS tagger semantic robustness with its computational parsimony.

References

1. Otlogetswe, T.J.: Corpus design for Setswana lexicography. Doctoral dissertation, University of Pretoria (2007)
2. Statssa.gov.za (2019). https://www.statssa.gov.za/publications/03-01-06/03-01-062016.pdf. Accessed 01 Apr 2019
3. Pretorius, R., Berg, A., Pretorius, L., Viljoen, B.: Setswana tokenisation and computational verb morphology: facing the challenge of a disjunctive orthography. In: Proceedings of the First Workshop on Language Technologies for African Languages, pp. 66–73. Association for Computational Linguistics (2009)
4. Palmer, D.D.: Tokenisation and sentence segmentation. In: Dale, R., Somers, H.L., Moisl, H. (eds.) Handbook of Natural Language Processing, pp. 11–35. Marcel Dekker, New York (2000)

5. Kanakaraddi, S.G., Nandyal, S.S.: Survey on parts of speech tagger techniques. In: International Conference on Current Trends towards Converging Technologies (ICCTCT), pp. 1–6. IEEE (2018)
6. Neves, M.: Introduction to Natural Language Processing (2016)
7. Morwal, S., Jahan, N., Chopra, D.: Named entity recognition using hidden Markov model (HMM). Int. J. Nat. Lang. Comput. (IJNLC) **1**(4), 15–23 (2012)
8. Paul, A., Purkayastha, B.S., Sarkar, S.: Hidden Markov model based part of speech tagging for Nepali language. In: 2015 International Symposium on Advanced Computing and Communication (ISACC), pp. 149–156. IEEE (2015)
9. Amri, S., Zenkouar, L., Outahajala, M.: Combination POS taggers on Amazigh texts. In: 3rd International Conference of Cloud Computing Technologies and Applications (CloudTech), pp. 1–6. IEEE (2017)
10. Sinha, P., Veyie, N.M., Purkayastha, B.S.: Enhancing the performance of part of speech tagging of Nepali language through hybrid approach. Int. J. Emerg. Techn. Adv. Eng. **5**(5), 354–359 (2015)
11. Freihat, A.A., Bella, G., Mubarak, H., Giunchiglia, F.: A single-model approach for Arabic segmentation, POS tagging, and named entity recognition. In: 2nd International Conference on Natural Language and Speech Processing (ICNLSP), pp. 1–8. IEEE (2018)
12. Ncube, D.N.: The morpheme in Setswana. Doctoral dissertation, University of Johannesburg (1994)
13. Le Roux, J.C.: A grammatical analysis of the Tswana adverbial. Doctoral dissertation, University of South Africa (2007)
14. Taljard, E., Faaß, G., Heid, U., Prinsloo, D.J.: On the development of a tagset for Northern Sotho with special reference to the issue of standardisation. Lit. J. Lit. Crit. Comp. Linguist. Literary Stud. **29**(1), 111–137 (2008)
15. Opennlp.apache.org, Apache OpenNLP Developer Documentation. Opennlp.apache.org/docs/1.9.1/manual/opennlp.html. Accessed 23 Apr 2019
16. Manning, C.: Maxent models and discriminative estimation. CS 224N lecture notes, Spring (2005)
17. Dii.unisi.it (2019). http://www.dii.unisi.it/~maggini/Teaching/TEL/slides/09%20-%20OpenNLP.pdf. Accessed 30 May 2019
18. Dalal, A., Nagaraj, K., Sawant, U., Shelke, S.: Hindi part-of-speech tagging and chunking: a maximum entropy approach. In: Proceeding of the NLPAI Machine Learning Competition (2006)

The Truth Rule: Definitional or Essential?

Maryam Ebrahimi Dinani(✉)

Institut Jean Nicod, PSL, Paris, France
maryam.ebrahimidinani@gmail.com

Abstract. This paper is about the norm of truth for assertion, which I henceforth call "The Truth Rule", and is formulated as follows: "One ought to assert only what is true". I argue that The Truth Rule as thus formulated is a norm for assertion in a specific sense. I defend the view that assertion is, *by its nature*, governed by the rule according to which one ought to assert only what is true.

Keywords: Assertion · Truth · Constitutive rules

1 Introduction

This paper is about the norm of truth for assertion, which I henceforth call "The Truth Rule", and is formulated as follows: "One ought to assert only what is true". Assertion as a speech act has many norms if we take 'norm' in the weak sense of any feature of the performances of an act which involves a dimension of evaluation and of appraisal. In this broad sense, there will be many norms for assertion, such as politeness, relevance, and prudence. It is uncontroversial in this sense of 'norm' that truth is a norm for assertion. What is controversial, however, is what kind of norm it is –whether it is a norm in a stronger sense– and what is its place with respect to the concept of assertion.

In this paper, I argue that truth is a norm for assertion in a specific sense. I defend the view that assertion is in an *essential* way – *by its nature* – governed by the rule according to which one ought to assert only what is true. The paper proceeds in two parts: I start the first part (Sect. 2) with M. Dummett, who argued for the importance of the norm of truth for assertion through an analogy with games. I then explain how this analogy is formulated in the literature in terms of what is known as "constitutive rules", and I discuss two distinct, but conflated, ways of characterizing constitutive rules from the works of T. Williamson and J. Searle. In the second part (Sect. 3), I introduce a distinction between two types of constitutive rules, which I call "definitional" vs. "essential" constitutive rules. I argue then for The Truth Rule as being of the essential type, and I give a possible explanation of the distinction within the institutional framework. I finish by looking at one possible account of the nature of the essential type of constitutive rules, through A. Reinach's account of social acts, and I argue that it could offer a promising insight about the essential aspect of that type of rules.

G. Bella and P. Bouquet (Eds.): CONTEXT 2019, LNAI 11939, pp. 74–83, 2019.
https://doi.org/10.1007/978-3-030-34974-5_7

2 Games, Speech Acts, and Constitutive Rules

2.1 Dummett and the Analogy with Games

In his 1959 paper, "Truth", M. Dummett argues that, in order to give a proper account of a concept, it is not sufficient to describe those circumstances in which we do, and those in which we do not, use the relevant word, "by describing the *usage* of that word; we must also give an account of the *point* of the concept, explain what we use the word *for*" (Dummett 1959: 143).

Applying this principle to the concept of truth, Dummett explains what the *point* of the concept of truth is, by making an analogy between the concepts of truth and falsity, and the concepts of winning and losing of a board game. In a game like chess, we could give a "formal description" (Dummett 1973: 296) of the rules of chess, by specifying the initial positions of the pieces and the permissible moves; as such, the game ends when there is no permissible move, where we distinguish between two kinds of final positions, "win" and "lose". This formal description of chess, by itself, does not provide us with a "theory of chess as an activity" (Dummett 1973: 296). It leaves out one crucial point: that it is the objective of a player to win. If a person has no previous understanding of the word 'win', the formal description of chess is not sufficient for her to understand what it is to play chess; she may be playing a *variant* of chess in which the formal description coincides with that of chess, but in which she is trying to be checkmated by the opponent rather than to checkmate her[1]. Hence, what should be added to the formal description of chess in order for this person to be said to be playing "chess" is the intention of arriving at a particular one of the end positions, the winning one.

At this point, one may object that, in the larger context of chess tournaments, some players adopt a strategy of playing to draw as black, since tournament statistics provide evidence that white is more likely to win than black but that there is a good probability of drawing. Similar points might apply to other games and their tournaments. One might then think that the principle that we aim at winning in competitive games is context dependent, insofar as the strategy of playing to draw as black in chess tournaments does not violate chess norms, or so the objection goes.[2]

In response, let us start by a point made by Dummett about games with stakes – where winning and losing are associated with certain definite consequences (ex. money, world cup, etc.). Dummett argues that in these cases it is possible to judge whether one is in fact playing a game by reference to those associated consequences, since in these cases the imposition of the penalty or the getting of the prize serves to distinguish winning from losing. According to Dummett, in such cases, we are not playing a game "for its own sake", this latter feature being characteristic of what he describes as games with no stake, where "we have no means of explaining the fact that checkmating your opponent constitutes winning otherwise than by saying that people have the custom of making

[1] In *Frege: Philosophy of Language*, in the chapter on "Assertion", Dummett argues, "there is no such thing as a game in which the object is to try to lose". We would better not say that two people are playing chess but that they have both agreed to try to lose, but we shall say that they are trying to win but that what they are playing is not chess, but a variant of it, in which players try to be checkmated rather than to checkmate the opponent's king (Dummett 1973: 320).

[2] Thanks to the anonymous referee for bringing out this objection.

these moves while trying to arrive at such an end-position" (Dummett 1973, p. 301). Indeed, in the case of games with stakes, there is no need to identify that particular game in terms of the players' aim of arriving at the winning position; even if the aim of the players may not be winning in those particular games, they can still be identified to be playing that game because of the consequences associated with the game; they would be playing it even if they did not aim at winning since the assignment of the associated consequences at the end of the game would make clear what game was being played.

I would like to deploy Dummett's remark – somehow modifying it[3] – and to suggest that when we are playing at a chess tournament, we are not playing amateur chess for its own sake; rather, we are playing highly institutionalized tournament chess – with the aim of getting the prize – so that if a player in a tournament chess does not try to "win" a particular match in the sense of checkmating the opponent's king, he can still be identified to be playing the tournament chess, because winning has become associated with getting the prize in the whole context of the tournament, and it is this aim of achieving the prize which guides the players' conduct in this context. Returning to the aforementioned objection, what I take as a reply is that the point of playing competitive games is aiming at winning. Now, in the case of the amateur game of chess, the point of the game is to checkmate the opponent's king. In the case of tournament games of chess, on the other hand, we are not playing amateur chess, and particular games in the tournament are played with a view to getting the prize. But nothing in all this implies that the aim of all those competitive games is not winning; the aim of the game is still winning, except that winning has gained another determination when we have to do with tournament games. Therefore, the idea that the point of competitive games is winning is not context-dependent.

Now, to go back to the analogy between the concepts of truth/falsity and the concepts of winning/losing of a board game, a person will not understand the concept of truth if all he has in hand is a method for determining the correct application of the terms to statements. If we want to give an adequate theory of truth, we should not presuppose that it is already known what it means to call a class of sentences "true" and the other class "false": we have to explain what this prior understanding consists of, *i.e.* what is the point of classifying sentences into true and false ones. According to Dummett, this principle is that according to which we aim at making true statements in our assertoric practices.

> It is part of the concept of winning a game that a player plays to win [...] Likewise, it is part of the concept of truth that we aim at making true statements. (Dummett 1959: 142-3)

In the Dummettian perspective, since it is part of the concept of truth that we aim at making true statements in our assertoric practices, an adequate theory of truth should explain this *conceptual* feature of truth. For Dummett, assertion is an activity in which

[3] Dummett's remark is only about the idea that our means of identifying what game is being played is through the players' intention of arriving at the winning position, but that this latter is not required in those cases where we have associated consequences, insofar as these consequences are in force at the end of the game and will make known which game was being played, whether a player played with the intention of winning or not.

"the utterance of a sentence, except in certain circumstances, is understood as being carried out with the intention of uttering a true sentence" (Dummett 1973: 302). This principle has become widely known as the Dummettian slogan "we aim at making true statements" (Dummett 1959: 142).

The Dummettian analogy is used by many authors in the debate surrounding the rules of speech acts and, in particular, assertion. T. Williamson and J. Searle both undertake the analogy between assertion and competitive games in order to argue for what is known as "a constitutive rule" for a speech act. In what follows, I explain how constitutive rules are to be conceived of in the debate surrounding the rules of speech acts, by reference to this analogy.

2.2 Williamson, Games, and Constitutive Rules

In *Knowledge and Its Limits*, Williamson appeals to the analogy between speech acts and competitive games, in order to argue that assertion is governed by a constitutive rule. He defines a constitutive rule as a rule which is *essential* to an act, such that "necessarily, the rule governs every performance of the act" (Williamson 2000: 239) and is due to its "specific nature" (Williamson 2000: 240). Games are typical examples through which the idea of constitutive rules is introduced. According to Williamson, games like chess are constituted by rules, which fix the permissible moves and justify praise or criticism with respect to the performance of those moves. Assertion is analogous to games, inasmuch as it falls under a constitutive rule, a rule which is essential to the speech act of assertion.

Williamson favors the knowledge rule –and not truth rule–, according to which one ought to assert only what one knows, but the point of the analogy between games and the speech act of assertion lays in the principle that assertion is governed by a constitutive rule which is "individuating", in the sense that necessarily, assertion is the unique speech act whose unique constitutive rule is a certain C rule (Williamson 2000: 241); all other norms for assertion result from the conjunction of the C rule and considerations which are not specific to assertion.

2.3 Searle, Games, and Constitutive Rules

In his *Speech Acts*, Searle distinguishes between what he calls "regulative" and "constitutive" rules.

> Regulative rules regulate a pre-existing activity, an activity whose existence is logically independent of the rules. Constitutive rules constitute (and also regulate) an activity the existence of which is logically dependent on the rules. (Searle 1969: 34)

To clarify the distinction, he takes the example of rules of etiquette as rules which regulate social relationships which exist independently of the rule – such as rules of eating with forks and knives on which the activity of eating is not logically dependent –, and rules of chess – such as castling – which not only regulate playing chess, but also define in what the practice of playing chess consists. They have a distinctive feature compared to merely-regulative rules, which is that they are almost "tautological in character", in the sense that the rules of castling, for example, are *definitional* of what castling is: since castling is nothing more than a move in chess, the rules of castling can appear "now as a rule, now as an analytic truth based on the meaning of 'castling in chess'" (Searle 1969: 34).

At this point, the question arises whether the sense of 'constitutive' as 'essential' –used by Williamson– differs from its sense as 'definitional'–used by Searle. In the second part of this paper, I argue that it does.

3 Two Types of Constitutive Rules

3.1 Definitional vs. Essential Rules

There is a difference in Williamson's conception of constitutive rules with respect to the Searlian conception, apparently based on the concept of violability of the rules. Williamson holds that obeying a constitutive rule is not a necessary condition for performing the rule-constituted act. According to Williamson, breaking a constitutive rule of a game does not amount to ceasing to play that game. If a person breaks a rule of castling, we just say that he has "cheated" (Williamson 2000: 238). Let us call the sense of 'constitutive' as used by Williamson "*essential*". This sense does not correspond exactly to the Searlian sense of 'constitutive' that we saw as "*definitional*" or "*analytic*". The difference lies in the fact that if a constitutive rule is definitional, as is in the Searlian sense, we do not engage in the act of which the rule is definitional if we do not act in accordance with the rule. In Williamson's framework, however, constitutive rules of speech acts, as well as constitutive rules of games, are not definitional in that sense; they are individuative and essential, insofar as they are unique rules of the practices –of which they are constitutive– that necessarily govern all performances of those practices. In his paper on assertion, J. MacFarlane also, alongside Williamson, holds that we have to distinguish between obeying a rule and being subject to a rule. For MacFarlane, if a rule is constitutive, we are subject to it, but this does not mean that we cannot disobey it: "A move may be subject to a rule either by obeying it or by being in violation of it." (MacFarlane 2011: 84, footnote 6) A constitutive rule, from this perspective, necessarily governs all performances of the act of which it is a constitutive rule, without it being a necessary condition for the *performance* of that act that we abide by the rule.

I want to argue that this debate about two ways of conceiving the notion of constitutive rules points towards a distinction within the category of constitutive rules. The rules of games as we know them are definitional rules: contrary to what Williamson and MacFarlane suggest, an act like castling in chess cannot be accomplished unless the rules that are constitutive of it are respected.[4] But assertion is governed by essential rules, a rule that the act presupposes even in particular cases where the rule is not

[4] The question arises at this point as to the possibility of cheating in a game. It seems to me that, with respect to the acts *within* practices with definitional constitutive rules, cheating amounts to getting out of the *act* of which the constitutive rule is definitional –and not out of the practice as a whole. If one 'castles' while the squares between the king and the rook involved in castling are occupied, one cannot be said to have castled, while one can be said to have cheated with respect to the game of chess as a whole.

respected. Indeed, MacFarlane is right to point out that we have to distinguish between being subject to a rule and obeying it, but he is wrong to conclude that, for *all* constitutive rules, a move may be subject to such a rule either by obeying it or by being in violation of it. Therefore, I shall introduce a distinction within the category of constitutive rules: a rule of castling is a *definitional* constitutive rule, in the sense that if the squares between the king and the rook involved in castling are occupied, the player cannot be said to be castling. On the other hand, the truth rule for assertion is an *essential* constitutive rule, insofar as one who lies breaks a constitutive rule of assertion, without thereby ceasing to make an assertion.

3.2 Constitutive Rules in the Institutional Framework

There is one way we could trace the distinction between essential and definitional rules based on the institutional framework. Searle himself grounds his theory of constitutive rules on such a framework, so that looking at this framework will also help us see where Searle fails to make the distinction.

Institutions are systems of constitutive rules that structure social interactions (Tummolini and Castelfranchi 2006: 1). Institutional facts are facts that presuppose certain institutional settings and brute facts are those which exist independently of human institutions. In other words, institutional facts are facts whose existence, unlike the existence of brute facts, presupposes the existence of certain human institutions (Searle 1969: 51). E. Anscombe takes the example of sending a piece of paper as the example of a brute fact while that of sending a bill as an institutional one. The reason is that "the institution of buying and selling is presupposed to the description 'sending a bill'" (Anscombe 1957: 72). Given that institutional facts presuppose institutions, they can only be explained in terms of the network of constitutive rules which underlie them. According to Searle, since every constitutive rule can be formulated in terms of "X counts as Y in context C" (Searle 1969: 36) –where C is the relevant context in which a rule takes its significance–, every institutional fact is underlain by a system of rules of this form. Let us take two examples of constitutive rules of chess.

(1a) In a game of chess, the king being attacked and there being no move to go out of check counts as *checkmate*.
(1b) In a game of chess, checkmating the opponent's king counts as *winning*.

Let us now take another pair of examples from Searle, concerning the speech act of promising.

(2a) In the speech act of promising, uttering "I hereby promise…" under certain conditions counts as *promising*.
(2b) In the speech act of promising, promising counts as the *undertaking of an obligation* to do some future act.

If we look at the two pairs of examples, we can identify a difference between the terms in italic in (1a) and (2a) on the one hand, and in (1b) and (2b) on the other. Whereas checkmating and promising seem to be concepts entirely defined from within a certain institution, the concepts of winning and of obligation seem to be logically independent of the institutions of chess and of promising. To take the example of obligation, first of all, it is not contained in the concept of obligation that a promisor has an obligation to keep his word; moreover, there are many other sources of obligation, such as the obligation generated by the act of borrowing something. Thus, concerning the institutional framework, I suggest a distinction between two kinds of concepts: intra-institutional concepts and trans-institutional concepts.[5] Given this framework, intra-institutional concepts are those that are entirely defined or that exist only in virtue of a rule within a certain institution. Trans-institutional concepts, on the other hand, are somehow floating concepts used in different institutions; to use D. Miller's terms, they exist "above and beyond" (Miller 1981: 188) the institution in the formulation of the rules of which they are used, and "do not derive from or are not constituted by rules of institution" (Miller 1981: 191). In other words, they do not take their meaning from within a particular institution, and the relation between the trans-institutional concept and the constitutive rule in the formulation of which the trans-institutional term is used is not analytic. Therefore, we have a choice to break the rule relating the institution to that trans-institutional concept which goes beyond the institution and which determines the significance and purpose of the whole institution, whereas if we stay intra-institutional, we do not have a choice but to be in accordance with the rule.

It seems then that there is a parallel between intra- vs. trans-institutional concepts and the definitional vs. essential constitutive rules. Consider three examples of essential constitutive rules –(1b), The Truth Rule, and (2b)– relating three pairs of concepts: competitive game of chess and winning, assertion and truth, and promising and obligation. The second element of each pair is what we have characterized as a trans-institutional concept. It seems that each time we have an essential rule, there is a trans-institutional concept used, which elevates the rule beyond the institution and which leaves us with the choice to disobey. In other words, essential constitutive rules involve at least one trans-institutional concept, whereas definitional constitutive rules do not involve any trans-institutional concept and typically involve intra-institutional concepts.

I have already characterized essential constitutive rules as not having a definitional or analytic tie to the concepts of competitive games, of assertion, and of promising –contrary to definitional rules–, since it is not self-contradictory to play a game with another aim than winning, neither contradictory to lie, nor to break one's promises. It now seems clear that essential rules do not have a definitional or analytic tie neither to the concepts of winning, of truth, and of obligation. It is not contained in these three concepts that we ought to abide by the essential rules which relate them to the corresponding concepts of competitive games, of assertion and of promising. Winning, truth, and obligation are logically independent of those rules, as they can be generated by other sources than competitive games, assertion, and obligation, respectively.

[5] For this distinction, I am indebted to a paper by D. Miller, who introduces a similar distinction between intra- vs. meta-institutional concepts. [See Miller (1981)] To go further, see Carnap (1950).

3.3 Essential Rules and Their *Essential* Aspect

I suggested that although The Truth Rule is neither definitionally tied to the concept of truth, nor to that of assertion, the latter is, by its nature and in an *essential* way, subject to this rule and governed by it, and this rule is presupposed in our assertoric practices. But what is the specificity of this type of constitutive rules, which makes them "essential" to the practices they govern and "presupposed" by them?

There are at least two ways we could give an explanation of this essential aspect. One is a naturalistic account *à la* Ruth Millikan, which I will skip here. The idea very roughly is that if the rule was not presupposed by the practice, the practice could not have existed; in particular, if it was not presupposed by our assertoric practices that one ought to assert only the truths, then communication would not have been possible in the first place.[6] But there is another possible explanation that I want to briefly refer to here: a realist explanation due to A. Reinach.

The realist phenomenologist Reinach has a concept of *social acts*, which is very similar to the concept of speech acts developed by Austin and Searle, and has as examples questioning, commanding, promising, etc. He analyzes the concept of promising in detail, and I will thus base my explanation on this case.

Reinach holds that claims and obligations are legal entities which constitute an ontological category of temporal objects of a special kind that exist independently of any human institution. These entities have two characteristics: First, they can only be generated through human acts, *i.e.* through social acts, since there should always be some "cause" for them (Reinach 1983: 15). According to Reinach, promising is one source of claims and obligations. Secondly, these legal entities and social acts are governed and related to each other by "essential laws". These are, in his terminology, universal laws which are in force in all performances of the social act in question, but which are not analytic, in the sense that, firstly, nothing is contained, for example, in the concept of obligation about the fact that an obligation dissolves as soon as the thing promised is done, and secondly, the contradictory of this statement, while being false, would not be a logical contradiction.

> We have found in promising an act all its own, and we claim that it lies in the *essence* of this act to bring forth claims and obligations. (Reinach 1983: 26. Emphasis added)

For Reinach, "the obligation is grounded in the nature of promising as an act" (Reinach 1983: 45), so that the act of promising *as such* produces obligation.

My purpose here is not to take side with Reinach concerning his realist account of legal entities; rather, what is important for my present purpose is to mention two interrelated characteristics of essential laws that Reinach explores and which seem helpful. The first one is their immediate intelligibility. Reinach says about the essential laws that they are self-evident, in the sense that "the intuitive grasping of them in an immediate insight can be achieved again and again, as soon as the knowing subject directs its attention to them" (Reinach 1983: 131). From this perspective, our knowledge of promising is different from our knowledge of constitutive rules of a game: as J. Crosby argues, we

[6] See Millikan (1984).

“pre-eminently understand something” when we are told that the obligation dissolves as soon as the thing promised is done, “but there is no trace of such understanding when we learn that the pawn is the only piece in chess which can move only by going forward and never backward” (Crosby 1990: 82). This takes us to a second characteristic of the essential laws in the Reinachian perspective, which is that “it makes simply no sense to speak of really forgetting them” (Reinach 1983: 131). Considering the nature of the act of promising is sufficient to rediscover the basic relations between promising, claim and obligation. As J. DuBois argues, “[t]his contrasts sharply with the non-intuitiveness of the basic rules of chess: the connection between a rook and its movement is simply stipulated, can easily be forgotten, and cannot be rediscovered simply by considering the form of the piece” (DuBois 2002: 338-9).

To sum up, the Reinachian conception of promising and of obligation puts forward a possible explanation of the nature of the relation which holds between them, *i.e.* an explanation of the essential aspect of what I called the “essential rule” of promising, according to which one ought to keep one’s promises. The intelligibility and non-forgettability of Reinachian essential laws seem to me to be illuminating, concerning the difference between the essential aspect vs. the stipulative aspect of the two types of constitutive rules.

To finish, let us take into account a possible objection about the modifiability of these essential rules with respect to the context of use. According to this line of thought, requirements to assert falsehoods need to be limited if language is to be viable but that does not mean they never modify the default norm of assertion; positive law and institutional rules have been known to require the assertion of falsehoods. For example, avoidance of scandal and disorder can easily lead a politician or an administrator to introduce legal or institutional requirements to assert falsehoods. The objection then consists in the question of whether such requirements cannot be said to modify the general norm of assertion and to render it context-dependent. (see Footnote 2) The same point may apply to the case of promising insofar as the essential rule that a claim is always and without exception generated by a promise can be modified in various contexts of civil law where we are faced with a factual impossibility of such a generation.

In response, we have to acknowledge that there may be institutional requirements to assert falsehoods, and the same point may apply to the essential rule of promising. Reinach’s reply to this objection would be that it presupposes that which it calls into question. Indeed, in these cases, the Truth Rule or the essential rule of promising would become inefficacious precisely *as a result of* an institutional requirement, or an obligation becomes efficacious *as a result of* such a requirement, and therefore the objection “rests on those very essential rules governing social acts the validity of which it calls into question”. (Reinach 1983: 116) My suggestion is that the institutional requirements or enactments may require modification of the efficaciousness of the essential rules only in highly institutionalized frameworks, but those enactments should be better seen as being themselves stipulative rules defined and operative only within those institutionalized forms of social acts, presupposing the violability of the essential rules without questioning their essential aspect underlying the un-institutionalized social acts types.

4 Conclusion

I argued, in this paper, that The Truth Rule is neither analytically tied to the concept of assertion, nor to that of truth. Nevertheless, it is a constitutive rule of the essential type: it is constitutive of assertion in the sense that it is presupposed by the speech act of assertion, and that the latter is governed, in an essential way, by The Truth Rule. However, The Truth Rule is not constitutive of truth in any interesting sense, so that an adequate theory of truth need not account for this principle. This bears against the normativist accounts of truth, which regard the Truth Rule as a conceptual aspect of truth to be accounted for in an adequate theory of truth. It seems to me that the explanation of The Truth Rule is, rather, part of an adequate theory of assertion, and that a Reinachian account of social acts could offer a promising explanation of the essential aspect of this rule. Of course, the whole enterprise requires that I show why other normativist accounts of assertion are less preferable to the one which takes The Truth Rule to be the essential constitutive rule of assertion – rather than the knowledge rule, or the reasonable-to-believe rule, etc. –, but that is a project for another paper.

References

Anscombe, G.E.M.: On brute facts. Analysis **18**(3), 69–72 (1957)

Carnap, R.: Empiricism, semantics and ontology. Revue Internationale de Philosophie **4**(11), 20–40 (1950)

Crosby, J.: Speech act theory and phenomenology. In: Burkhardt, A. (ed.) Speech Acts, Meaning, and Intentions, pp. 62–88. The Gruyter, Berlin (1990)

DuBois, J.M.: Adolf Reinach: metaethics and the philosophy of law. In: Drummond, J.J., Embree, L. (eds.) Phenomenological Approaches to Moral Philosophy: A Handbook. CTPH, pp. 327–346. Springer, Dordrecht (2002). https://doi.org/10.1007/978-94-015-9924-5_17

Dummett, M.: Truth. Proc. Aristot. Soc. **59**(1), 141–162 (1959)

Dummett, M.: Frege, Philosophy of Language. Harper & Row, New York (1973)

MacFarlane, J.: What is assertion? In: Brown, J., Cappelen, H. (eds.) Assertion: New Philosophical Essays, pp. 80–97. Oxford University Press, New York (2011)

Miller, D.: Constitutive rules and essential rules. Philos. Stud. **39**(2), 183–197 (1981)

Millikan, R.G.: Language, Thought, and Other Biological Categories. The MIT Press, Cambridge (1984)

Reinach, A.: The apriori foundations of the civil law. In: Crosby, J.F. (tr.) Irving Texas. International Academy of Philosophy Press (1983)

Searle, J.: Speech Acts: An Essay in the Philosophy of Language. Cambridge University Press, London (1969)

Tummolini, L., Castelfranchi, C.: The cognitive and behavioral mediation of institutions. Cogn. Syst. Res. **7**(2–3), 307–323 (2006)

Williamson, T.: Knowledge and Its Limits. Oxford University Press, New York (2000)

Supporting Privacy Control and Personalized Data Usage Explanations in a Context-Based Adaptive Collaboration Environment

Mandy Goram(✉) and Dirk Veiel

Faculty of Mathematics and Computer Science, FernUniversität in Hagen, 58084 Hagen, Germany
{mandy.goram,dirk.veiel}@fernuni-hagen.de

Abstract. The General Data Protection Regulation, e.g., provides the "right of access by the data subject" and demands explanations of data usages, i.e. explanations where and for what purpose personal data is being processed. Supporting this kind of privacy control and related personalized explanations of data usage in context-based adaptive collaboration environments are big challenges. Currently, users cannot retrace the usage and the storage of their personal data in context-based adaptive collaboration environments. We address the aforementioned challenges by developing a context-based adaptive collaboration platform, the CONTact platform, that can be linked to or integrated into different kinds of collaboration environments (e.g., meinDorf55+, a novel community support system for elderly). The CONTact platform supports users with privacy control and personalized explanations of data usages. In this paper we present an excerpt of our extended domain model and two sample situations when privacy control and personalized explanations get relevant. We use a sample ontology that is based on our domain model to illustrate the related processes and rules. Using our approach users can control their data usage and are able to get personalized explanations of their data usage in a context-based adaptive collaboration environment. This helps us observing legal regulations, e.g. privacy laws like the GDPR.

Keywords: Context-based · Adaptive · Collaboration environment · Privacy control · Personalized explanations · Legal regulations · GDPR

1 Introduction

Considering legal regulations has become an important aspect of software development. The General Data Protection Regulation (GDPR) demands comprehensibility of personal data processing and provides the "right of access by the data subject"[1]. Due to that software providers must be able to reveal what data is stored and processed by their applications and services. The ongoing trend to personalize content and applications requires the development of more sophisticated approaches. These should take the current situation of their users into account and provide adequate support.

[1] https://gdpr-info.eu/art-15-gdpr/.

G. Bella and P. Bouquet (Eds.): CONTEXT 2019, LNAI 11939, pp. 84–97, 2019.
https://doi.org/10.1007/978-3-030-34974-5_8

Context-aware systems are able to support personalization with regard to the current situation of related users. To support users in certain situations (e.g. create documents in a collaborative work environment) the system must be aware of the user's situation and the related socio-technical environment, i.e. the context. Dey [1, P. 5] defines that "Context is any information that can be used to characterize the situation of an entity". Considering the user's context, a system becomes context-aware, as soon as it "uses context to provide relevant information and/or services to the user, where relevancy depends on the user's task" [1, P. 5]. The disadvantage of context-aware systems is the predefined and fixed context to a certain domain and some few situations [2]. Because it is not possible to predefine all possible situations of the users and their interactions (i.e. at design time), the systems are limited in supporting users. [3] present an approach that uses a formal context model representing the socio-technical system. This enables modelling of and acting on different kinds of interactions and situations of a specific domain even after the final shipment of the underlying application(s). They support this kind of opportunity by adding an extra level of abstraction (i.e. a formal context model) and by separating the application from the so-called Adaptation Runtime Environment. Such a formal context model is part of a context-based system. It describes the relationships between objects which are relevant and significant for the current situation.

These kinds of systems are very complex and need extensive information about the situation including the user which shall be supported. The "Significant complexity issues challenge designers of context-aware systems with privacy control" [4, P. 59]. From the legal perspective of the GDPR the user must be able to restrict or even decline the data usage. From a user's perspective he or she wants to decide who will be able to access personal information and when it will be shared or processed. Due to that the privacy control is strongly related to intelligible explanations. That demands a way to explain system processes and data usage, to help users to understand the current situation [5]. "The dynamic aspect of context implies that it is not possible to plan in advance the whole explanatory dialogue" [5, P. 123]. Our understanding of personalized explanations is that they "serve to clarify and make something understandable" [6, P. 498] to the user in a specific situation like the relevance of the GDPR and its consequences to the usage of the system.

According to [7], supporting user friendly intelligible and comprehensive explanations in context-based adaptive systems is a big challenge. They are important for a personalized system to support user acceptance and user trust [6]. Additionally, legal regulations make it necessary to explain data storage and data processing of personal information in a system.

We use a scenario to illustrate the above requirements. Alice uses a context-based adaptive collaboration environment which uses and stores personal information about her in order to, e.g., support personalization. So, we have to answer two questions:

1. Q1: How can Alice agree that her personal information can be stored in and processed by the system and its associated functionalities and services?
2. Q2: How can the "right of access by the data subject" (see Footnote 1) be realized for Alice?

Q1 also includes that every change affecting her agreement, i.e. changing data storage location or usage in the system, invalidates it. Therefore, Alice must accept the changes and agree upon it once again. Q2 implies that Q1 has to be answered as well.

Currently, users cannot retrace the usage and the storage of their personal data in context-based adaptive collaboration environments. According to [5] it is not possible to place explanations to every situation in the system. Furthermore, it is not possible to agree to the usage of personal data for individual functions and applications. In the case of rejection, the entire system can no longer be used.

We address the aforementioned problems and challenges by developing a context-based adaptive collaboration environment supporting user control, comprehensibility and intelligibility. In this paper we present an approach

(1) to give privacy control back to the users according to Q1, and
(2) to create personalized explanations of data usages according to Q2

in our context-based adaptive collaboration environment, based on the CONTact platform (c.f. [3]). We use two typical scenarios ("Compliance by Design" and data usage explanation, cf. Sect. 4) to illustrate our approach consisting of (1) an extended domain model for legal regulations, (2) two process models, and (3) two related rules.

The paper is structures as follows: in Sect. 2 we present related work. We illustrate our extended domain model for legal regulations in Sect. 3, before we use the above scenario to present our answers to Q1 and Q2 in Sect. 4. We discuss our results in Sect. 5. Finally, we present some conclusions and future work in Sect. 6.

2 Related Work

Due to the development of mobile devices and applications as well as the development of personal recommendation systems and intelligent assistants [8], which support users at work or in private areas, many sensitive and personal data sets need to be saved, analyzed and processed. Since many years, researchers realized that the intensive use of sensitive and personal data is a challenge for data privacy. Privacy protection especially concerns the development of personalized application, e.g. collaborative environments, intelligent tutoring systems, (embedded) recommender systems, intelligent assistant systems and mobile assistants in smart devices, cars and even smart cities [9].

So far, research has raised questions concerning the data usage and data processing in systems and techniques mostly from the ethical-moral perspective [10] or from the perspective of supporting user trust [10, 11]. By the GDPR data collection and data usage must be considered also due to the legal necessity [10, 11]. This already applies to the planning and design of a system which is intended to process personal data.

Scientists who work on the design of personalized, adaptive environments focus on the mapping of user and domain-specific aspects. Some of them consider context information to support the users in certain situations. One promising technology on modelling context is ontology [2]. An ontology is a formal specification of a certain domain which describe a set of concepts, relationships and formal axioms that restrict the interpretation of concept instances [12]. The formal concepts can become a common ground to describe a specific domain which can be shared and reused. Most of the concepts do not consider privacy control or intelligibility of the personal data usage what became so important through the GDPR.

[13] present an approach to support the intelligibility of complex context-aware systems. They point out that intelligibility must be accompanied by a control function for the user. In their work they present an extension of the Context Toolkit. "The Context Toolkit aims at facilitating the development and deployment of context-aware applications."[2] With a programming abstraction they support developers and designers to create explanations to support intelligibility and user control in context-aware applications build by the Context Toolkit. For that, they integrate meaningful explanations in the application Situation by exposing the internal processing of context-aware applications.

Enhancements to the explanation component in the Context Toolkit can generate explanations of the behavior of more popular machine learning techniques and enriched explanations for user control [14, 15]. According to [13] and [14], we consider user control and explanations about the context and the internal processing in our context-based collaboration environment with focus on integration and explanation of external policies. The Context Toolkit and its extensions [13–15] does not reveal any relation to data privacy compliant declarations of data usage and also does not provide information on whether context-based collaborative environments are supported.

Supporting privacy control in context-aware systems is the approach of [4]. They present annotations in information spaces to classify personal and sensitive information. The privacy tagging is used to mark privacy related information that can be identified during processing. The access of a user defined information space is used as a contextual trigger to ask for permission of the owner. The approach support users to get back control on their personal information.

Similar to our approach is the work of [16]. His approach considers the user privacy preferences in context-aware webservices. Therefore, he introduces the policy language Consumer Privacy Language (CPL). The CPL is used to specify the user's privacy preferences, who can insert their privacy setting through a web application. These preferences are considered during the webservice invocation. An adaptation mechanism uses the privacy preferences to get access to context information on a per case basis. The mechanism is integrated in the webservice infrastructure that applies the user's privacy preferences and manages the service execution. [17] extended the privacy module of the Linked Unified Service Description Language (USDL). The privacy module is used to describe privacy policies for the use of any webservice. For that they focus on the service provider and how the provider can communicate the policies considering a service. By using Linked Data they provide the opportunity to link policies and place them in context. The extension can use and include existing privacy policies to answer questions about what personal data is collected from the user, what the service provider does with the collected data and to whom it will shared. The approaches of [16, 17] focus on supporting privacy of user while using webservices. An interesting aspect is the separation of private and non-private data on the conceptual layer. Neither [16] nor [17] describe if and how to support an integrated collaborative environment and so they do not consider the requirements of a personalized collaboration environment. They also do not present

[2] http://contexttoolkit.sourceforge.net.

how a user can get access to a personalized explanation of stored and processed data in the system. In [17] the authors do not describe how users can accept or decline the data usage for certain applications or services and what consequences are related to it. Our approach considers, that users can make decisions about the data processing (accept or decline). For that, we integrate external policies, which are important for the situation, in the context and analyze which of the policies must applied in the specific situation.

Privacy and privacy control come along with intelligible explanations. Explanations are needed to help users to understand why and how their data is used in the system and to whom it will be accessible [18].

[19] present a generic four-layer framework for modelling context in a collaboration environment, a generic adaptation process, and a collaboration domain model for describing collaboration environments and collaboration situations. [3] implements the framework, using an extended domain model and the related adaptation process. The resulting CONTact platform is able to sense and formalize users' interaction with the system at runtime, and to adapt according to the user's current collaboration situation. These adaptations may confuse users. Therefore, [20] use context enriched explanations to help them understand the adaptation behavior. [3] and [20] take the aspects of the comprehensibility of system behavior, decisions and data processing into account, but do not satisfy the legal requirements. Furthermore, the explanations provided are not presented in a way that is intelligible to the users. So far, there are no known context-based collaborative systems that support comprehensibility and intelligibility for users.

No approach is known to us for context-based collaborative systems that considers the requirements of the GDPR and taking up the topic Compliance by Design.

3 Domain Model: Legal Regulations

In this section we introduce the domain model and explain its concepts and relationships. We used the OWL 2 Web Ontology Language (informally OWL 2) and the Protégé Ontology Editor for modelling. In this paper we focus on concepts of user control, comprehensibility and intelligibility by considering the legal requirements.

3.1 Context Modelling

We use our approach presented in [19] consisting of the generic four-layer framework for modelling context in a collaboration environment and the related collaboration domain model for describing collaboration environments and collaboration situations. The framework contains the knowledge layer, the state layer, the contextualized state layer and the adaptation layer [19]. The knowledge layer describes a domain model with abstract (e.g. classes, properties) and concrete (e.g. individuals) predefined knowledge, mapped to corresponding concepts and relations. The state layer uses sensing rules to instantiate related concepts and relationships from the domain model (cf. knowledge layer) to represent the current collaboration environment of all users. The contextualized state layer applies contextualization strategies to extract a subset from the state (cf. state layer) and/or domain model (cf. knowledge layer) which are relevant for the current collaboration situation. This creates a contextualized state (the context). The adaptation layer evaluates the adaptation rules and executes applicable adaptation rules. This leads to the adapted state that is mapped to the collaboration environment.

To address the GDPR, we extended our domain model (cf. [19]). Figure 1 shows an excerpt of our resulting ontology (i.e. domain model and relevant instances from the state required to illustrate our approach). For readability reasons, we omitted concepts and relationships, and focused on the concepts, relationships, and instances helping to describe situations, when user control, comprehensibility and intelligibility is needed. Therefore, we use Alice who has already created an account in the app meinDorf55+ (a novel community support system for elderly) which demands personal information.

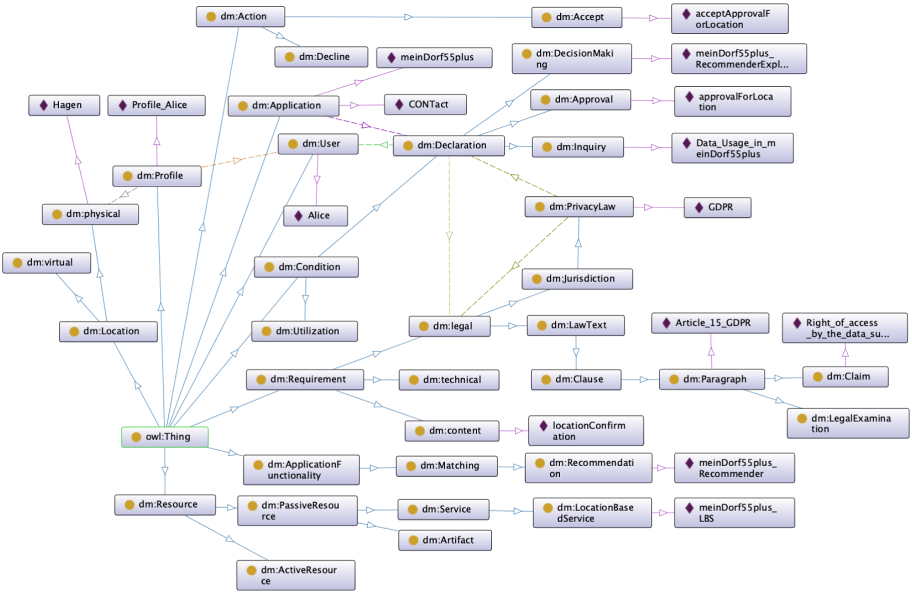

Fig. 1. Ontology representing legal and comprehensibility concepts and relations

As Fig. 1 shows, *Alice* is an instance of the concept *dm:User* in our sample ontology. *CONTact* (representing the CONTact platform) and *meinDorf55plus* (representing the novel community support system for elderly) are instances of the concept *dm:Application*. The concept *dm:Profile* is related to *dm:User* and includes the address represented as *dm:physical* that is a subclass of *dm:Location*. Despite physical locations we support *dm:virtual* as a subclass of *dm:Location*, e.g. to support URLs. Applications usually provide different kinds of functionalities. We map these to related concepts of *dm:ApplicationFunctionality* and *dm:Resource* when modelling the related opportunities. The concept *dm:Resource* can be either *dm:PassiveResource* or *dm:ActiveResource*. A *dm:PassiveResource* can be split up into a *dm:Service* and *dm:Artifact*.

Using the presented concepts and relationships we can create instances in our ontology representing related situations in our context-based adaptive collaboration environment. In case of Alice that means that *Alice* is an instance of *dm:User* and

Profile_Alice is an instance of *dm:Profile*. As soon as Alice tries to use a special *dm:ApplicationFunctionality* of *meinDorf55plus*, e.g. *meinDorf55plus_LBS* as an instance of *dm:LocationBasedService* or *meinDorf55plus_Recommender* as an instance of *dm:Recommendation*, she has to provide her address (in our ontology *Hagen* as an instance of *dm:physical*). This is when the legal regulations have to be represented in our domain model.

Figure 1 shows the concepts *dm:Requirement, dm:Condition* and *dm:Declaration* and their dependencies which are used for adaptation, user control and comprehensible explanations to the users. The aforementioned concepts are used to answer the questions *What happened?* (*dm:Requirement*), *Why does it happened?* (*dm:Condition*) and *What kind of explanation should be provided?* (*dm:Declaration*) in the specific context.

3.2 Concept dm:Requirement

Requirements are conditions for applications and define what an application (*dm:Application*) or application functionality (*dm:ApplicationFunctionality*) must check and take into account during processing. The requirements are no fixed set of rules instead they are used at runtime to find out what the application has to do in the current situation. Therefore, requirements can be seen as external policies which must be considered by an application (we use the term rule to illustrate that a related policy can be implemented in our CONTact platform). Requirements can be technical conditions (*dm:technical*), content definitions (*dm:content*) and legal regulations (*dm:legal*). These three aspects are separate domain models that are subordinated to the concept *dm:Requirement*. Technical requirements can be hardware resources that limit the execution of certain functionalities, e.g. by using mobile devices with less powerful hardware. The application has to react to this, e.g. by organizing a provision via other devices (e.g. by computing on servers). Content definitions can result from the domain of an application or a service. Figure 1 shows *locationConfirmation*, a content related requirement of the *dm:Application* instance *meinDorf55plus* that provides a location-based service (*dm:LocationBasedService*) represented in the instance *meinDorf55plus_LBS*. For that *meinDorf55plus_LBS* needs a conformation of the users location which is requested by the instance *locationConfirmation* of the concept *dm:content*.

The ontology shows an excerpt from the legal domain model *dm:legal*. It describes the German jurisdiction by depicting its taxonomy as part of the concept *dm:Jurisdiction*. The law taxonomy has different legal areas, e.g. the privacy law or the civil law. For readability reasons, Fig. 1 only contains the privacy law (*dm:PrivacyLaw*). The instance *GDPR* of the concept *dm:PrivacyLaw* represents the applicable law. Furthermore, the legal domain model depicts the general structure of the legal texts through the concept *dm:LawText* including its clauses (*dm:Clause*) and paragraphs (*dm:Paragraph*). The instance *Article_15_GDPR* of the concept *dm:Paragraph* is used to identify the claim. The instance *Right_of_access_by_the_data_subject* of the concepts *dm:Claim* represents the claim which is derived from Article 15 (*dm:Paragraph*). A paragraph can either represent a claim (*dm:Claim*) or an explanation of the right (*dm:LegalExplanation*). Both determine the activities of an application.

3.3 Concept dm:Condition

Conditions are derived from the requirements. The concept *dm:Condition* is intended to verify the correctness and legitimacy of the processing. The legitimacy arises, e.g. from the legal regulations of privacy law like in Fig. 1. Conditions are a set of abstract rules which are defined in the application to map the external requirements to the application processing. At development time not all rules are known, so they are based on the concepts and domain models of *dm:Requirement* for the specific purpose of the application. The rules have the form *WHEN condition part THEN action block*. At runtime, the application uses these constructs to check which situation it is in, which actions has to be executed, and which conditions must be fulfilled for continuing processing. The conditions, on their part, can trigger a cascade of checks that are given on the basis of the requirements of the respective domain *dm:technical*, *dm:content* or *dm:legal*. In Fig. 1 the *locationConfirmation* caused a check of legal requirements that results in the creation of different kinds *of dm:Declaration* instances.

The condition for the use of certain application functionalities (*dm:ApplicationFunctionality*) maybe also be motivated from a legal perspective. Thus, a direct interaction with the user is maybe not necessary (e.g. encrypted data transmission). The concept *dm:Utilization* of the domain model can be used for that kind of required functionality.

3.4 Concept dm:Declaration

Declarations are the interface to users which can support comprehensibility and user control. As shown in Fig. 1, the provision of an explanation depends on the requirements (e.g. legal regulations). According to Article 15 of the GDPR, data subjects whose data are collected and processed have a right to obtain information about the usage. This includes the purposes of the processing, the categories of personal data processed, the recipients to whom the data are disclosed, the duration of the storage, the existence of a right of appeal and an overview of the origin of the data, if not collected from the data subject. In addition, Article 15 declares, the data subject has the right to limit the processing by the data processor. Furthermore, a right of objection against the processing exists at any time.

Addressing Q1 and Q2, our domain model contains the concept *dm:Declaration* to be able to represent the right to obtain information about data usage. Depending on the current context the concept *dm:Declaration* is used to provide comprehensible explanations (*dm:Inquiry*), demand an approval (*dm:Approval*) or to explain processing (*dm:DecisionMaking*). The user can accept (*dm:Accept*) or decline (*dm:Decline*) the usage of his/her data by the system through an approval (answering Q1). Approvals are needed to execute an action (*dm:Action*) and depend on the requirements for the application, e.g. when personal information shall be transmitted to a third party it must be approved by the user. Figure 1 shows the instance *approvalForLocation* of the concept *dm:Approval*, which is needed to approve the usage of the users location by himself or herself for the content requirement *locationConfirmation*. Accepting it leads to the creation of the instance *acceptApprovalForLocation* of *dm:Accept* which stores all relevant information to the approved data usage.

The instance *Right_of_access_by_the_data_subject* of the concept *dm:Claim* caused the creation of the *dm:Inquiry* instance *Data_Usage_in_meinDorf* (cf. Q2). Information must also be provided on whether and how automated decision-making, including profiling, takes place. According to Article 22[3] (1) and (4), meaningful information on the logic involved, the significance and the intended impact of such processing for the data subject must be provided. This requirement is considered separately in the domain model through the concept *dm:DecisionMaking*. It is used when application functionalities for decision-making, such as a personalized recommendation (*dm:Recommendation*), is performed based on user data. The instance *meinDorf55plus_RecommenderExplanation* (answering Q2) of *dm:DecisionMaking* results from the instance *meinDorf55plus_Recommender* of the concept *dm:Recommendation* which is a subclass of *dm:Matching*.

4 Scenarios

As illustrated in the above sections, collaboration environments have to support explanations where and for what purpose personal data is being processed. We use the above scenario to illustrate our rule-based approach of supporting 'Compliance by Design', i.e. giving users control over their personal data being processed by our CONTact platform (cf. Q1). The second scenario describes how we use our formal context model for creating explanations to support the aforementioned mandatory feature (cf. Q2).

In Fig. 2 we present the scenario 'Compliance by Design' where we attempt to give users control over their personal data being processed by related applications.

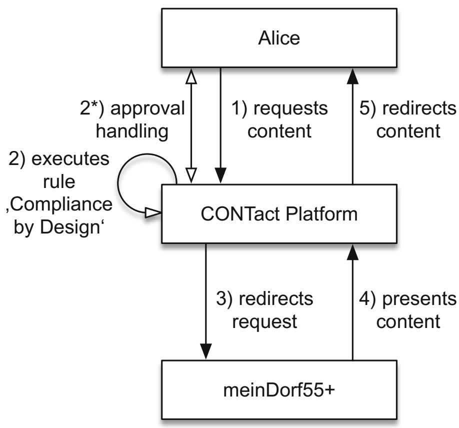

Fig. 2. Process of user interactions – Compliance by Design

The user (in our scenario Alice) requests content from the corresponding application 'meinDorf55+' through the CONTact platform (cf. (1–3)). After receiving the request,

[3] https://gdpr-info.eu/art-22-gdpr/.

the CONTact platform checks whether or not personal data will be processed (i.e. interpreting related annotations at source code level). When personal data will be processed, the rule (cf. (2)) presented in Listing 1 is executed to ensure that user (in our scenario Alice) has approved the usage of her/his personal data (represented as white arrows in Fig. 2). When the user has not authorized the data usage previously, she/he will be prompted to do so (cf. (2*)). After approving the data usage, the related request will be processed, and the content will be presented to the user (cf. (3–5)).

```
rule "Compliance by Design"
  when
    user: getUserInContext("dm:User")
    app: getApplicationInContext(user, "dm:Application")
    req: getRequirementInContext(app, "dm:Requirement")
    appr: requestApproval(user, app, req)
  then
    createOrUpdateAcceptedApproval(appr)
    notify(user, appr)
  end
end
```

Listing 1 Rule "Compliance by Design"

Listing 1 uses pseudocode to illustrate our approach to implement "Compliance by Design". The rule consists of a condition part (`when` to `then`) and an action block (`then` to `end`). `getUserInContext` retrieves the user interacting with the CONTact platform (in our scenario Alice). The function `getApplicationInContext` determines the application used by the user which is of type *dm:Application* (in our scenario meinDorf55plus). The function `getRequirementInContext` retrieves all instances and relations connected to the domain concept *dm:Requirement* of the given application. The function `requestApproval` uses the context information about the user, the application and the requirement and ensures that the user has approved the data usage. We distinguish two different situations:

(I) When the user approved the data usage beforehand, the return value of the function is empty.
(II) When there is no or an inapplicable approval instance present in the current context, the approval is requested from the user.

  a. When the user declines the data usage, the return value of the function is empty.
  b. When the user accepts the data usage, the approval information will be returned.

The action part of the above rule is executed as soon as all conditions are met (i.e. the returned information are not empty). First, we create or update the approval instance in the current context. Next, we notify the user about it. This shows our answer of Q1.

Figure 3 shows the process of creating explanations of data usages. In our sample scenario Alice wants to know, where and how her personal data is being processed. She requests the information about data usage from the CONTact platform (cf. (1)).

The CONTact platform redirects the request to the application ‘meinDorf55+’ (cf. (2)) where the data is stored and used. The related application has all the information about the requested data usage and reports it to CONTact platform (cf. (3)). Based on the returned data the CONTact platform checks which conditions are affected and requests an inquiry from the Requirements Handler (cf. (4) & (8)), that determines which legal requirements are affected and requests related templates from the Explanation Template Builder (PrivacyLaw) (cf. (5–7)). The CONTact platform uses the inquiry (cf. (8)) and reported data usage (cf. (3)) and executes the rule ‘data usage explanation’ (cf. Listing 1) to create a personalized explanation about Alice’s data usage and present it to her (cf. (9–10)).

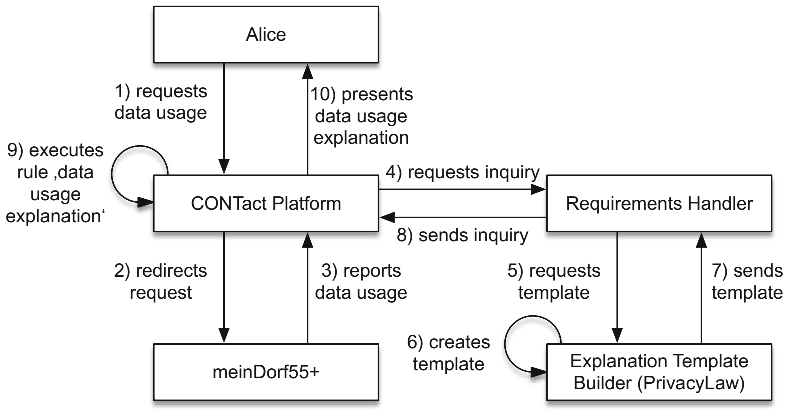

Fig. 3. Process of request and create an explanation of data usage

Listing 2 illustrates the creation of a personalized explanation about the data usage of a user (in our scenario Alice). In Listing 2, `getUserInContext` retrieves the user interacting with the CONTact platform (in our scenario Alice). The function `getApplicationInContext` determines the application used by the user which is of type *dm:Application* (in our scenario meinDorf55plus).

```
rule "data usage explanation"
  when
    user: getUserInContext("dm:User")
    app: getApplicationInContext(user, "dm:Application")
    req: getRequirementInContext(app, "dm:legal")
    tmpl: getExplanationTemplate(user, app, req)
    expl: createExplanationInContext(tmpl)
  then
    present(user, expl)
  end
end
```

Listing 2 Rule "data usage explanation"

The function `getRequirementInContext` retrieves all instances and relations connected to the domain concept *dm:legal* of the given application. `getExplanationTemplate` uses the context information about the user, the application and the requirement and generates the related explanation template that the function `createExplanationInContext` uses to create the related explanation. The action part of the above rule is executed as soon as all conditions are met (i.e. the returned information are not empty). After applying the rule to the CONTact platform, the personalized explanation is presented to the user (e.g. Alice). This illustrates our answer of Q2.

5 Discussion

The presented domain model shows the connection between applications and legal regulations in the context-based adaptive collaborative environment. The CONTact platform can be linked to or integrated into different kinds of collaboration environments that was illustrated with the novel community support system meinDorf55+. We explained the process of checking legal regulations from the privacy law GDPR while using a service that requires personal information about the user.

This paper does not cover some outstanding aspects. (I) Due to the limitation of the paper we could not explain in detail the connection of external policies with the application by the concept *dm:Condition*. (II) The presented extended domain model is only an excerpt. We focused on describing only specific concepts of the jurisdictions in it. (III) The extended domain model is only a basis for comprehensibility and personalized intelligible explanation of system processes. Users should be able to understand why something happens and how it happens in a personalized, adaptive system. The challenges of the comprehensibility of system processes includes their presentation and intelligibility. Presenting only technical information is not sufficient [7, 15]. In context-aware systems explanations "need to have access to information about complex real-world concepts that are not necessarily core to the application" [15, P. 166]. The mentioned explanation building process (cf. Fig. 3) is responsible for creating personalized explanations, e.g. when legal regulations demand it. The legal concepts of the domain model can be used to support intelligible legal explanations by the system. The intelligibility can be facilitated by the deposit of target-group-specific texts (e.g. texts created by experts) and explanations through integrated and linked dictionaries. The resulting templates could be used to provide explanations at runtime by creating instances of the concept *dm:Declaration*. (IV) While context-based adaptive collaborative environment needs adaption rules this paper does not cover it. Regarding to our four-layer context model [18], we modeled our domain model independently from adaptation rules.

6 Conclusions and Future Work

In this paper we presented an approach that enables us to answer the two questions how a user can agree that her/his personal information can be stored in and processed by the system, and how the "right of access by the data subject" (see Footnote 1) can be realized for a user. We used a sample scenario where Alice uses a context-based adaptive collaboration environment based on the CONTact platform which uses and stores personal information about her in order to support personalization. We illustrated when the two questions, mentioned above, get relevant.

Our approach is based on three main contributions: 1. an extended domain model for legal regulations (cf. Fig. 1) and 2. two processes (cf. Figs 2 and 3) and 3. related rules (cf. Listings 1 and 2). This enables us to support user control, comprehensibility and intelligibility in a context-based adaptive collaboration environment, based on the CONTact platform. We introduced the domain model and explained its concepts (especially the concepts *dm:Requirement, dm:Condition* and *dm:Declaration*) to facilitate user control, comprehensibility and intelligibility. For readability reasons we omitted other concepts and relationships that are part of our domain model, e.g., considering legal regulations other than privacy laws.

We presented our approach to give privacy control back to the users (answer to Q1), and to create personalized explanations of data usages (answer to Q2) in our context-based adaptive collaboration environment. We used the above scenario to illustrate our rule-based approach of supporting 'Compliance by Design', i.e. giving users control over their personal data being processed by our CONTact platform, and how we use our formal context model for creating personalized explanations of data usages. This helps us observing legal regulations, e.g. privacy laws like the GDPR.

We argue that the presented approach answers the above questions, but also that it does not represent the developed approach in full detail. For readability or space reasons we presented only an excerpt of our developed domain model, i.e. we have to omit the other concepts and relationships for legal regulations, and, e.g., the representation of external policies of an application using the concept *dm:Condition*.

In the next step we will investigate the challenges of comprehensibility including presentation and intelligibility. Thereby, we have to support related explanations when the user has to approve the usage of her/his personal data (cf. Fig. 2). This will lead to a combination of the two processes and related rules presented in this paper. Furthermore, we have to investigate, how the explanations have to be created, personalized and presented so that users are able to understand the meaning of the presented text and the consequences of accepting or declining the approval. This will mainly influence the presented explanation builder process (cf. Fig. 3).

Acknowledgment. The project is supported (was supported) by funds of the Federal Ministry of Food and Agriculture (BMEL) based on a decision of the Parliament of the Federal Republic of Germany via the Federal Office for Agriculture and Food (BLE) under the rural development programme.

References

1. Dey, A.K.: Understanding and using context. Pers. Ubiquit. Comput. **5**(1), 4–7 (2001)
2. Abarca, M.G., Alarcon, R.A., Barria, R., Fuller, D.: Context-based e-Learning composition and adaptation. In: Meersman, R., Tari, Z., Herrero, P. (eds.) OTM 2006. LNCS, vol. 4278, pp. 1976–1985. Springer, Heidelberg (2006). https://doi.org/10.1007/11915072_106
3. Veiel, D., Haake, J.M., Lukosch, S., Kolfschoten, G.: On the acceptance of automatic facilitation in a context-adaptive group support system. In: 2013 46th Hawaii International Conference on System Sciences, pp. 509–518. IEEE (2013)

4. Jiang, X., Landay, J.A.: Modeling privacy control in context-aware systems. IEEE Pervasive Comput. **1**(3), 59–63 (2002)
5. Brezillon, P.J.: Contextualized explanations. In: Proceedings of International Conference on Expert Systems for Development, pp. 119–124. IEEE (1994)
6. Gregor, S., Benbasat, I.: Explanations from intelligent systems: theoretical foundations and implications for practice. MIS Q. **23**(4), 497–530 (1999)
7. Dey, A.K.: Explanations in context-aware systems. In: Proceedings of the Fourth International Conference on Explanation-Aware Computing (EXACT 2009), pp. 84–93 (2009)
8. Tahir, H., Brézillon, P.: Contextual graphs platform as a basis for designing a context-based intelligent assistant system. In: Brézillon, P., Blackburn, P., Dapoigny, R. (eds.) CONTEXT 2013. LNCS (LNAI), vol. 8175, pp. 259–273. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-40972-1_20
9. Nigon, J., Verstaevel, N., Boes, J., Migeon, F., Gleizes, M.-P.: Smart is a matter of context. In: Brézillon, P., Turner, R., Penco, C. (eds.) CONTEXT 2017. LNCS (LNAI), vol. 10257, pp. 189–202. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-57837-8_15
10. Bohn, J., Coroamă, V., Langheinrich, M., Mattern, F., Rohs, M.: Living in a world of smart everyday objects—social, economic, and ethical implications. Hum. Ecol. Risk Assess. **10**(5), 763–785 (2004)
11. Wachter, S.: Normative challenges of identification in the Internet of Things: privacy, profiling, discrimination, and the GDPR. Comput. Law Secur. Rev. **34**(3), 436–449 (2018)
12. Guarino, N.: Formal ontology in information systems. In: Proceedings of the First International Conference (FOIS 1998), Trento, Italy, 6–8 June 1998, vol. 46. IOS Press (1998)
13. Dey, A.K., Newberger, A.: Support for context-aware intelligibility and control. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 859–868. ACM (2009)
14. Lim, B.Y., Dey, A.K.: Toolkit to support intelligibility in context-aware applications. In: Proceedings of the 12th ACM International Conference on Ubiquitous Computing, pp. 13–22. ACM (2010)
15. Lim, B.Y., Dey, A.K.: Design of an intelligible mobile context-aware application. In: Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services, pp. 157–166. ACM (2011)
16. Kapitsaki, G.M.: Reflecting user privacy preferences in context-aware web services. In: 2013 IEEE 20th International Conference on Web Services, pp. 123–130. IEEE (2013)
17. Kapitsaki, G., Ioannou, J., Cardoso, J., Pedrinaci, C.: Linked USDL privacy: describing privacy policies for services. In: 2018 IEEE International Conference on Web Services (ICWS), pp. 50–57. IEEE (2018)
18. Bellotti, V., Sellen, A.: Design for privacy in ubiquitous computing environments. In: de Michelis, G., Simone, C., Schmidt, K. (eds.) Proceedings of the Third European Conference on Computer-Supported Cooperative Work 13–17 September 1993, Milan, Italy ECSCW 1993, pp. 77–92. Springer, Dordrecht (1993)
19. Haake, J.M., Hussein, T., Joop, B., Lukosch, S., Veiel, D., Ziegler, J.: Modeling and exploiting context for adaptive collaboration. Int. J. Coop. Inf. Syst. **19**(01n02), 71–120 (2010)
20. Hussain, S.S., Veiel, D., Haake, J.M., Lukosch, S.: Facilitating understanding of team-based adaptation policies. In: 6th International Conference on Collaborative Computing: Networking, Applications and Worksharing (CollaborateCom 2010), pp. 1–8. IEEE (2010)

Conceptual Puzzle Pieces

An Image Schema Experiment on Object Conceptualisation

Maria M. Hedblom and Oliver Kutz(✉)

Conceptual and Cognitive Modelling Research Group (CORE),
KRDB Research Centre for Knowledge and Data, Faculty of Computer Science,
Free University of Bozen-Bolzano, Bolzano, Italy
{mhedblom,okutz}@unibz.it

Abstract. Image schemas were introduced as mental generalisations learned from the sensorimotor experiences in infancy that in adulthood shape language formation and conceptualisations. So far, little empirical research has been devoted to investigate to which degree image schemas are involved in object conceptualisation more concretely. To address this, this experimental study investigates the relationship between abstract image schemas and their involvement in conceptualisations of common, everyday objects. The experimental set-up asks participants to describe objects using abstract representations of image schemas. The results from the study support the claim that image-schematic thinking is prevalent in the conceptualisation of objects, thus providing empirical evidence for the idea that image schemas can serve as conceptual building blocks for the meaning of objects.

Keywords: Image schemas · Affordances · Common sense · Conceptual structure

1 Introduction

Stemming from the ideas behind embodied cognition, the theory of image schemas proposes that there exists a number of pre-linguistic abstract patterns learned from repeated sensorimotor experiences in early infancy. While they are derived from particular contexts, such as 'plates resting on tables' or 'being inside the crib,' they are generatisations of the relevant spatio-temporal relationships (here SUPPORT respectively CONTAINMENT). These generalisations are then used to reason about similar situations in different contexts, for instance in abstract language, in the conceptualisation of concepts, and when considering the affordances of physical objects [19,23]. To demonstrate the power of this reasoning, [22] proposed that the concept underlying the term 'transportation' can be captured simply by considering the behaviours of the two image schemas SOURCE_PATH_GOAL (SPG) and SUPPORT (alternatively CONTAINMENT). Likewise, abstract concepts such as 'marriage' could be described using

This experiment is part of the research presented in [13].

G. Bella and P. Bouquet (Eds.): CONTEXT 2019, LNAI 11939, pp. 98–111, 2019.
https://doi.org/10.1007/978-3-030-34974-5_9

a combination of the image schemas Link and SPG since a common conceptualisation of marriage is that of two 'parts' moving together along the axis of time [25]. This line of modelling is in accordance with the 'image schema profiles' of [26], which are a form of unsorted image schema combinations that can be used to describe events. Extendeding this line of thinking, [5,15,18] include also *complex event conceptualisations*, arguing that conceptualisations of complex events can formally be described using temporally structured collections of image schemas.

This straightforward line of reasoning is appealing both from a philosophical and a formal direction when modelling concepts and events, in particular in relation to transferring valuable information to new contexts. However, so far few non-linguistic studies have been devoted to evaluate the role of image schemas in conceptualisations [9]. To aid in this research endeavour, this paper presents an experimental study that investigates the relationship between a series of image schemas and the conceptualisation of objects.

2 Transfer of Information Through Image Schemas

The commonly studied image schema Containment is learnt at an early age as movement In and Out of containers is a common sight in many different contexts. From these contexts children can generalise that objects can be inside/outside of each other, that they can cross borders and they learn the implicit rules of containers. For instance, that a container needs to be larger than the 'containee'. Similarly, the image schema Verticality is believed to be acquired through experiences with the body's vertical axis and the perceivable effect gravity has on falling objects [19]. From conceptual patterns such as these, the infant can make predictions in new contexts that are similar to previous experiences. Through this form of analogical transfer, a child can early conclude that if a cup can contain coffee, a glass can contain water. In adult language, image schemas often constitute the conceptual skeleton for metaphors and analogies. The expressions 'to fall in love' and to have an 'open marriage,' both utilise the notion of the Containment schema; and the concepts 'career ladder' and 'fall from grace' are based on an understanding of Verticality and the conceptual metaphor 'UP is GOOD' [21].

Partly due to the interdisciplinary nature of the research field, there currently exists no coherent and agreed upon list of image schemas. This problem is amplified further by the heterogeneous way image schemas seem to manifest. For instance, image schemas by their nature undergo spatio-temporal transformations [26]. This means that the image schema itself is not an exclusively static notion but involves dynamic transformation as well [29]. The image schema Containment for example, includes in addition to the notion of 'enclosure' also the event transformations of 'going In' and 'going Out' (see [14] for a formal investigation of Containment). Purely looking at the classic definition of Containment, namely the 'inside-border-outside' relationship [3,19] could identify no less than eight different kinds of Containment through a corpus study.

This indicate that image schemas are not isolated theories that easily can be defined, but instead a complex web of associated notions and transformations.

Tightly connected to the theory of image schemas is the theory of *affordances* introduced by [10]. Affordances are the potentialities for action for an object in a particular context and environment, independent of an individual's ability to recognise those potentialities. Gibson proposed that affordances constitute the meaning of an object. For instance, a 'chair' is a chair if it affords 'sitting' and a 'coffee cup' is a coffee cup if it allows a user the possibility to 'drink coffee' from it. In terms of image schemas, this would correspond to the image schemas Support respectively Containment. The reality of such a direct and inter-subjective link between an everyday object and (specific bundles of) affordances resp. image schemas is the central hypothesis tested in the experimental setup in this paper. We here rely on the fact that objects can generally be described using their functionality and corresponding image schematic structure, and investigate inhowfar we can find statistically significant agreement patterns.

Establishing such patterns should have a significant impact on the formal representation of everyday concepts and objects, particularly when shifting contexts to e.g. a more abstract setting or to a different common sense domain. However, this hypothesis needs to be evaluated more strongly within behavioural psychology and not only analysed from the direction of cognitive linguistics. Below we discuss some related research in this direction.

2.1 Related Work on Conceptualisation

A classic work on conceptualisation is [20], who investigated infants' understanding of objects though a series of experiments on object occlusion. Their study shows that infants, already in the early months, understand the relationship between 'behind' and 'in front.' In terms of image schemas, their work demonstrates also that children at this early age have a conceptualisation of Link as they can register that two parts moving in unison behind an occlusion belong to the same object.

[1], and the follow up study [2], performed experiments on music conceptualisation in relation to cognitive metaphor theory in different contexts. Important findings were that musical concepts are often conceptualised by using visuo-spatial conceptual metaphors. For instance, musical scales are directly related to UpDown or Verticality. This means that the mind, either through learned behaviour or through some other cognitive adaptation, ascribes image-schematic structures to abstract patterns in sound.

Another significant contribution on the link between conceptualisation and image schemas is the research by [24]. In their book, they present theoretical support for the notion that image schemas construct the conceptual foundation for abstract mathematical concepts. For instance, addition and subtraction are according to the authors perceived as movements along a path, in other words SPG. Also, Venn diagrams used to describe set-theoric notions are related to a direct visual representation of the Containment schema. In their work they eventually discuss increasingly abstract concepts including tracing down the conceptualisation of ‘infinity’ and ‘zero’ from embodied experience and image-schematic structures.

Looking at spatial categories for ontology building, [22] uses ontological properties of image schemas to formally construct concepts’ underlying meanings. His work takes a straightforward approach to how image schemas can be used as conceptual building blocks for concept definitions independent of context. In [12], the authors present a crowdsourcing study on how to gather image-schematic structures for different linguistic expressions similar to the previous approach.

Following the ideas underlying the invariance principle[1] [30], the research in [17] motivates how image schemas are part of creative thinking through conceptual blending[2] [8] by considering them as the most generic and important building blocks for concepts and, therefore, playing a central role in the creative transformation of knowledge.

3 Experimental Foundation

3.1 Hypothesis and Purpose

This work hypothesizes that image schemas are conceptual building blocks used in the conceptualisation of objects, concepts, and events. Assuming this, it must be possible to investigate to which degree image schemas are involved in concept generation and understanding. This study tests that hypothesis by looking at instances of common objects and their conceptual connection to image schemas.

The purpose of the study is twofold. First, we want to confirm the hypothesis that image-schematic thinking plays a pivotal role in conceptualising common objects. Second, if this is indeed the case, we want to see what the differences in importance is, of assigning specific image schemas to specific object.

3.2 Material and Its Motivation

Representing Image Schemas. The first important obstacle to deal with was to select a feasible number of image schemas for the study. This presented two problems. First, which image schemas should be chosen given the large number of image-schematic structures proposed in the literature. Second, as image schemas are used to model abstract conceptual patterns, how should they be re-presented and presented in a meaningful way in the study setup.

The selection of image schemas was based primarily on two criteria. First, their commonality in the literature, with the motivation that the more commonly studied an image is, the more reliable (or at least agreed upon) their image-schematic structure/semantics was. Second, the image schemas needed to

[1] The invariance principle states that analogical transfer is based on both source and target domains being structured by the same image schema.

[2] In short, conceptual blending is a theoretical framework built on the idea that concept invention is a process of selectively combining existing knowledge domains into new domains.

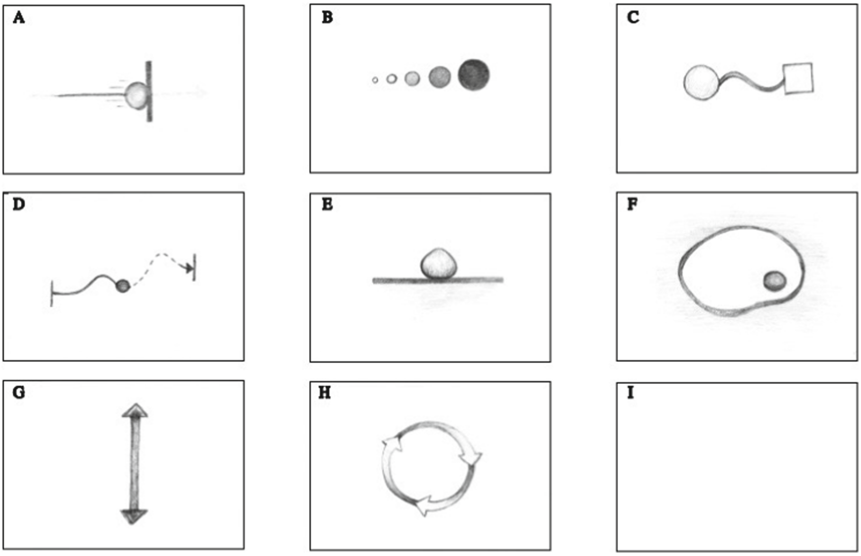

Fig. 1. The image schema illustrations as used in the experiment. From *A*−*I* as follows: *A*: Blockage; *B*: Scale; *C*: Link; *D*: Source_Path_Goal (SPG); *E*: Support; *F*: Containment; *G*: Verticality; *H*: Cycle; *I*: the empty set.

be presentable in such a way that it could be intuitively understood what the image schema entailed. Aiming in part for a simple visual representation, this disqualified for our purposes certain image schemas that are either too complex for unambiguous pictorial representation, or are predominantly dynamic rather than static.[3] The representation of the image schemas thus used two methods, consequently dividing the participants into two groups. The first used just the linguistic phrasing of the image schemas, whilst the second group was confronted with a basic visualisation made with graphite pencil on white sketch paper. Figure 1 contains the visual illustrations of the eight selected image schemas.

Objects for Assessment. While deciding which objects to choose, we deemed important that they were objects that children come into contact with early on in conceptual development. This is important as the conceptual core of the presented concepts should be a rather basic conceptualisation, and be sheltered from cultural contexts and influence as much as possible. A second aspect was that in order to avoid priming the participants with language, the objects should be presented visually rather than by written description. Therefore, flashcards for language learning were used.[4]

[3] Looking at dynamic image schemas employing moving images for representation is a promising direction for a follow-up study.

[4] http://www.kids-pages.com.

Focusing solely on nouns, the selection process excluded all verbs and adjectives. Likewise, all animate objects as well as human roles were eliminated as the association and conceptualisation to these categories may be clouded by personal and cultural experience. To get an as unbiased sample as possible, 44 objects were selected 'randomly' from the available flashcards applying the discussed restrictions (see next section for a list).

Pilot Study and Establishing a Gold Standard. As a pilot, three image schema experts performed the experiment by assigning image schemas to the objects, generating a series of image schema categories. The results can be found in Table 1 which shows the objects after they had been sorted into their respective categories. Note that, on occasion, objects occur more than once as a consequence of the experts assigning multiple image schemas per object. The table is also missing three objects that were included in the experiment (camera, lamp, pacifier) since these had not been assigned any particular, agreed upon, image schemas by the experts.

Table 1. The objects sorted into image-schematic category by majority of the experts.

Image schema category	Objects
VERTICALITY	Colour pyramid, stiletto shoes, ladder, plant sprout, sunflower, skyscraper, stairs, tree
CYCLE	Clock, screw, sunflower, washing machine
CONTAINMENT	Banana, bathtub, boiled eggs, car, cherries, computer, guitar, hat, house, mirror, oven glove, pants, school bag, skyscraper, strainer, wardrobe, washing machine
SPG	Air plane, car, garden path, lightning, ruler, stairs, wheelbarrow
SUPPORT	Bed, play blocks, chair, plate, stiletto shoes, sofa, wheelbarrow
LINK	Ankle, cherries, computer, lightning, pliers
BLOCKAGE	Oven glove, strainer, umbrella
SCALE	Colour pyramid, fire, plant sprout, ruler, thermometer

3.3 Participants and Experimental Groups

The experiment consisted of 25 participants gathered using a convenience sampling. From these, four participants had to be eliminated due to not following the instructions of the experiment. The remaining 21 participants (females: 28.6%, males: 71.4%) came from various cultural backgrounds, had twelve different mother-tongues and ages ranging between 25 to 60 (*mean age:* 36.2, *median age:* 32, *SD:* 10.68, *variance:* 108.77).

As mentioned above, the participants were divided into two groups. Ten participants were presented with the illustrations of the image schemas (Fig. 1) and eleven participants were instead presented with the names of the image schemas.

In order to avoid a possible bias towards particular image schemas based on placement on the sheets, each group was divided into an additional three groups, where the image schemas had been spatially re-arranged. Before all the data was analysed, this data was aligned to make sure that all answers were based on the same material.

4 Method

The experiment started with a brief oral introduction including encouraging the participants to carefully read the written instructions. Written instructions had been selected to avoid accidentally providing the participants with different instructions.

The experiment consisted of the participants to familiar themselves with the alternatives $A - I$ on the image schema sheet, followed by flipping through the 44 flashcards[5] and to 'describe' the object on them by matching them to one or more of the abstract image schemas. They were explicitly asked not to focus on visual attributes of the illustrations nor of the objects, but instead to "think holistically about the object." The experiment also required the participants to write a short motivation statement to explain the 'reasoning' that led to their choices.

4.1 Methods of Analysis

Analysis of the Method Behind Object Conceptualisation. The study aimed to investigate whether participants used image-schematic thinking to conceptualise the objects. To determine this, only the data in the motivations was used, in a qualitative analysis, as the assigned image schemas were considered irrelevant for the mode of thinking.

Presented below are the four major methods that we analysed to be at work when conceptualising the objects:

- **(i) Image schemas:** if the motivation contained the abstract, spatio-temporal motion or relationship found in image schemas. Examples: Stiletto shoes: VERTICALITY and CONTAINMENT, *increase height of person, contain feet*; Umbrella: BLOCKAGE, *blocks rain and sun.*
- **(ii) Association:** if the motivation made association to similar concepts and objects to that on the flashcard. Examples: Lightning: CYCLE, *the water cycle*; Ankle: SPG, *running towards a goal.*

[5] Note that the three objects excluded from the gold standard were still present in the experiment to allow for a larger span of interpretation.

(iii) Visual/attribute cues: if the motivation made direct visual or attribute connections between the object and the image schemas. Examples: Camera: LINK, *lens is round, picture is square*; Boiled eggs: CONTAINMENT, *illustration looks like an omelette.*

(iv) Other: This was used when none of the previous methods were deemed applicable. Examples: Clock: BLOCKAGE, *if it falls it breaks*; Guitar: SUPPORT, *supports a singer.*

Analysis of Image Schemas Attributed to Objects. The second research question was to determine if it is possible to assign particular image schemas to certain objects. For this part, the data from both groups was merged, motivated by assuming that the illustrations and the terms could be treated equally. This was approached by, similarly to the expert pilot study, generating image-schematic categories from the majority of the participants. At the same time, three other aspects were looked at more closely: First, the objects that had the highest assignment of 'nothing;' Second, the most consistently defined objects; and third, the objects that had more than 50% of a *particular combination* of assigned image schemas.

The objects that best matched these criteria were then presented and discussed to find the commonality amongst the objects that had the highest image-schematic structure.

5 Results

5.1 Assignment Method

Table 2 shows the distribution in percent between the different assignments methods for respective participation groups.

Table 2. Distribution of method for assigning the image schemas to the objects per participant group.

Assignment method:	Illustrations	Terms
Image schemas	70.82	65.48
Association	14.66	14.88
Visual/attribute cues	14.65	7.14
Other	10.23	12.50

The results show a dominance in using a method of abstract image-schematic thinking when describing the common objects in approximately 2/3 of the time regardless of them being presented with illustrations or terms. This result gives a strong indication that the participants were thinking abstractly enough and in line with the goals of the experiment.

5.2 The Image-Schematic Structure of the Objects

Table 3 shows the image schema categories where at least 50% of the participants agreed upon a particular image schema for the same object. While the table demonstrates a great reduction in assignment in comparison to the one made by the experts (see Table 1), it does represent a near perfect mapping. Out of the objects that the participants agreed upon, all but one ('plant sprout') had been assigned the same image schemas by the experts, illustrating that while the participants had no trained knowledge of the concept of image schemas, their intuitions closely align with those of the experts.

Table 3. The objects sorted into image-schematic categories by at least 50% of the participants.

Image schema category	Objects
VERTICALITY	Ladder, skyscraper, stairs, tree
CYCLE	Clock, plant sprout, screw, washing machine
CONTAINMENT	Bathtub, boiled eggs, house, school bag
SPG	Air plane, car, garden path
SUPPORT	Bed, chair, sofa, wheelbarrow
LINK	Computer
BLOCKAGE	
SCALE	Ruler

The Highest Non-Image-Schematic Objects. Whilst the experts could not find an agreement on three objects, this number was higher for the participants. The following objects had six or more 'no image schema assigned:' Camera, lamp, fire, hat, banana, oven glove, guitar, lightning, pants and strainer. Out of these objects, 'camera' and 'lamp' were among the objects not defined also by the experts.

The Most Consistently Assigned Objects. After counting the number of assessed image schemas to each object and per person, a few objects ranked higher in agreement of the assigned image schema. The objects that had at least 2/3 of the participants agreeing on the image schema were: chair (SUPPORT), garden path (SOURCE_PATH_GOAL), sofa (SUPPORT), ladder (VERTICALITY), bathtub (CONTAINMENT), washing machine (CYCLE) and stairs (VERTICALITY).

The Objects with Multiple Image Schemas. For some objects, the pattern for assigning image schemas were spread widely amongst the different alternatives to assign image schemas. However, for several of the objects the assigned image schemas occasionally were arranged in patterns in which more than one image schema played a central role in its conceptual description. The objects which had two (or on occasion three) assigned image schemas that 'in combination' had been assigned by at least 50% of the participants, are: wheelbarrow

(Support, Containment, Source_Path_Goal), sunflower (Verticality, Cycle), stiletto shoes (Support, Verticality), play blocks (Verticality, Link) and ankle (Support, Link).

6 Discussion

6.1 Method Discussion

Sample. The participants were gathered through a convenience sampling. This resulted in a higher than average level of education of the participants, which in turn could have resulted in unintended 'over thinking.' However, given the purposes of the experiment, we believe that the possible effects of this are minimal and that they can be disregarded.

Likewise, the gender distribution is uneven. However, since the experiment does not presume any gender difference in cognitive conceptualisation (supported by e.g. [27]) and the cognitive mechanisms investigated ought to not be influenced by any potentially existing gender-cultural differences, it is believed that this uneven distribution can be disregarded as well.

The divergence in nationality, consequently also in native language, and the varied age of the participants are thought to produce a fairly solid sample. Naturally the sample size lies in the lower margin with only 21 participants whose performance could be counted into the analysis of the experiment. In order to properly assess the generalisability of the results, further studies need to be conducted.

Material. Regarding the image schema illustrations, the results illuminated a few issues with some of the illustrations. The biggest challenge of making the illustrations was to capture the whole family of notions involved in the image schemas, meaning that Containment should also include the notions of In and Out, and Verticality should include vertical movement and/or relative position in either direction of UpDown. Likewise, Source_Path_Goal was required to cover not only movement, but the source and the goal as well, as brought into light and discussed in [11,16]. The experiment used a set-up with static illustrations, suppressing the dynamic aspects of the image schemas. To balance this issue, the instructions contained an explanatory text: *"The nine illustrations are meant as capturing a mental 'idea' and while this abstract content should remain you may perform transformations to apply it to the context of the object."* However, it is not clear whether these clarifications were interpreted in the intended way. For instance, one of the participants violated the Verticality principle by transforming it into 'horizontality,' rather than preserving the verticality through other means of transformations.

Additionally, the results indicated that the image schema illustrations might have been a bit too abstract. The participant's written motivations occasionally demonstrated misapprehension of some of the illustrations, where Link was the illustration to gain the most incoherent interpretations. Naturally, this had negative effects on the results, producing outliers.

For further and similar experiments, the image schema illustrations presented in this study may be used as a guide, but ought to be mildly modified in order to better capture the dynamics of the underlying spatio-temporal relationships.

As previously motivated, the objects had been chosen because of their commonality in everyday life, varying from simple (e.g. chair, house) objects to increasingly complex objects (e.g. camera, washing machine). Their visual representation utilised pre-designed flashcards to have a homogeneous design. The goal of choosing objects with a coherent visual representation was to lower the possible problems due to participants being distracted and associating the objects on the cards with particular visual characteristics. While most of the pictures caused no misapprehension in the subjects, two of the illustrations appeared to have been borderline cases: the 'skyscraper,' to which several participants asked what the picture portrayed, and the 'ankle' which some participants (as illuminated in the participants' motivation) had interpreted as a 'foot.'

6.2 Result Discussion

Method Behind Object Conceptualisation. With approximately 2/3 of the image schema assignments determined through abstract image-schematic thinking, the results provide support towards objects being conceptualised in accordance with the main hypothesis of this experiment, namely that image schemas lie at the foundation of, and give structure to, the meaning of concepts [19,23]. These findings are also in accordance with those found in the related work on image-schematic structures in music conceptualisation, carried out by [2], who showed that music conceptualisation is often based on visuo-spatio-metaphors.

Image Schemas Assigned to Objects. The results show a near perfect correlation between the expert assessment (taken as gold standard) and the most commonly assigned image schemas per object. While the experts demonstrated a superior level of detail in terms of which image schemas were assigned, among the objects where the majority of the participants assigned the same image schema, there was only one instance that did not correspond to the expert's choice. While this is an encouraging result underpinning that image schemas can be seen as conceptual building blocks, the rather large variance in choices needs attention. One reason for this might be due to the image schemas not being comprehended completely. A second reason could also lie in the observed high reluctance among some of the participants to assign more than one image schema per object. This naturally resulted in a smaller set of image schemas to be distributed over the objects than in the more generously assigned image schemas found amongst the experts.

Regarding that some objects had a higher number of 'no image schema assigned,' might be naturally due to the objects being perceived as more complex. For example, the underlying conceptualisation of an object such as a 'camera' might be far more affected by associations and the 'complex' usage than objects with far more straight-forward usages such as a 'chair.' Indeed, complex technical artefacts have low image-schematic content in an expert assessment, too.

Likewise, objects such as 'banana' and 'cherries' whose primary function (for humans) is to be eaten may also not carry clear image-schematic content in terms of spatio-temporal relationships as used in this study, but rather have other affordance-based conceptual primitives associated, relating to nutrition and providing physical energy[6].

The objects that were most coherently assigned image schemas were those objects where the usage of the object, and people's contextual experience, are more or less homogeneous amongst individuals. The possible uses and experiences a person has with a 'chair' is more or less identical in all (adult) individuals. Likewise, different modes of transportation are heavily associated with the notion of going from one place to another; therefore, concepts such as car, air plane and garden path naturally are associated to the Source_Path_Goal schema in accordance with the ideas presented by [22].

7 Conclusion and Outlook

The notion of 'image schema' is central to transferring information across contexts and as such has been for decades an influential notion for conceptual metaphor theory, and more broadly for cognitive linguistics. More recently, image schemas have been increasingly common in more formal domains, particularly in methods to solve some of the problems with information transfer from one context to another. Namely, the straightforward way image schemas are learned by humans from different contexts—only to be generalised into patterns that can be adapted into a whole range of different scenarios—is the main reason they hold interest as conceptual building blocks for computational natural language processing and cognitively inspired robotics, to name a few.

In the presented study, we demonstrated that image-schematic thinking is used not only for analogical transfer and metaphors, but is also found in the focused conceptualisation of objects. Including image schemas annotations in object descriptions could thus be useful also for computational approaches to metaphor and natural language understanding [4,31], tool substitution in robotics [28] as well as computational creativity and concept invention [6,7,17].

Future work will have to confirm the findings in more refined set-ups by extending the investigation to include more dynamic presentations of image schemas and address the multi-modality of image-schemas beyond the basic spatio-temporal interpretation.

6 In accordance with embodied cognition and the multi-modal nature of image schemas, it is possible that primitives such as those found in taste and bodily reactions to food should also be included in the research field of image schemas. However, to the authors' knowledge little such research exists as of yet.

Acknowledgements. The authors thank the reviewers for their useful comments. We would also like to thank Mihailo Antović for his assistance regarding data analysis and the study of image schemas. Our thanks also extend to John Bateman, Tony Veale and Rafael Peñaloza for valuable input during discussions on experimental set-up and analysis methods.

References

1. Antović, M.: Musical metaphors in Serbian and Romani children: an empirical study. Metaphor Symb. **24**(3), 184–202 (2009)
2. Antović, M., Bennett, A., Turner, M.: Running in circles or moving along lines: conceptualization of musical elements in sighted and blind children. Musicae Scientiae **17**(2), 229–245 (2013)
3. Bennett, B., Cialone, C.: Corpus guided sense cluster analysis: a methodology for ontology development (with examples from the spatial domain). In: Garbacz, P., Kutz, O. (eds.) 8th International Conference on Formal Ontology in Information Systems (FOIS), Frontiers in Artificial Intelligence and Applications, vol. 267, pp. 213–226. IOS Press (2014)
4. Bergen, B.K., Chang, N.: Embodied construction grammar in simulation-based language understanding. In: Östman, J.O., Fried, M. (eds.) Construction Grammars: Cognitive Grounding and Theoretical Extensions, pp. 147–190. John Benjamins Publishing (2005)
5. Besold, T.R., Hedblom, M.M., Kutz, O.: A narrative in three acts: using combinations of image schemas to model events. Biol. Inspired Cogn. Archit. **19**, 10–20 (2016)
6. Confalonieri, R., Kutz, O.: Blending under deconstruction: the roles of logic, ontology, and cognition in computational concept invention. Ann. Math. Artif. Intell. (2019)
7. Confalonieri, R., et al. (eds.): Concept Invention: Foundations, Implementation Social Aspects and Applications. Computational Synthesis and Creative Systems. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-65602-1
8. Fauconnier, G., Turner, M.: Conceptual integration networks. Cogn. Sci. **22**(2), 133–187 (1998)
9. Gibbs, R.W., Colston, H.L.: The cognitive psychological reality of image schemas and their transformations. Cogn. Linguist. (Includes Cogn. Linguist. Bibliogr.) **6**(4), 347–378 (1995)
10. Gibson, J.J.: The Ecological Approach to Visual Perception. Houghton Mifflin, Boston (1979)
11. Gromann, D., Hedblom, M.M.: Breaking down finance: a method for concept simplification by identifying movement structures from the image schema path-following. In: Proceedings of the Second Joint Ontology Workshops (JOWO), CEUR-WS online proceedings, Annecy, France, vol. 1660 (2016)
12. Gromann, D., Macbeth, J.C.: Crowdsourcing image schemas. In: Proceedings of TriCoLore, CEUR-WS, Bolzano, Italy, vol. 2347 (2019)
13. Hedblom, M.M.: Image Schemas and Concept Invention: Cognitive, Logical, and Linguistic Investigations. Cognitive Technologies, Springer, Heidelberg (2019). Forthcoming
14. Hedblom, M.M., Gromann, D., Kutz, O.: In, out and through: formalising some dynamic aspects of the image schema containment. In: Proc. 33rd Annual ACM Symposium on Applied Computing, Pau, France, pp. 918–925 (2018)

15. Hedblom, M.M., Kutz, O., Mossakowski, T., Neuhaus, F.: Between contact and support: introducing a logic for image schemas and directed movement. In: Esposito, F., Basili, R., Ferilli, S., Lisi, F.A. (eds.) AI*IA 2017: Advances in Artificial Intelligence, pp. 256–268 (2017)
16. Hedblom, M.M., Kutz, O., Neuhaus, F.: Choosing the right path: image schema theory as a foundation for concept invention. J. Artif. Gen. Intell. **6**(1), 22–54 (2015)
17. Hedblom, M.M., Kutz, O., Neuhaus, F.: Image schemas in computational conceptual blending. Cogn. Syst. Res. **39**, 42–57 (2016)
18. Hedblom, M.M., Kutz, O., Peñaloza, R., Guizzardi, G.: Image Schema Combinations and Complex Events. KI - Künstliche Intelligenz, July 2019. Special Issue on Cognitive Reasoning, https://doi.org/10.1007/s13218-019-00605-1
19. Johnson, M.: The Body in the Mind: The Bodily Basis of Meaning, Imagination, and Reason. The University of Chicago Press, Chicago and London (1987)
20. Kellman, P.J., Spelke, E.S.: Perception of partly occluded objects in infancy. Cogn. Psychol. **15**, 483–524 (1983)
21. Kövecses, Z.: Metaphor: A Practical Introduction. Oxford University Press, USA (2010)
22. Kuhn, W.: An image-schematic account of spatial categories. In: Winter, S., Duckham, M., Kulik, L., Kuipers, B. (eds.) COSIT 2007. LNCS, vol. 4736, pp. 152–168. Springer, Heidelberg (2007). https://doi.org/10.1007/978-3-540-74788-8_10
23. Lakoff, G.: Women, Fire, and Dangerous Things. What Categories Reveal about the Mind. The University of Chicago Press, Chicago (1987)
24. Lakoff, G., Núñez, R.: Where Mathematics Comes From: How the Embodied Mind Brings Mathematics into Being. Basic Books, New York (2000)
25. Mandler, J.M.: The Foundations of Mind : Origins of Conceptual Thought: Origins of Conceptual Though. Oxford University Press, New York (2004)
26. Oakley, T.: Image schema. In: Geeraerts, D., Cuyckens, H. (eds.) The Oxford Handbook of Cognitive Linguistics, pp. 214–235. Oxford University Press, Oxford (2010)
27. Richardson, J.T.E., Caplan, P.J., Crawford, M., Hyde, J.S.: Gender Differences in Human Cognition. Oxford University Press, New York (1997)
28. Thosar, M., et al.: From multi-modal property dataset to robot-centric conceptual knowledge about household objects. arXiv preprint arXiv:1906.11114 (2019)
29. Tseng, M.Y.: Exploring image schemas as a critical concept : toward a critical-cognitive linguistic account of image-schematic interactions. J. Lit. Semant. **36**, 135–157 (2007)
30. Turner, M.: An image-schematic constraint on metaphor. In: Conceptualizations and Mental Processing in Language, pp. 291–306 (1993)
31. Veale, T., Shutova, E., Klebanov, B.B.: Metaphor: A Computational Perspective. Synthesis Lectures on Human Language Technologies. Morgan & Claypool Publishers, London (2016)

Measuring Insight in the Classroom

John Hegarty[1(✉)] and Regis Maubrey[2]

[1] Ipag Business School, Paris, France
john.hegarty@ipag.fr
[2] Greenway International, Paris, France

Abstract. Organizations learn by comparing practices of employees under different aspects and leverage the lessons learned in improved procedures [9]. This paper reports on a pilot research project that shows how the theory of practice-based organizational learning can be transposed to the classroom. Students' insights about their learning practice are linked to, and reflected in, better learning performance in the subject matter. Peer-controlled self-evaluation is used to measure subject-matter understanding on a scale inspired by Lonergan's cognitional theory [12]. The pilot research project presented here was undertaken in the 2018/2019 academic year in a French business school in a management class. A multi-class research project in the same school is planned by the authors for the academic year 2019/2020.

Keywords: Insight · Practice-based learning · Peer-controlled self-evaluation

1 Introduction

Twenty years ago, a CONTEXT paper explicated the pragmatic roots of context going all the way back to Aristotle [7]: context is used by task realizers to separate activity from background, enabling attention to be focused on the task at hand. Another paper, in the same CONTEXT conference, showed how artificial intelligence researchers used context to represent actual practice and not just the procedural description of an activity [5]. Building on these practical approaches to context, one of the authors of this paper, showed how representing the unfolding realization of tasks in organizations opens up opportunities for practice-based organizational learning and performance improvement [8].

Independently, over the same period of time, the other author studied the role of socio-linguistics in the transmission of knowledge among adults [14]. Recently, the authors, who both teach in a business school in Paris, and share an interest in Bernard Lonergan's cognitional theory [11, 12], have begun a research program dedicated to "Insight in Learning and Technology Transfer". This paper is their first joint publication and it reports on a pilot research project that is part of the wider program.

The theory of practical understanding that underpins the contextual graph approach [6] is brought under a wider theory of knowledge as a dynamic structure of experiencing, understanding, and judging [11, 12]. We show how contextual graphs may be used by

G. Bella and P. Bouquet (Eds.): CONTEXT 2019, LNAI 11939, pp. 112–119, 2019.
https://doi.org/10.1007/978-3-030-34974-5_10

learners to express their insights into their own practice of learning, and, how peer-controlled self-evaluation may be used to measure insight into the subject matter being learnt. Specifically, this paper reports on a pilot research project that investigates the relationship between insights into learning practice and insights into content learned as they occur in the learning of a management activity in a business school.

Hereafter, the paper is organized as follows. Section 2 presents the relevant literature on insight in the classroom, Sect. 3 presents the methodology used in the pilot research project, Sect. 4 presents the results obtained, and Sect. 5 concludes and indicates avenues for future research.

2 Insight in the Classroom

In this section we present the relevant literature. There is a large body of research on insight, the so-called *Aha!* Experience [3, 16], and also on measuring learning objectives [13], but there is little research on formally measuring insight in the classroom, perhaps because of a methodological bias against self-reports. But, researchers, who are themselves teachers, see their students having insights, or reporting on their insights in the classroom. The research presented here starts from that fact and asks how to help students measure their insight and to adapt their learning practice so as to improve learning effectiveness and efficiency.

This raises two questions: first, how to present content in a manner suitable for self-evaluation? and, second, how to represent learning practice in a way that students can share their best practices?

Section 2.1 presents a theory of insight and its relation to the learning of content [12] and derives conditions for the effective presentation of content suitable for self-evaluation, together with an example of such content.

2.1 Insight into the Structuring of Content - Didactics

In the critical realist tradition that inspires this research, insight is understanding, understanding correctly is knowing, and knowing intends being (reality) [11]. Knowledge is a dynamic structure (self-assembling, self-constituting) with many distinct and irreducible activities, not experience alone, not understanding alone, not judgment alone [12].

The implications of this approach to learning are three conditions on the learner. "To have an insight, one has to be in the process of learning or, at least, one has to reenact in oneself previous processes of learning. While this is not particularly difficult it does require (1) the authenticity that is ready to get down to the elements of a subject, (2) close attention to instances of one's own understanding and, equally, one's own failing to understand, and (3) the repeated use of personal experiments in which, at first, one is genuinely puzzled and then catches on."

Transposed to a business school classroom, these three types of activities (experiencing, understanding and judging) may be practiced by students in situation-based learning, where they are asked to put themselves in the decision-making situation of the manager or entrepreneur confronted by a particular opportunity or problem. In undergraduate classes, the problem is often explicitly provided by the teacher, whereas, in post-graduate classes, the student must find the problem based on the data provided about the activity and the environment.

It follows that the content engaged with by the student must be amenable to situated reasoning. Figure 1 presents a model of content that is eminently suitable for a management class; the Lemoigne model of the organization as an active open system in a changing environment [10] is extended, firstly, by using Porter's value chain to represent generically all possible operating system activities [1], and secondly, to include Key Performance Indicators (KPI's) for those activities together with what the authors call Key Situation Indicators (KSI's) that contextualize the performance of the focal activity with respect to factors (political, economic, socio-cultural and technological) and actors (stakeholders) that bear on the performance of that activity.

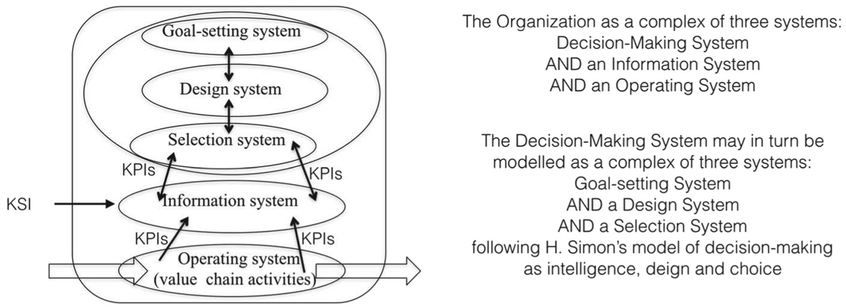

Fig. 1. Example of content: Lemoigne Model of the Organization (extended)

The model is an example of knowledge that is hierarchical/chunked [15] and is situated in a specific context (case study). In order to be able to self-evaluate, in a transparent and comparable way, students must be presented with learning content structured in summary form. This structuring of the content is sometimes called didactics. Some of the issues are the number of items to be evaluated, the number of levels of chunking chosen in the hierarchical representation, and the order of presentation.

In the learning table, a simple student self-evaluation tool, illustrated in Table 1 below, the lines represent such a hierarchy of Learning Objectives (LO) provided by the teacher. In fact, they correspond to an intermediate level of chunking between the more detailed learning objective hierarchy in the course textbook and the less detailed summary level in the course syllabus. In general, students as they master a subject will develop their own personalized hierarchies.

Section 2.2 presents the theory of insight applied to the learning of learning.

2.2 Insight into the Practice of Learning - Pedagogy

Critical to getting students to observe their own learning practice, is that they understand the difference between procedures/plans and practices, the former are generic with respect to the latter; procedures are safe ways of realizing a task in general, practices are instances of the realization of the task in specific circumstances. This can be explained using an extended version of the planning and understanding model [17] illustrated in Fig. 2.

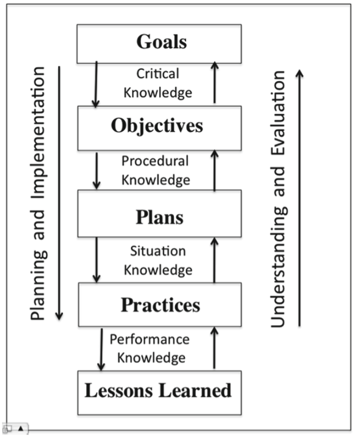

Fig. 2. A model of planning and understanding (after [17])

The four types of knowledge illustrated in Fig. 2 are all deployed by managers in the realization of their activity qua manager, but the management student can apply the very same categories to learning about learning. Starting with generic Learning Goals, the student with the help of the teacher must specify Learning Objectives and come up with a Learning Plan, corresponding to a syllabus for a given course. Each student will go about learning in a different way, or the same student in different circumstances will go about learning in different ways and it is this actual Learning Practice that is the source of lessons learned in learning. Aspectual differences in practice correspond to performance indicators that can be used to guide the selection of best learning practices both for the individual and across communities of learning practice when students work together in teams.

3 Methodology

The pilot research project, in the domain of Education, presented here was undertaken in the 2018/2019 academic year in a management class. The literature reviewed in the preceding section led us to prepare a self-evaluation table for the students and on a weekly basis discuss with students the progress of their insight into the subject matter on the one hand and into their learning practice on the other hand. We asked what strategies support students' practice-based learning of content in the classroom? And what structures should the learning content take to support students' insight into content in the classroom?

In this section, we present the peer-controlled self-evaluation tool and the pedagogical and didactical issues associated with using the table in the classroom.

Table 1. A Learning Table – a Peer-controlled Self-evaluation Form

Course:	Purchasing and Supply Chain Management (PSCM)	Student:			
		I don't know	I know with help	I know	I really know, I understand
Learning Objective LO 1	The role of purchasing in the value chain				
Learning Objective LO 1.1	The role and importance of the purchasing and supply chain functions in the value chain				
Learning Objective LO 1.2	The difference between concepts such as ordering, purchasing, buying				
Learning Objective LO 1.3	New developments in the practice of PSCM				
Learning Objective LO 2	Industrial Buying Behavior: Decision-making in purchasing				
Learning Objective LO 2.1	The major differences between organizational and consumer buying				
Learning Objective LO 2.2	The key elements of the purchasing process				
Learning Objective LO 2.3	How to model organizational buying behavior				

The rows in Table 1 show the learning objectives, the four columns show the evaluation of the level of understanding, a simple four-point scale is chosen: "I don't know," "I know with help," "I know," and "I understand". Students assess on an ongoing basis (each week) how they are doing in terms of understanding the material. Understanding is measured as the ability to transmit their knowledge of a given learning objective to a peer; hence it is peer-controlled, but it is the students themselves that ascribe their level of understanding.

The impact of an improvement on a particular learning objective is immediately seen by the student. The form is structured in line with the syllabus of the course. The hierarchical structure is a matter for the teacher. In textbook-based learning this may correspond to chapters and in-chapter learning objectives chosen by the teacher.

The effect of over-confidence bias is reduced by peer-control in which students explain the subject matter under question to their peers and reassess their level of understanding in light of feedback.

Each student prepared an updated version for each class and in the last class the teacher validated the evaluation, checking the sincerity of the students.

A sincerity check consisted in the teacher asking the students questions to allow them to demonstrate their knowledge. If the student had marked "know with help" then they could use their notes and textbook as help in answering the question. If they marked "know" tout court then they had to answer without help, for example by giving the definition, without being required to explain. If they marked "understand" they had to be able to answer in a way that demonstrated their understanding, through reformulation or giving an example of use.

Section 4 presents the results of the pilot research project.

4 Results: Pilot Study on Learning in a Business School

In this section, we present the qualitative results of the pilot research project.

Once students started measuring their insight, they spontaneously sought to improve their performance which brings them to the question of their learning practice. Even though the students had been introduced to the literature on contextual graphs, often they just discussed among themselves their learning practices. In their discussions they demonstrated clear awareness of the difference between procedure and practice. They effectively formed spontaneous communities of learning practice with their peers.

Students quickly discovered the link between their learning practices and learning outcomes. They discovered how peer control helped them to improve their self-evaluation and this led to improved confidence in their judgment. They understood how certain learning practices worked better for them and that insights into their own learning practice were levers to improve their learning performance.

Throughout the semester students were able to observe their progress and gained confidence in their ability to pass the final exam. And indeed, no student failed the class, as would be expected from Bloom's findings that having enough time was the critical determinant of success in learning across learning abilities [2].

A peer-controlled self-evaluation form was tested in different classes. Students learned how to discriminate different levels of understanding and to detect changes in their level of understanding (insight) at multiple levels of consolidation. This observable chunking of knowledge [15] led students to have a more sophisticated approach to questions and answers. By the end of the course, students were able to spontaneously reconstruct the hierarchy of learning objectives, indeed some students developed their own personalized structure for the subject body of knowledge.

Potential benefits to schools that encourage students to engage in peer-controlled self-evaluation in an environment of shared learning practice include:

(1) better use of class-time. The implementation of flipped classes and blended learning for example is particularly easy in an environment where students are curious to learn. And,
(2) more effective communities of teaching practice. When students share learning practice, teachers discover new opportunities to improve their pedagogy, and when students share their ways of chunking knowledge, teachers discover new opportunities to improve their didactics.

Communities of teaching practice enable teachers develop better forms of peer-controlled self-evaluation for their students. Researchers, strong in didactics, contribute insights into the structure of the body of knowledge that help students gain theoretical insights in the subject. Practitioners, strong in pedagogy, contribute insights in situations of application that help students gain practical insights. Peer-controlled self-evaluation applies equally to teaching practice and the authors have found themselves engaging in a spontaneous community of teaching practice with their colleagues. Better contextualization leads to better evaluation of learning (and teaching). And better evaluation leads to better learning (and teaching).

5 Conclusions and Future Work

The feasibility of the characterization of learning in levels of understanding for different chunks of knowledge was confirmed in a pilot study in a management class. This opens up a new area of research and practice in the field of education, insofar as more attention in the classroom can be focused on understanding rather than memorizing. There are also implications for evaluation of learning outcomes without recourse to traditional grading as the simple ad hoc learning table presented in this paper shows.

The conditions for the acceptability of self-evaluation by students were determined; as long as students could go back and redo their self-evaluation as the course continued, they were enthusiastic about the approach. The results of the pilot study were sufficiently encouraging to begin the multi-class research project. Future research on a cross-case study of three business schools involved in the same learning activity (management) is under preparation. An insight-driven context-aware intelligent tutoring system could be an interesting development out of this work.

The feasibility of using peer-controlled self-evaluation to measure insight in the classroom and the necessity for a control of sincerity by the teacher raises analogous questions with respect to insight in management and insight in technology transfer. In the field of management, in particular for the activity of situation assessment, an important concern of the board of directors is to assess the capacity of management to really understand the situation in which the company finds itself. And in technology transfer, problems of adaptation of technology can be avoided by better understanding of the situation into which the technology is being implanted.

The research presented in this paper addressed insight in the classroom, and the authors suggest both insight in management and insight in technology transfer as avenues of future research. Given the great strides in AI and Cognitive Science since the first Context conference twenty years ago, perhaps the next twenty years will see neural imaging of insight being leveraged insightfully in all of these areas.

References

1. Arduin, P.-E., Grundstein, M., Rosenthal-Sabroux, C.: Information and Knowledge Systems. Wiley-ISTE, London (2015)
2. Bloom, B.S.: Time and learning. Am. Psychol. **29**(9), 682–688 (1974)
3. Bowden, E.M., Jung-Beeman, F.J., Kounios, J.: New approaches to demystifying insight. Trends Cogn. Sci. **9**(7), 322–328 (2005)
4. Brézillon, P., Pomerol, J.-C.: Is Context Tacit Knowledge? CONTEXT (2001)
5. Brézillon, P., Pomerol, J.-C.: Contextual Knowledge and Proceduralized Context, AAAI Technical Report WS-99-14 (1999)
6. Brézillon, P.: Task-realization models in contextual graphs. In: Dey, A., Kokinov, B., Leake, D., Turner, R. (eds.) CONTEXT 2005. LNCS (LNAI), vol. 3554, pp. 55–68. Springer, Heidelberg (2005). https://doi.org/10.1007/11508373_5

7. Edmonds, B.: The pragmatic roots of context. In: Bouquet, P., Benerecetti, M., Serafini, L., Brézillon, P., Castellani, F. (eds.) CONTEXT 1999. LNCS (LNAI), vol. 1688, pp. 119–132. Springer, Heidelberg (1999). https://doi.org/10.1007/3-540-48315-2_10
8. Hegarty, J., Brézillon, P., Adam, F.: The role of context in practice-based organizational learning and performance improvement. In: Brézillon, P., Blackburn, P., Dapoigny, R. (eds.) CONTEXT 2013. LNCS (LNAI), vol. 8175, pp. 59–72. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-40972-1_5
9. Hegarty, G.J.: Leveraging lessons learned in organizations through implementing practice-based organizational learning and performance improvement - An opportunity for context-based intelligent assistant support (CIAS). Ph.D. Thesis, University College Cork (2013)
10. Lemoigne, J.-L.: La modélisation des systèmes complexes, Dunod (1999)
11. Lonergan, B.: Insight - a study of human understanding. In: Crowe, F., Doran, R. (eds.) Collected Works of Bernard Lonergan, University of Toronto Press, Canada (1988)
12. Lonergan, B.: Cognitional structure. In: Crowe, F., Doran, R. (eds.) Collected Works of Bernard Lonergan, vol. 3, pp. 205–221. University of Toronto Press, Canada (1988)
13. Marples, R. (ed.): The Aims of Education. Routledge International Studies in the Philosophy of Eduction, London (1999)
14. Maubrey, R.: Etude du rôle des interactions socio-langagières dans la transmission de savoirs entre adultes: la situation de formation trans-culturelle en agriculture: cas de stagiaires africains, doctoral thesis, University of Paris 8, Vincennes Saint-Denis (1999)
15. Miller, G.: The magical number seven, plus or minus two: some limits on our capacity for processing information. Psychol. Rev. **63**, 81–97 (1956)
16. Sternberg, R.J., Davidson, J.E. (eds.): The nature of insight. MIT Press, Cambridge (1995)
17. Wilensky, R.: Planning and Understanding. Addison-Wesley, Canada (1983)

Evaluation of Computer-Tailored Motivational Messaging in a Health Promotion Context

Jens E. d'Hondt(✉), Raoul C. Y. Nuijten, and Pieter M. E. Van Gorp

Eindhoven Unitversity of Technology, Eindhoven 5612 AZ, The Netherlands
j.e.dhondt@student.tue.nl, {r.c.y.nuijten,p.m.e.v.gorp}@tue.nl

Abstract. Persuasive messages have recently been shown to be more effective when tailored to the personality and preferences of the recipient. However, much of the literature on adaptive persuasion has evaluated the effectiveness of persuasive attempts by the direct reactions to those attempts instead of changes on the longer term (e.g. lifestyle changes). Results of this study suggest that adaptive persuasion improves attitudes towards persuasive attempts, but does not necessarily cause a change in longer term behavior. This was found through a randomized controlled trial evaluating the implementation an adaptive persuasive system in a health promotion intervention. This article provides a detailed description of this evaluation and encourages the research community to (1) become more skeptical towards the longer term effectiveness of adaptive persuasive techniques and (2) design more explicitly for longer term changes in behavior.

Keywords: Persuasion · Individual differences · Health promotion

1 Introduction

Disruptive developments in computing power and data collection throughout the last decade have allowed suppliers to obtain and process more information on customer interactions, both relating to the customer itself and the circumstances in which the interactions take place. This information gain does not only aid in designing products better suited to customer needs, it also allows for tailoring promotional efforts to individual customers [1]. Currently, research on personalized promotion has focused most on creating artificial agents and feeding them with information on context variables like individual's valuation of product attributes and prior purchase behavior in order to improve their decisions on product recommendations (also known as recommender systems). However, integration of information on other variables such as price elasticity or personality traits is far less discussed in literature, and is scarcely implemented in situations where effective product promotion is fundamental, such as marketing and e-commerce [11]. The idea of using information on people's personality traits in order to optimize promotional efforts arose from the fact that it was recently found that there exist individual differences in susceptibility to certain principles of persuasion and therefore also in responses to different persuasion strategies [10].

G. Bella and P. Bouquet (Eds.): CONTEXT 2019, LNAI 11939, pp. 120–133, 2019.
https://doi.org/10.1007/978-3-030-34974-5_11

These differences may entail improvements in persuasive approaches when tailored correctly to the recipient, but may also encompass negative results when striking the wrong chord [10].

Tailoring persuasion strategies to the personality of the recipient (adaptive persuasion) has shown to be effective in a variety of contexts, e.g. marketing [9], medication [3] and health care [15]. However, what is remarkable about these studies is that they often only measure the effectiveness of persuasive approaches on short-term behavior and attitudes (e.g. direct feedback or click-through rates [9]) without measuring their effect on long-term behavior. Consequently, while the positive impact of adaptive persuasion on the direct attitudes towards persuasive approached has been shown, it's ability to foster behavioral change still remains unknown. This is, however, important in case adaptive persuasion is desired to be used in contexts like health promotion and medication, where habits are attempted to be changed in order to ensure medication adherence. We have reason to question this ability as a prominent framework on behavioral change, the COM-B framework by Michie et al. [14] suggests that there are three antecedents of behavioral change; Capability, Opportunity and Motivation, and interventions aimed at inducing behavioral change ought to provide support which cultivates compliance to all conditions. This implies that informing artificial agents on the personality of the recipient may help to motivate a person to act, but it may fail to foster behavioral change in case no contextual information is provided regarding the person's capability or opportunity to do so. Therefore, the current study evaluates if adaptive persuasion does not only improve attitudes towards persuasive attempts, but also has the potential to induce behavioral change. This was done by conducting a randomized controlled trial evaluating the implementation of an adaptive persuasive messaging system (APMS) in a health promotion intervention for employees and students of the University of Technology in Eindhoven. In this paper, Sect. 2 will focus on the design of the APMS and methods used to evaluate the system. Section 3 will present the results of the evaluation, subsequently discussing these results and drawing conclusions in the final section.

2 Design Process of an APMS

Currently, many of the existing health promotion interventions include some combination of a health tracking device and a mHealth application to log activities and review progress. In this context, the APMS could be integrated into the back-end of such applications, using messages to 'nudge' the user into performing more activities and to remind him/her of certain developments on the platform. Information on in-app behavior and direct feedback could then be used to personalize nudges in order to increase their effectiveness, resulting in greater engagement with the application, and ultimately greater adoption of healthy behaviors. In the case of this study, the APMS was known to be implemented in a mHealth application called GameBus[1]. GameBus was built according to the

[1] https://www.gamebus.eu/.

principles of gamification and provides an environment to host and participate in digital competitions in which points can be earned by performing activities related to living a healthy lifestyle. Since GameBus did not yet contain any form of active messaging towards its users, the platform could be naturally extended with the purpose of studying persuasive messaging personalization. Figure 3 provides a very simplified flow diagram representing the overall user journey of health promotion applications like GameBus, highlighting the aimed position of the APMS in that journey. Similarly, Fig. 2 provides a graphical overview of the design of the study.

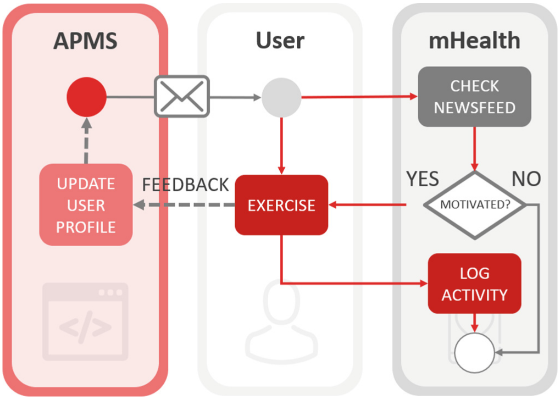

Fig. 1. Visual representation of general user journey when using mHealth applications

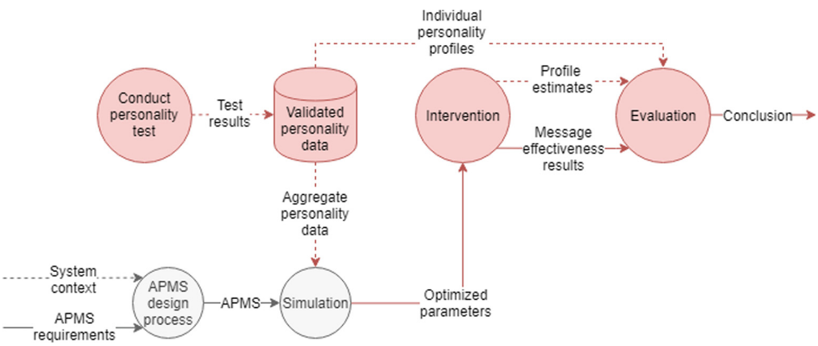

Fig. 2. Graphical representation of the study design

2.1 System Requirements

Built on the requirements of *adaptive persuasive technologies* as presented by Kaptein and Van Halteren [12], the authors suggest five requirement for the

APMS to be capable of effectively adapting to individual differences in responses of recipients. Firstly, the system should be able to identify individual users and maintain a specific user profile indicating the probability of success of different influence strategies. Secondly, it should be able to frame messages to be congruent with a specific influence strategy. Thirdly, at time of constructing the message, it should have a clear protocol of choosing the influence strategy used in the message. Fourthly, when the message has been sent, predetermined success measure(s) are needed to assess the effectiveness of the approach. And lastly, after the effectiveness of an approach has been assessed, the system needs specific learning rules in order to update the user profile and optimize long-term message effectiveness.

2.2 System Design

User Profiling. Prior studies on adaptive persuasion report profiling users on different psychological traits such as Need for Cognition [7], susceptibility to different influence principles (e.g. those presented by Cialdini [2]) or the Big Five personality dimensions (which proposes that a person's personality can be modeled by his/her adherence to each of five different traits; extraversion, agreeableness, conscientiousness, neuroticism and openness) [5]. Profiling based on the latter is deemed most appropriate in this context compared to profiling based on Cialdini's principles as we look to induce long-term behavioral change, which reduces the relevance of principles encouraging immediate action like Cialdini's scarcity principle. In addition, there is a greater body of research discussing the variation in motivational systems reflected by each of the Big Five personality traits [9] compared to Need for Cognition. This helps to ensure unambiguous message framing in later stages of the design process. Therefore, users are profiled based on their personality expressed by their estimated adherence to each the Big Five personality traits. In line with this decision, individual user profiles are modeled as a collection of Beta-Binomial models [21] each representing the probability distribution of the value of p_m, indicating the probability that the recipient's response to a message framed congruent with trait m is positive. Note that this implies we are modeling the probability distribution of the value of p_m, i.e. the probability of a probability. This method is suitable here as any persuasive interaction can be seen as a binomial random process, for which prior information (i.e. responses to previous messages) is included to iteratively enhance estimations for the probability of success p_m [12]. See Fig. 3 for an example profile. In the figure, every curve corresponds to the probability density function of p_m of a message framed congruent with trait m (being one of the Big 5 personality traits).

Message Framing. In order to preserve significant variation in message content, five distinct motivational message elements were drafted for each of the five personality traits. When constructing an email, a message element was chosen and inserted into the body of an email. Besides the persuasive element, emails

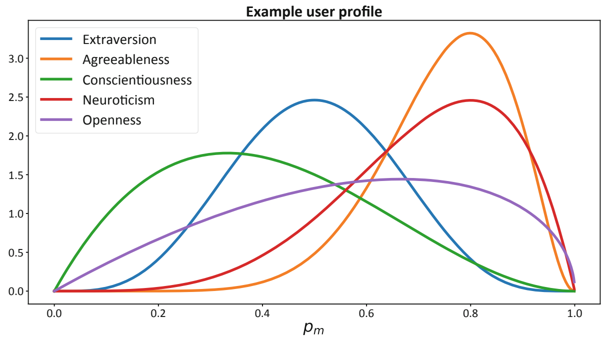

Fig. 3. User profile example

included a personalized preamble and the promotion of a random GameBus activity[2].

Influence Strategy Choice. The algorithm used to determine the optimal persuasion strategy of a given message was derived from the decision rules presented in a study by Kaptein and Parvinen [11]. In the study, they use the concept of *Randomized Probability matching* (RPM) as introduced by Scott [18] to deal with the *explore/exploit dilemma*, and James-Stein shrinkage [20] of individual user profiles to an average profile of the whole sample population to reduce the uncertainty on user profiles of individuals of whom little data has yet been obtained. We refer to the paper of Kaptein and Parvinen [11] for a basic explanation of the algorithm. Both solutions were used in the current study with the modification that α- and β-values corresponding to the Beta-Binomial models (Beta(α, β)) in user profiles were shrunk using the following set of equations. Note that Eq. 3 was not derived from prior research on RPM and James Stein Shrinkage, but was added to ensure a decrease in the amount of shrinkage over time;

$$\alpha^*_{t,i} = \overline{\alpha}_t + c_i(\alpha_{t,i} - \overline{\alpha}_t) \tag{1}$$

$$\beta^*_{t,i} = \overline{\beta}_t + c_i(\beta_{t,i} - \overline{\beta}_t) \tag{2}$$

$$c_i = 1 - e^{-\lambda * n_i} \tag{3}$$

where:

$\alpha_{t,i} = \alpha$ in Beta(α, β) for personality t, participant i
$\overline{\alpha}_t = \alpha$ in Beta(α, β) for personality t in average profile
$\beta_{t,i} = \beta$ in Beta(α, β) for personality t, participant i
$\overline{\beta}_t = \beta$ in Beta(α, β) for personality t in average profile
n_i = number of profile updates for participant i
λ = profile maturation constant

[2] See https://doi.org/10.6084/m9.figshare.8289044.v1 for the list of persuasive elements as used by the APMS.

This implies that upon strategy selection a profile is used which essentially is a mix between an average profile derived from all reported message feedback and the individual profile of the prospective recipient. The amount of information 'borrowed' from the average profile decreases as a function of the number of profile updates of a participant and the profile maturation constant λ. Following the principles of RPM, strategies are eventually chosen by taking draws from each of the Beta-Binomial distributions in the shrunk profile, choosing the strategy associated with the highest draw.

Success Measure. Message effectiveness was evaluated using two sources of feedback from the recipient. Firstly, users were able to provide direct feedback on the message through a feedback form included in the email. In order to capture additional (contextual) information, participants could not only indicate if the message content appealed to them, but also why (not). Hence, when asked whether a message suited their preferences, participants could either express their satisfaction with the message (and confirm whether they thought they would soon perform the activity, or not), or express their dissatisfaction with one or multiple (contextual) aspects. In case participants indicated they disliked the message due to it's timing or frequency, they were automatically offered the opportunity to update their message preferences. Only message feedback which includes specified sentiment towards the used influence strategy triggered a model update corresponding to the influence strategy used in the message. Secondly, message effectiveness was indirectly measured using information on the participant's performed activities. To illustrate, upon logging an activity into GameBus, participants were requested to indicate what triggered them into performing that activity. A message was deemed effective in case it's promoted activity was logged within 3 days after the message was sent, indicating it was triggered by a message.

Learning Rule. Consequent to a persuasive approach's effectiveness being reported, the parameters of the recipient's corresponding Beta-Binomial distribution were updated in the following way:

$$\begin{aligned} \alpha_{ti} &:= \alpha_{ti} + a \\ \beta_{ti} &:= \beta_{ti} + (1 - a) \\ \text{with } a &= \begin{Bmatrix} 1 \text{ if feedback} == \text{positive} \\ 0 \text{ if feedback} == \text{negative} \end{Bmatrix} \end{aligned} \tag{4}$$

As the point estimate $\hat{\mu}_t$ of the effectiveness of a message framed congruent with personality t follows $\hat{\mu}_t = \frac{\alpha_t}{\alpha_t+\beta_t}$ with variance $\hat{\sigma}_t^2 = \frac{\alpha_t\beta_t(\alpha_t+\beta_t+1)}{(\alpha_t+\beta_t)^2(\alpha_t+\beta_t+1)}$, the estimated effectiveness was increased/decreased depending on a while the variance of the estimate decreases each time new information is gained through feedback.

System Overview. In the discussed example, the usages of contemporary database, web and messaging middleware allowed the APMS to be deployed on a separate network node other than the system on which activity registrations were taking place (in this case GameBus). Thanks to this approach, the APMS is fully modular and should be able to be build on top of other platforms in a relatively easy manner. Additionally, this also allows the full source code to be shared for further use - see https://github.com/JdHondt/APMS.git.

2.3 Trial Simulation

In order to validate the system's functionalities and to allow for hyper-parameter tweaking (i.e. the profile maturation constant λ), the trial was simulated with conditions similar to the randomized controlled trial through which the system was ought to be evaluated. 150 personality profiles were randomly generated based on the means and standard deviations of personality profiles obtained via a personality test (following the Mini-IPIP design by Donnelan et al. (2006) [4]) included in an intake survey conducted on participants of the randomized controlled trial ($n = 36$). Note that the results of the personality test as part of the intake survey were solely used for evaluation purposes and simulation, and did not serve as direct input for the algorithm's profile development. Reactions to persuasive approaches were simulated by taking a draw from a Bernoulli-distribution with $p = p_m$, m being the trait towards which the message was framed. To evaluate the impact of adaptive persuasion, the APMS updated the profile estimates of only half of the generated profiles, effectively simulating two treatment groups. The profile maturation constant λ and Big Five trait value variance σ_p of the generated profiles (i.e. individual differences in personality) were varied across simulations in order to analyze the ideal value for λ given the length of the intervention and different values of σ_p. 20 simulations were run per λ, σ_p combination to reduce the variability in results.

Simulation Results. The results as presented in Fig. 4 verify the effectiveness of RPM in increasing the expected 'performance' of messages. This can be concluded by both the positive differences for all conditions in the left figure and the increasing performance over time of the adaptive condition in the right figure. Furthermore, the λ optimization graph clearly highlight the positive relationship between the σ_p and λ. This relationship is logical as high variability in individual profiles diminishes the usefulness of shrinkage, profiting from earlier influences of individual profile estimates. As the result of the personality survey indicated a personality variance of around $\sigma_p = 0.08$, the authors decided to go with a λ of 0.13, implying strategy selection will be mainly based on the average profile considering the short duration of the intervention.

2.4 Evaluation Methodology - Health Promotion Intervention

The APMS was evaluated by implementing it in a health promotion intervention similar to that of Nuijten et al. [16], which included a 6-week digital health com-

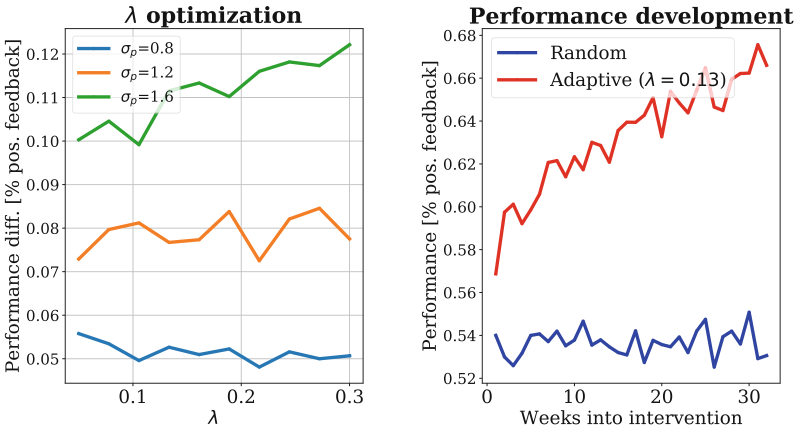

Fig. 4. Simulation results

petition among a sample of students and employees of the Eindhoven University of Technology ($N = 149$). Both an individual and a team-based competition were held among different university departments with prizes being awarded to the entities that ranked best of their competition at the end of the trial. Additionally, weekly rewards were raffled among those who reached a weekly point target. Points were earned by completing photo-based challenges involving engagement in some form of physical activity and taking a photo (so called *FitPic*) as proof of completion[3]. The rationale behind the selection of competition elements discussed above was to provide significant stimuli for all Big Five personality types and their characteristics. To illustrate, both the competition reward structure as GameBus' news feed functionality served as stimuli for participants scoring high on the Extraversion dimension as they are especially sensitive to rewards and social attention [19] or Neuroticism as it is associated with a strong sensitivity to uncertainty and threats [8]. The concept of a team-based competition was introduced to support agreeable individuals as they value communal goals and interpersonal harmony [6]. Similarly, an individual competition with only one winner served as a stimulant for conscientious individuals who value achievement [17]. Lastly, as open individuals value creativity and intellectual stimulation [13], challenges were framed in a flexible manner that offered opportunity for creativity and one's own interpretation.

Conditions. To evaluate the effectiveness of personalized persuasive messaging as opposed to using arbitrary persuasion strategies in messages, participants were distributed over two treatment groups:

1. *Control:* Random choice of persuasion strategy upon messages creation. Each strategy had equal chance of being chosen.

[3] See https://doi.org/10.6084/m9.figshare.8288882.v1 for the list of challenges.

2. *Treatment:* Persuasion strategy used in messages chosen based on both feedback received by the prospective recipient on previous messages as well as feedback received by all participants (including participants assigned to control) on messages sent to them.

This between-subjects design was chosen over a within-subjects design to compensate for a learning curve related to GameBus which appeared from previous studies using the application. The first intervention week served as a learning period for participants to get comfortable with the GameBus platform and elements of the competition. In this week both control and treatment received randomized messages and no message feedback was gathered. Adaptive messaging was initialized starting from the second week.

2.5 Statistical Analysis

To evaluate the effectiveness of tailored messages over messages with a random persuasion strategy, multiple generalized linear (mixed effect) models were fitted on the data for different effectiveness metrics and compared to each other using χ^2 tests. Additionally, in case a model reported a significant improvement towards a simpler model, it's fixed effects were analyzed to see what conclusions could be made on the relationship between the fixed effects and target variables. As for the effectiveness metrics, the following variables were chosen:

- *Message Feedback (MF):* Sum of feedback values f_i.
- *Performed Activities (PA):* Number of activities logged in GameBus.
- *Message Success (MS):* Percentage of messages causing an activity being performed[4].

All target variables were measured on an individual level and aggregated by week. f_i was computed following;

$$f_i = \left\{ \begin{array}{rl} 1 & \text{if feedback} == \text{positive} \\ 0 & \text{if feedback} == \text{negative (not content-related)} \\ -1 & \text{if feedback} == \text{negative (content-related)} \end{array} \right\} \tag{5}$$

As for the models, all models are extensions on a so called "null" model. This model describes an overall intercept on the target variable, with the addition of individual-level intercepts as a random effect. Fitting this model basically tests for individual differences in the values for the effectiveness metrics, regardless of the individual's treatment group. All model comparisons were done towards the null model. The models were fitted using Poisson error distributions for both PA and MS target variables, and Gaussian error distributions for target variable MF. Models are presented in Table 1. Both *group* features in Model C and D

[4] Activity-message causation in MS was determined the same way message effectiveness was measured using information on the performed activities of participants; by flagging messages of which the promoted activity was performed within 3 days, indicating a message as it's trigger.

are binary variables indicating a participant's belonging to the treatment group. The β-values correspond to β_0 (intercept) for Model A and the weights of time, group and $time \times group$ for Models B, C and D, respectively.

3 Results

At the start of the intervention 149 people were registered as eligible participants by direct registration or indirectly as they were ex-participants of Nuijten et al. [16] and did not opt-out for this study. These participants were randomized to either the intervention ($n_i = 75$) or control group ($n_c = 74$). At the end of the intervention, 16 participants had actively withdrawn their participation and 111 were excluded from the data analysis as they did not comply to the requirements of completing the intake survey and providing feedback on at least one message. Of the remaining participants ($n_i = 12$, $n_c = 10$), the majority was male (63.64%) with a mean age of 32.2 ± 6.49 and a baseline BMI of $23.6 \pm 2.59\,\text{kg/m}^2$.

Table 1. Table showing fixed effects and model comparisons for different target variables

	MF		PA		MS	
	χ^2	β	χ^2	β	χ^2	β
Model A: null model	-	0.33^{**}	-	-0.55	-	-1.10
Model B: A + time	2.23	0.07	0.29	-0.03	0	0
Model C: A + group	8.69^{**}	0.54^{**}	0.53	-0.06	0	0
Model D: A + time × group	11.5^{**}	0.07	2.55	0.13	0.35	-0.12

$^{*}p < 0.05$
$^{**}p < 0.01$
$^{***}p < 0.001$

3.1 System Evaluation

When analyzing the results shown in Table 1, we see a significant positive intercept for model A on MF, indicating a general tendency towards positive message evaluations of participants (regardless of treatment group). Moreover, significant increases in model fit are reported for models C and D on MF, with a significant positive fixed effect for the *group* variable in Model C, but an insignificant β-value for the interaction term on *time* and *group* in Model D. All other model fits are reported to be insignificant. The insignificant β-value for $time \times group$ indicates that there does not seem to be a significant difference in MF development over time between groups, which one would not expect considering the fact that user profile estimates (in theory) should improve over time, resulting increased MF values as illustrated in the simulation results. In view of this discrepancy, Mann-Whitney tests were performed for both week 1 and 6 of the intervention, indicating an insignificant difference in MF in week 1 ($p = 0.119$), but a significant difference in week 6 ($p = 0.035$).

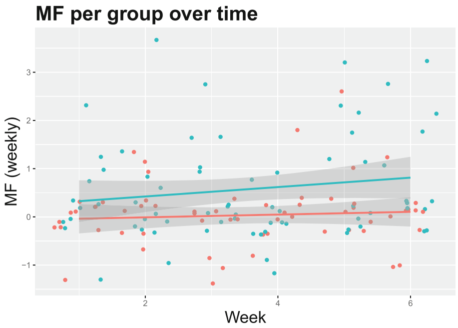

Fig. 5. Overview of MF over time

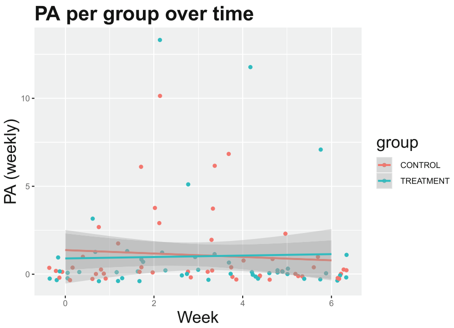

Fig. 6. Overview of PA over time

4 Discussion and Conclusions

We conclude that participants receiving personalized messages evaluate messages more positively than those receiving messages framed towards a random personality. These results are in line with those reported in prior research. The insignificant difference in MF in week 1 opposed to the significant difference in week 6 hints towards a between-group MF development difference, but does not support the corresponding hypothesis which might be due to the limited duration of the intervention. From the insignificant fits for models on both PA and MS we can conclude that despite their superior evaluation of messages, participants in the treatment group are not reported to perform more activities. The poor fits and effects of models with MS were mainly a consequence of the fact that only 9 activities were reported indicating a message as activity trigger.

In this study, we presented an extension on the requirements of adaptive persuasive systems which designers can use to create and implement artificial agents which adapt to individual differences in the preferences of users. An important purpose of this study was to evaluate if adaptive persuasive systems have the ability to induce behavioral change in a health promotion context, as the authors felt that this was insufficiently studied by prior research. The reported results suggest that although these systems may induce elevated attitudes towards persuasive approaches, they do not necessarily cause a change in health behavior. This highlights the multidimensionality in the antecedents of behavioral change, and should encourage researchers and enterprises to create artificial agents which are not only sensitive to context-variables which influence people's motivation to act, but also to those which impact people's capability and opportunity to adopt the desired behavior. In essence, while guidelines are offered to implement adaptive persuasion, we emphasize that effective persuasion alone will not be enough to induce behavioral change in a health promotion context, and sufficient regard has to go out to other support elements that increase capability and opportunity to act.

4.1 Limitations

There are several limitations to this study. One includes it's small sample size, which resulted from high drop-out rates and low participant engagement with the study. According to a post-intervention survey ($n = 12$), the main causes of this inactivity included a substantial barrier to log daily activities, as this was ought to be done manually and annoyance with the email messages, their frequency was felt to be too high. A second limitation involves the Influence Strategy Choice method used in the APMS. Here, considering time constraints, RPM with James-Stein shrinkage was used due to its simplicity and fast implementation. However, while the main attractiveness of this method comes from its short computation time, other more computational intensive methods like Gittins Indices or Multilevel Hierarchical models are estimated to have higher predictive performance. These models could be more suited for this problem as message effectiveness predictions are not required to be computed real-time. Lastly, a problem the authors dealt with during design, was the difficulty in linking activity to message, effectively identifying which approach was successful. This problem is similar to the credit assignment problem discussed in research on Reinforcement Learning, for which currently no definitive solution is known. Unfortunately, this has limited the system's ability to improve user profiles based on direct actions, consequently having to base profiles mainly on message feedback.

4.2 Future Research

In light of the discrepancy between positive evaluation of persuasive approaches and reported behavior found in this study, future research should be conducted evaluating the effectiveness of more complete persuasive systems that combine personalization in terms of promoted product or service with personalization of persuasion strategy to assess if a synergy between methods does bring about behavioral change. Such a system could be implemented in an enlarged replication of the trial discussed this study, with the addition of a more effective recruitment strategy to ensure larger sample sizes and increased participant engagement. Lastly, the current evaluation could be expanded by including an analysis comparing the individual profile estimates which resulted from the intervention with the profiles derived from the personality test conducted pre-intervention (indicated by the top dotted arrow in Fig. 2). This could not only provide insights into the system's profile prediction accuracy, it could also shed light on potential differences between reported personality and performed behavior.

Acknowledgements. This work is part of the research program 'Gamification for Overweight Prevention and Active Lifestyle' (443001101), which is partly financed by the Netherlands Organisation for Health Research and Development (ZonMw). The authors also thank Kalliopi Zervanou for feedback on an early version of this work

References

1. Allenby, G.M., Rossi, P.E.: Marketing models of consumer heterogene- Ity. J. Econometrics **89**, 57–78 (1999)
2. Cialdini, R., Trost, M., Newsom, J.: The development of a valid mea- sure and discovery of surprising behavioural implications. J. Pers. Soc. Psychol. **69**(2), 318–328 (1995)
3. Dijkstra, A.: Working mechanisms of computer-tailored health education: evidence from smoking cessation. Health Educ. Res. **20**(5), 527–539 (2005). https://doi.org/10.1093/her/cyh014. ISSN: 02681153
4. Brent Donnellan, M., et al.: The Mini-IPIP scales: tiny-yet-effective mea- sures of the Big Five factors of personality. Psychol. Assess. **18**(2), 192–203 (2006). https://doi.org/10.1037/1040-3590.18.2.192. ISSN: 10403590
5. Goldberg, L.R.: An alternative "Description of Personality": the big- five factor structure. J. Pers. Soc. Psychol. **59**(6), 1216–1229 (1990)
6. Graziano, W.G., Eisenberg, N.: Agreeableness: a Dimension of Personality (1997). http://cachescan.bcub.ro/e-book/E1/580591/795-870.pdf
7. Haugtvedt, C.P., Petty, R.E., Cacioppo, J.T.: Need for cognition and advertising: understanding the role of personality variables in consumer behavior. J. Consum. Psychol. **1**(3), 239–260 (1992). https://doi.org/10.1016/S1057-7408(08)80038-1. ISSN: 10577408. 1
8. Hirsh, J.B., Inzlicht, M.: The devil you know: neuroticism predicts neural response to uncertainty. Psychol. Sci. **19**(10), 962–967 (2008). https://doi.org/10.1111/j.1467-9280.2008.02183. ISSN: 09567976
9. Hirsh, J.B., Kang, S.K., Bodenhausen, G.V.: Personalized persuasion: tailoring persuasive appeals to recipients personality traits. Psychol. Sci. **23**(6), 578–581 (2012). https://doi.org/10.1177/0956797611436349. ISSN: 14679280
10. Kaptein, M., Eckles, D.: Heterogeneity in the effects of online persuasion. J. Interact. Mark. **26**(3), 176–188 (2012). https://doi.org/10.1016/j.intmar.2012.02.002. ISSN: 10949968
11. Kaptein, M., Parvinen, P.: Advancing e-commerce personalization: process framework and case stud. Int. J. Electron. Commer. **19**(3), 7–33 (2015). https://doi.org/10.1080/10864415.2015.1000216. ISSN: 15579301
12. Kaptein, M., Van Halteren, A.: Adaptive persuasive messaging to increase service retention: using persuasion profiles to increase the effectiveness of email reminders. Pers. Ubiquitous Comput. **17**(6), 1173–1185 (2013). https://doi.org/10.1007/s00779-012-0585-3. ISSN: 16174909
13. McCrae, R.R., Costa, P.T., Costa Jr., P.T.: Conceptions and correlates of openness to experience. In: Handbook of Personality Psychology May (1997), pp. 825–847. ISSN: 14222795. https://doi.org/10.1016/B978-012134645-4/50032-9
14. Michie, S., van Stralen, M.M., West, R.: The behaviour change wheel: a new method for characterising and designing behaviour change interventions. Implementation Sci. **6**(1), 42 (2011). https://doi.org/10.1186/1748-5908-6-42. ISSN: 17485908
15. Morris, M., Guilak, F.: Mobile heart health : project high-light. IEEE Pervasive Comput. 57–61 (2009). https://doi.org/10.1145/1280120.1280170
16. Nuijten, R.C.Y., et al.: Evaluation of the impact of extrinsic rewards on user engagement in a health promotion context (2019)
17. Roberts, B.W., et al.: The structure of conscientiousness: an empirical investigation based on seven major personality questionnaires. Pers.l Psychol. **58**(1), 103–139 (2005). https://doi.org/10.1111/j.1744-6570.2005.00301.x. ISSN: 0031–5826

18. Scott, S.L.: A modern Bayesion look at multi-armed bandit. Appl. Stoch. Models Bus. Ind. **26**, 157–164 (2007). https://doi.org/10.1002/asmb. ISSN: 1524–1904
19. Shao, L., et al.: Cross-cultural evidence for the fundamental features of extraversion. J. Pers. Soc. Psychol. **79**(3), 452–468 (2005). https://doi.org/10.1037/0022-3514.79.3.452. ISSN: 0022–3514
20. Stein, C.: Inadmissibility of the usual estimator for the mean of a multi-variate normal distribution. Proc. Third Berkeley Symp. Math. Stat. Probab. **1**(4), 197–206 (1955). ISSN: 0097-0433
21. Wilcox, R.R.: A Review of the beta-binomial model and its extensions. J. Educ. Stat. **6**(1), 3–32 (1981)

Racist Language, Speaker Responsibility and Hearer Authority

Palle Leth(✉)

Stockholm University, Stockholm, Sweden
palle.leth@philosophy.su.se

Abstract. Actual interpretive practice, whether colloquial or formal, is underdetermined with respect to whether qualifications of utterances as racist are based on the speaker's attitude, communicative intention or the meaning of the utterance and whether in the latter case the racist meaning can be implicit or must be explicit. The focus on the speaker's expression of evaluation and attitude in current theorizing about slurs involves a similar underdetermination. A common problem is that the speaker's responsibility is referred to the speaker's own authority. The interactional model of utterance interpretation which forms the theoretical background of the present paper permits us to elaborate a novel and robust conception of speaker responsibility which situates it within the hearer's authority. The speaker's responsibility depends on the hearer's most reasonable interpretation of the utterance in such a way that the speaker's actual attitude and intention are irrelevant and also whether the meaning was explicitly or only implicitly conveyed.

Keywords: Racism · Slurs · Utterance meaning · Communicative intentions · Contextualism · Semantics · Pragmatics · Accountability

1 Introduction

Some forms of racist language are overt. A speaker may use derogatory terms and explicitly express ideas according to which some people are inferior to other people on account of the colour of their skin. In these cases words seemingly speak for themselves. Other forms of racist language are covert only. Since the decline of racism as a socially accepted doctrine these cases are by far the most frequent. The words used suggest racism only against the background of various contextual factors (accompanying gestures, tone of voice, circumstances of the utterance, background assumptions, historical situation). Whether there is racism or not is not so easily nailed down. When taken to task, the speaker will typically deny having said anything racist. The problem of settling whether an utterance is racist or not is not, on closer inspection, unique with covert forms of racism. Even speakers of overtly racist language may deny having said something racist, alleging that derogatory terms are neutral terms for them or that they were not speaking seriously or that they were speaking unpremeditatedly.

Checking racist language is a prerequisite of democratic societies. Yet the interpretive practice displayed in conversational contexts, social networks and media as well as

G. Bella and P. Bouquet (Eds.): CONTEXT 2019, LNAI 11939, pp. 134–146, 2019.
https://doi.org/10.1007/978-3-030-34974-5_12

in courts is marked by a considerable degree of uncertainty with respect to the speaker's responsibility for racist utterances. It is certainly natural that different standards of assessment are applied in formal and informal contexts respectively, but the variety is actually rather arbitrary. The fundamental question is the same: does the speaker's responsibility fall within the speaker's authority or the hearer's? In other words, is it a matter of the context of production or of the context of reception?

In this paper I set out to state the conditions of holding speakers responsible for racist language. I will first deal with racist utterances in general. The overall question is: what does it take to qualify an utterance as racist? To determine whether it is the speaker with her privileged access to her communicative intention and general attitudes or the hearer with her access to public features and the perceived injury who has authority over the meaning of an allegedly racist utterance is necessary in order to hold speakers accountable, disarm subterfuges and effectively prevent the occurrence of racist language. I will second deal with the particular phenomenon of racial slurs. The overall question is: what makes a slur a slur? According to most philosophical and linguistic accounts of slurs, slurs are offensive because speakers express a negative attitude towards the targets by the evaluative meaning component of slurs. If focus is on the speaker's attitude, the speaker's avoiding liability is not sufficiently prevented, in so far as the speaker has authority concerning her attitude. The alternative account of slurs which I propose focuses on the damage brought about by slurs, which situates the offensiveness of slurs within the hearer's authority.

2 Background

According to most philosophical conceptions of utterance responsibility a speaker is responsible only for the meaning which she explicitly commits herself to. This implies that accountability depends on the speaker's communicative intention as well as the literal meaning of her utterance. Implicitly conveyed meaning, like covert racism, is in principle deniable by the speaker (Soames 2008; Camp 2008). Recently, however, philosophers and linguists have come to recognize that some denials are rather implausible (Lee and Pinker 2010; Asher and Lascarides 2013; Pinker et al. 2008; Sternau et al. 2016). Yet the conviction persists that a speaker can avoid liability for anything which she did not explicitly state (Fricker 2012; Stokke 2018; Camp 2018).

In contrast to philosophical theory, the general tendency in legal practice is to hold the speaker accountable for the meaning which a reasonable hearer would be justified in taking as the meaning conveyed (Burger 1973; Robertson and Nicol 2002; Quinn 2015). However, whether the hearer should take only linguistic or conventional meaning into consideration or may also pay attention to various contextual features such as previous discourse and situational setting is up to the judgment of individual judges. An additional issue is which importance is due to the establishment of the speaker's communicative intention. Here again court decisions vary.

There is thus no consensus, neither in the philosophical literature nor in legal theory or practice (not to speak of less formal contexts), on what speaker responsibility really amounts to. The variation is not accounted for by different standards of precision and evaluation for different kinds of contexts, but depends on individual convictions and preferences.

Slurs have become an increasingly researched topic in philosophy and linguistics in recent years (Hornsby 2001; Williamson 2009; Hom 2010; Camp 2013; DiFranco 2014; Bach 2018). The majority of accounts take slurs to carry a negative evaluation of their referents or targets. Most of the discussion is concerned with whether the evaluative dimension of slurs is conveyed by semantic meaning, conventional implicature, presupposition or associated perspective. Only a few accounts take the offensiveness of slurs not to be explained by some feature of their meaning (Anderson and Lepore 2013a, 2013b; Blakemore 2015; Bolinger 2017; Nunberg 2018).

The theoretical background of the paper is what I call the interactional model of utterance interpretation (see Leth 2019, forthcoming a and b). The basic idea is to derive principles and constraints on interpretation from the implications of the conversational interpretive interaction of speakers and hearers. Confronted with an utterance, the hearer has basically three options. She may be concerned with either (1) what the speaker wants to say (collaborative approach), (2) what the speaker is most reasonably taken to say (conflictual approach) or (3) what the speaker could be imagined to say (playful approach).

(1) Opting for the speaker's *intended meaning* is the hearer's default option. The sentence's linguistic material and diverse contextual cues serve as an indication of the speaker's intention and help the hearer to form an hypothesis of the speaker's intention. If the hearer is made aware, by the speaker's clarification or further discourse, of her interpretation's divergence from the speaker's intention, the hearer will go along with the speaker's intention as it is eventually manifested to her.
(2) The hearer opts for the *most reasonable meaning* of the utterance typically when normative consequences are at stake (assertion, promise). In this case the hearer holds the speaker responsible for the utterance irrespective of the speaker's actual intention; the indication is evaluated as such, i.e. linguistic or compositional meaning and contextual factors – among which whether the context is informal or formal is especially important – are taken into account in an all-things-considered judgment concerning what the utterance was most reasonably taken to convey.
(3) The hearer opts for an *imagined meaning* of the utterance when she takes interest in a merely possible meaning, usually for the purpose of merriment. In this case, the hearer imagines the utterance to have some meaning which she knows corresponds neither to the speaker's intended meaning nor to the most reasonably assigned meaning, yet this imagined meaning is not unreasonable; it is a meaning the utterance could possibly indicate.

It is the second option which will form the basis of the accounts which I will develop in the present paper.

3 Racist Utterances

3.1 Problems About Charges of Racism

What is involved in qualifying an utterance as racist? In actual practice, a charge of racism often corresponds to one or the other of the following claims.

(1) The speaker by making her utterance displayed an attitude towards a certain group of people, an attitude which is to be qualified as racist.
(2) The speaker by her utterance conveyed a communicative intention which was racist.
(3) The speaker's utterance is, on the basis of its linguistic meaning or diverse contextual factors, the carrier of a racist meaning.

Thus charges of racism are either concerned with the speaker's attitude, the speaker's communicative intention or the meaning of the speaker's utterance. In the latter case one may hold that the utterance is racist on account of the linguistic and conventional meaning of the sentence used to make the utterance, or that it is in virtue of contextual factors that the utterance carries a racist meaning.

Let us consider an actual case. In 2016, the Mayor of Clay County in West Virginia, U.S.A., made an utterance of the seemingly innocent sentence 'Just made my day Pam' in connection with a friend's arguably racist Facebook comment. The mayor was subsequently accused of racism on all of the three accounts above. People made claims as to her general attitude ('She is clearly a racist!!!!'), as well as to her communicative intention (the Mayor was said to have 'showed she agreed with her [friend's] comments') and the meaning of her utterance ('Making racist remarks is not how we do things in [C]lay').[1] The Mayor then posted a statement in her defence in which she denied, in turn, the utterance meaning to be racist ('I was referring to my day being made for change in the White House!'), having had a racist intention ('My comment was not intended to be racist at all.'), and also to have a racist attitude ('Those who know me know that I'm not of any way racist!').[2]

A problem with taking an utterance to be racist for the reasons considered above is that the speaker typically has first person authority with respect to her own attitude as well as to her communicative intention and also to the meaning of her utterance, especially if the racist meaning is only implicit and must be recovered from the context, as in the case under consideration. But even in case the utterance contains explicit racism, it still, seemingly, depends on the speaker whether her statement was meant to be serious or somehow ironical e.g., hence the speaker preserves her first person authority (cf. Rattansi 2007).

The outcome of such charges and such defence is quite unpredictable. The intuitions of ordinary people and media vary according to the circumstances. In most cases it is not settled whether the utterance was racist or not. The speaker typically apologizes and she may even resign from her position on account of the breach of confidence which her utterance occasioned, independently of whether the utterance really was racist.

If an accusation of racism is considered in court the question of whether the utterance was racist or not must certainly be definitely settled. But if we look at court decisions, the uncertainty is actually the same as in ordinary contexts. According to Swedish law, e.g., a person may be convicted of agitation against ethnic group if she 'expresses contempt for an ethnic group'. As we can see, the formulation is underdetermined with respect to

[1] http://www.thepetitionsite.com/sv/takeaction/571/113/900/.
[2] https://www.washingtonpost.com/news/post-nation/wp/2016/11/14/ape-in-heels-w-va-officials-under-fire-after-comments-about-michelle-obama/?utm_term=.f24d17261f79.

whether it is the utterance, the intention or the attitude which is at stake and leaves open for denials backed up by first person authority.

Courts usually set out to establish what they label the objective meaning of the utterance. However, there is no consensus on what objective meaning amounts to. It is unclear whether utterance meaning is a matter of conventional, literal meaning or whether contextual factors should also be permitted to constitute the meaning of the utterance (Young 1991; Onishi 2016; Friedersdorf 2017). For one judge '[w]ords have their own meaning and do not necessarily require a context within which to acquire meaning'. For another judge words 'must of necessity be viewed in the context in which they are used for the purposes of determining whether they amounted to a racist slur' (Gaibie 2012). Other judges are more laconic, only stating that the utterance is 'to be considered as an expression of contempt' (Karlström 2007), without indicating on what grounds.

If the racist utterance is treated as a *crimen iniuria* the speaker's intention to convey something racist must be established in addition to the meaning of her utterance. Proving intentions is notoriously difficult and it happens that judges take the intention as a matter of course: 'the words speak for themselves and they conju[n]ctively demonstrate the intention to be racist' (Gaibie 2012). It also happens that the question of the speaker's intent is evaded by declaring that she was indifferent as to the effect and therefore acted as if she had the intention in question (indifferent intent). It also happens that courts take the speaker's behaviour in court into consideration and if she shows no regret this speaks for her having had a racist intention (Theron 2018). Requiring the speaker's intention might be considered problematic in any case, in so far as whether or not the speaker had a racist intention, the damage made unto the hearer is equally serious (cf. Botha 2018).

We can see that there is hardly any agreement, neither in ordinary nor in court practice, on what constitutes or qualifies an utterance as racist. A general problem about the grounds invoked is their referring the issue to the authority of the speaker herself. Hearers and courts can at best have inconclusive evidence to support claims as to utterance meaning, communicative intentions and attitudes and such claims may typically be defeated by the speaker herself. It is true that neither courts nor ordinary people take the speaker's denial as decisive and often distrust the speaker's sincerity, but then they had better base their charges of racism on grounds which do not invoke the speaker's authority at all.

3.2 Speaker Responsibility as Dependent on the Hearer's Interpretation

What is needed is a robust notion of speaker responsibility, which holds the speaker responsible even if the racism of the utterance is only covert and the speaker herself has privileged access to her intention and general attitude. I believe that the implications of conversational interpretive interaction will provide us with such a notion.

According to the interactional model, a hearer has basically three options when confronted with an utterance of the speaker's. She may take an interest in an imagined, merely possible meaning of the utterance (usually for purposes of merriment); she may take an interest in the speaker's intended meaning (her default option); or she may take an interest in the most reasonable interpretation of the utterance (in case she wants to hold the speaker responsible for her utterance).

Taking the speaker to task for having made a racist utterance corresponds to the last of the options mentioned. It amounts to holding the speaker responsible for the most reasonable interpretation of her utterance. When opting for the most reasonable interpretation of the utterance, the hearer's claim is simply that she had good reasons to understand the speaker's utterance the way she did. There are two notable implications of this claim.

(1) The speaker's actual intention and actual attitude are irrelevant. What is at issue is whether the hearer had good reasons to take the speaker to convey what she took her to convey. The hearer's interpretation is based on a consideration of the linguistic and contextual cues which were available to her. It is the epistemic warrantability of the hearer's interpretation which matters, not what the speaker's actual intention was. The hearer's interpretation might very well correspond to the best hypothesis about the speaker's intention and as such constitute the most reasonable interpretation of the speaker's utterance, even though, as a matter of fact, the speaker's actual intention was different.
(2) Whether the meaning the hearer took the speaker to convey was explicitly or implicitly conveyed is not decisive. What matters is whether the hearer had good reasons to take the utterance to convey what she took it to convey. In order to settle this question the specific circumstances (such as previous discourse and situation) of the individual utterance must be taken into consideration. It cannot be regimented in advance on principled grounds what belongs and what does not belong to the most reasonable interpretation of an utterance, even though it is certainly true that it may be more difficult to argue that implicit meaning actually is part of the most reasonable interpretation.

These general lessons from interpretive interaction have definite bearing on the qualification of utterances as racist. When the hearer holds the speaker responsible for her utterance, she enforces her own interpretation of the utterance. The hearer's claim is not that it corresponds to the speaker's intention or attitude, nor that it is confined to the literal, semantic or explicit meaning of the utterance. Consequently, the speaker's first person authority concerning her attitude, intention and the meaning of her utterance is irrelevant. What is at stake in the case of racist utterances is solely whether the speaker made an utterance the most reasonable interpretation of which is racist, i.e. the epistemic evaluation of the warrantability of the hearer's interpretation is decisive.

If we now consider the case of the Mayor of Clay again, we can say that the charges of racism against her were unnecessarily concerned with her attitude, her intention and the meaning of her utterance. The petitioners' claim had better be that hearers had good reasons, even the best reasons, to take her utterance as racist. For the petitioners themselves have authority over what is the most reasonable interpretation of the utterance, since the question as to what it is independent of the Mayor's actual attitude, of her actual intention and also of such a thing as the actual meaning of her utterance. If we want the Mayor to resign from her position the most pertinent thing to say is that it is unacceptable for a person in an official position to make an utterance which is most reasonably taken to be racist (cf. Richard 2018).

In many court decisions we find good specimens of careful, balanced and nuanced reasoning as to what is the most reasonable interpretation of an utterance. However, from the viewpoint of the present conception of utterance responsibility, courts should refrain from labelling what is nothing but the most reasonable interpretation of the utterance the objective meaning of the utterance. They should also admit that literal meaning is not necessary for qualifying an utterance as racist. Finally, they need not make any imputations of intention.

A consequence of this view is that a speaker may be convicted of having made a racist utterance, even though she had no intention to be racist. This amounts to strict liability: the speaker is convicted on account of the damage produced, not her intention. Some may find this repugnant: people should not be convicted of consequences which they had no intention to bring about.

Strict liability is indeed controversial (Simester 2005), but seems here to be compatible both with our aim and our actual practice. We have seen that courts impute liability even if the speaker denies having had any intention to injure, taking the intention more or less for granted or imputing indifferent intent. So even if they profess not to practice strict liability, this is what they in fact often do. Additionally it seems to me that presumably a speaker would better accept being convicted on account of her utterance, if the court did not also make assumptions about the true objective meaning of her utterance or about her actual intentions and real attitude. A speaker is certainly prepared to accept that the hearer might have had good reasons to take the utterance as racist, even if she claims not to have made a racist utterance, nor to have had a racist intention or to be a racist. If strict liability is adopted the speaker need not feel offended by being imputed intentions and attitudes which she claims not to have. A strict liability approach is also pertinent given our aim to prevent the kind of damage which racist utterances result in. It very reasonably imposes uttermost precaution on the part of speakers.

The interactional model of utterance interpretation thus permits us to clearly state the conditions under which an utterance should be qualified as racist, to completely avoid the interference of first person authority and thus block the possibility of subterfuges and to have an effective means of checking the making of utterances which reasonably are perceived as racist. In this way charges of racism are on a safe ground.

4 Slurs

4.1 Slurs as Conveying Speaker Attitudes

Slurs or derogatory terms are lexical items designating groups of people on account of their ethnic, sexual or social identity the use of which is offensive to the referents of the terms. I will focus on racial slurs. If we compare a slur with its neutral counterpart they appear to have the same extension, in so far as they designate the same group of people, but slurs appear to carry an additional component of meaning, a negatively evaluative dimension, which accounts for their offensiveness. Linguists and philosophers of language have recently begun to identify the mechanism which makes a slur a slur. Most theorists agree that speakers by the use of a slur express that its referents or targets are inferior and despicable on account of their belonging to a certain group of people and that this is what explains the offensiveness of slurs: targets are humiliated and dehumanized on account of the speakers' expression of contempt. The debated question is how this contempt gets expressed, whether it is by means of semantic content (e.g. Hom 2008; Hom and May 2013), conventional implicature (e.g. Potts 2012; Whiting 2013), presupposition (e.g. Schlencker 2007; Cepollaro 2015; García-Carpintero 2017) or an associated perspective (Camp 2013). There is as yet no consensus on how this question should be resolved.

A common problem with these accounts is that they disregard the instability of the evaluative dimension of slurs. Salient facts about slurs, however, speak against there being any stable and uniform evaluative component of slurs.

Many racial slurs started as neutral designations and have taken on their offensive character only subsequently. Some slurs are appropriated by the targets and lose much of their offensiveness consequently. Here standard accounts would have to explain how the negative evaluation or the racist attitude both gets into slurs and also may get out of them. Further, some terms function both as derogatory terms and as neutral terms, e.g. the K-word, and here standard accounts would presumably have to posit ambiguity. Finally, some speakers use slurs claiming not to have any racist attitude. Such speakers may be ignorant or confused, but nevertheless they are not clearly contradicting themselves (cf. Camp 2013). They may for example hold that the imposition of a novel designation for a certain group is a manoeuvre of political correctness and constitutes no more than a vain effort to ameliorate the conditions of the group in question which is without any real significance. This attitude does not, however, prevent targets from being offended.

Standard accounts may certainly be able to accommodate facts like these, but at least these facts complicate the picture of the negative evaluation dimension which they take as their starting point. I believe that these facts point to an essential feature of slurs, which should not be given a merely *ad hoc* treatment.

Two other problems about standard accounts have to do with their focus on the speaker's attitude. Because of this focus, I think that offensiveness is not accurately explained and also that the issue of responsibility is not properly addressed.

First, standard accounts assume that the offensiveness of slurs is to be explained by means of the speaker's expression of her negative evaluation of the targets or of her racist attitude towards them. Now to make an utterance which most reasonably is interpreted as expressing the idea that a certain group of people are inferior on account of the colour of their skin – and most uses of slurs are reasonably interpreted in this way – certainly amounts to racism and is as such condemnable. But such linguistic behaviour is hardly offensive *per se*. I am not automatically offended by your expressing your racist ideas, even if I happen to be the target. I may think the speaker has nasty ideas and is blameworthy without taking personal offense. For me to take offense, it is necessary that the speaker has launched a particular attack on me, has manifested disrespect of my person in a more direct way, not only that she has given expression to a general attitude towards people of the group I belong to. Standard accounts tend to disregard that offense cannot be explained solely by reference to the speaker's attitude, as if the target's reaction were not also essential for the offensiveness of slurs. For this reason it seems to me that accounts which lay emphasis on the speaker's expression of negative evaluation are not able to account for the particular offensiveness of slurs.

Second, when the focus is on the speaker's negative attitude, focus is naturally shifted from the damage and injury made to the targets. When a target is offended by the speaker's use of a slur, the target's charge is not primarily concerned with the speaker's having a negative attitude towards the target, but with the fact that the use of the slur is painful to the target. If the charge were concerned with the speaker's attitude, it would be concerned with something over which the speaker herself has first person authority: naturally the speaker is in a privileged position to know about her own attitudes. If the focus is on the damage made unto the target, it is within the authority of the target: the target is a privileged position to know about the injury experienced. For this reason, I believe that if focus is on the speaker's attitude, the speaker will more easily avoid liability.

4.2 Slurs as Unauthorized Designations

I take offensiveness to be the core phenomenon of slurs (like e.g. Richard 2008; Jeshion 2013; Bolinger 2017). The novel account of slurs which I propose effectively explains their offensiveness and readily accommodates the other facts concerning instability considered above. It also avoids the vexed issue of the linguistic status of the evaluative dimension. Further, it avoids the speaker's first person authority.

On my proposal, the basic fact which makes a term a slur is that its targets or referents for some reason or other do not want to be designated by it. A slur is ultimately a designation which is unauthorized by its referents. The offense committed by a speaker using a slur is comparable to the offense committed by a speaker using a proper name for a person which that person has not sanctioned. Since it is inimical to call people by names they do not want to be called by, using a slur is an offensive act which dehumanizes the target in not treating her with due respect. This account of the offensiveness of slurs does not presuppose that any negative attitude is carried by the slur itself. The use of a slur does not serve to express the speaker's negative evaluation, but rather serves to directly attack the target.

On this account, the emergence of a slur may be pictured in the following way. Certain people – the targets – experience that other members of the community – the speakers – have a negative attitude towards them. The speakers regularly talk about the targets in unpleasant ways ('What you would expect from a … ?') and point out their membership in contexts where it is quite irrelevant ('We were served by a … today.'). The problem is not with the designation used but with its contextual occurrence, i.e. that the term occurs in negatively charged settings, which reflects the speakers' negative evaluation of the targets. In order to remedy the situation a little, the targets propose a new designation for their group. The introduction of this new term will not change the speakers' attitudes of course, but will give them an opportunity to reflect on their behaviour. Speakers who adopt the new designation will thereby show that they treat targets with dignity in respecting their will to be designated by the term they prefer. Speakers who do not adopt the new designation will effectively show that they do not care about the targets. The new term thus functions as a shibboleth which marks out friends from foes. In this way one designation is replaced by a novel designation, not because the old term as such carries a negative evaluation, but because the use of the new term is a signal of respect. The use of the old term is offensive because it amounts to disrespecting the will of the targets, their reasonable demand to be called by the name they prefer. Will the introduction of the new term and the constitution of the old term as a slur remedy the targets' situation? It is certainly no more than a first step in an attempted reformation of speakers' attitudes.

Another way for slurs to emerge is by the speakers' specific design of offensive designations for targets which they despise. By the speakers' having recourse to unpleasant sounds, morphemes and connotations, these terms may seem to be invested with the speakers' contempt. Thus their being slurs would not depend on their being unauthorized by their targets, but on their being directly expressive of the speakers' attitudes. Targets may even be unaware of the existence of these terms. Rather than conceiving of designations like these as being expressive of negative evaluation, however, we could conceive of them as being calculated to offend their referents. For a term to be a slur, it might suffice that it is potentially unauthorized. Speakers may create designations which they trust would not be approved by their referents.

This picture easily explains how originally neutral designations may become slurs. It also explains how a speaker may say '… are nice people' without involving herself in contradiction, because she may be unaware that targets do not want the designation to be used. Targets may nevertheless have good reasons to take offense, since the slur, even though it does not by itself convey a negative evaluation, is disrespectful. Appropriation is also less mysterious on the present account. Targets appropriating a slur do not change the linguistic meaning of the term, but disarm speakers by depriving them of one means of offending them. Finally, since the offensiveness of slurs does not involve the speaker's attitude, she will not avoid liability by making claims about her non-racist attitude.

As already mentioned some people think the imposition of a novel designation for a certain group replacing the allegedly derogatory term previously in use is quite vain. They may hold that the old term contained no negative evaluation and also think that the change of designations does not change the real situation for the people in question. In the light of the present account such a change of designations should not be seen as a manoeuvre of political correctness, but as corresponding to how the referents want to be designated themselves. People should stop using the old term, not because it carries a negative evaluation, but because the referents want to be called by the new term. To persist in using the old term is not to be critical of political correctness, but to fail to pay fellow citizens due respect.

4.3 Comparison with Other Accounts

My account could be compared with two major accounts of slurs which also do not take meaning to be what makes a slur a slur. Nunberg (2018) thinks that the primary function of slurs is not to offend targets, but to be group-affiliating: speakers use slurs to enforce their own sense of belonging to a community, a community which is defined by its antagonism to the targets' community. It seems to me, however, that the group-affiliating function presupposes offensiveness. If the speakers using a slur to reinforce their identity as allegedly superior to the targets were aware that the targets would not object to their being designated by the term in question the group-affiliating function of using the slur would presumably be lost.

Anderson's and Lepore's (2013a, b) prohibition account is similar to the present account. What makes a slur a slur, according to them, is that the slur is a prohibited term. But it does not seem to me that prohibition takes us far enough. If slurs are prohibited terms, slurring consists in infringing a rule, which certainly is transgressive, but hardly offensive. A target is not humiliated and dehumanized by the speaker's simply breaking a rule, but by the speaker's not respecting the target's will. Slurring is not injurious because it is prohibited; rather slurring is prohibited because it is injurious.

My account of slurs, like my account of the speaker's responsibility for racist utterances, removes the speaker's first person authority. The interactional model clearly situates the authority in cases of accountability with the hearer. A slur is not primarily a matter of a speaker's attitude, but of the injury made to the target. My account of slurs as unauthorized designations clearly explains the damage caused by the use of slurs.

5 Conclusion

Most conceptions of speaker responsibility as well as most accounts of slurs take a speaker oriented perspective. The speaker's responsibility is tied to her intention and attitude. I have attempted to show that such accounts are problematic in so far as they locate the speaker's responsibility within the speaker's own authority. I have argued that utterance responsibility must be tied to the hearer's interpretation which is independent of the speaker's actual intention or attitude and also the distinction between explicitly and implicitly conveyed meaning. The speaker may be accountable for any interpretation of her utterance, provided that the hearer has the best reasons to take the utterance the way she does. This provides a firm basis for checking racist language.

Acknowledgements. I would like to thank three anonymous referees for valuable suggestions.

References

Anderson, L., Lepore, E.: Slurring words. Noûs **47**(1), 25–48 (2013a)

Anderson, L., Lepore, E.: What did you call me? Slurs as prohibited words. Anal. Philos. **54**(3), 350–363 (2013b)

Asher, N., Lascarides, A.: Strategic conversation. Semant. Pragmatics **6**, 1–62 (2013)

Bach, K.: Loaded words: on the semantics and pragmatics of slurs. In: Sosa, D. (ed.) Bad Words: Philosophical Perspectives on Slurs, pp. 60–76. Oxford University Press, Oxford (2018)

Blakemore, D.: Slurs and expletives: a case against a general account of expressive meaning. Lang. Sci. **52**, 22–35 (2015)

Bolinger, R.J.: The pragmatics of slurs. Noûs **51**(3), 439–463 (2017)

Botha, J.: Why precisely formulated laws regulating racist speech are needed. Business Day (2018). https://www.businesslive.co.za/bd/opinion/2018-10-10-why-precisely-formulated-laws-regulating-racist-speech-are-needed/

Burger, C.J.: Bronston v. United States, 409 U.S. 352 (1973). https://supreme.justia.com/-cases/federal/us/409/352/case.html

Camp, E.: Showing, telling, and seeing: metaphor and 'poetic' language. In: The Baltic International Yearbook of Cognition, Logic and Communication, vol. 3, pp. 1–24 (2008)

Camp, E.: Slurring perspectives. Anal. Philos. **54**(3), 330–349 (2013)
Camp, E.: Insinuation, common ground, and the conversational record. In: Fogal, D., et al. (eds.) New Work on Speech Acts, pp. 40–66. Oxford University Press, Oxford (2018)
Cepollaro, B.: In defence of a presuppositional account of slurs. Lang. Sci. **52**, 36–45 (2015)
DiFranco, R.: Pejorative Language. Internet Encyclopedia of Philosophy (2014). https://www.iep.utm.edu/pejorati/#SH3f
Fricker, E.: Stating and insinuating. Proc. Aristotelian Soc. **86**(1), 61–94 (2012)
Friedersdorf, C.: Words Which by Their Very Utterance Inflict Injury. The Atlantic (2017). https://www.theatlantic.com/politics/archive/2017/04/words-which-by-their-very-utterance-inflict-injury/523344/
Gaibie, S.: Modikwa Mining Personnel Services v S.E. Ramaepadi and Others. JR 1904/2010. The Labour Court of South Africa, Johannesburg (2012)
García-Carpintero, M.: Pejoratives, contexts and presuppositions. In: Brézillon, P., Turner, R., Penco, C. (eds.) CONTEXT 2017. LNCS (LNAI), vol. 10257, pp. 15–24. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-57837-8_2
Hom, C.: The semantics of racial epithets. J. Philos. **105**(8), 416–440 (2008)
Hom, C., May, R.: Moral and semantic innocence. Anal. Philos. **54**(3), 293–313 (2013)
Hom, C.: Pejoratives. Philos. Compass **5**(2), 164–185 (2010)
Hornsby, J.: Meaning and uselessness: how to think about derogatory words. Midwest Stud. Philos. **25**, 128–141 (2001)
Jeshion, R.: Expressivism and the offensiveness of slurs. Philos. Perspect. **27**(1), 231–259 (2013)
Karlström, I.: Dom nr 20/07 i mål nr A 54/06. Arbetsdomstolen, Stockholm (2007)
Lee, J.J., Pinker, S.: Rationales for indirect speech: the theory of the strategic speaker. Psychol. Rev. **117**(3), 785–807 (2010)
Leth, P.: Is there any use for a notion of the correct interpretation of an utterance? In: Stalmaszczyk, P. (ed.) Philosophical Insights into Pragmatics, pp. 83–106. De Gruyter, Berlin (2019)
Leth, P.: Speakers, hearers and demonstrative reference. In: Ciecierski, T., Grabarczyk, P. (eds.) The Architecture of Context and Context-Sensitivity. Springer, Heidelberg (forthcoming a)
Leth, P.: Utterance Interpretation and Actual Intentions. Axiomathes (forthcoming b)
Nunberg, G.: The social life of slurs. In: Fogal, D., et al. (eds.) New Work on Speech Acts, pp. 237–295. Oxford University Press, Oxford (2018)
Onishi, N.: Jail Time for Using South Africa's Worst Racial Slur? The New York Times (2016). https://www.nytimes.com/2016/10/28/world/africa/south-africa-hate-speech.html
Pinker, S., Nowak, M.A., Lee, J.J.: The logic of indirect speech. Proc. Natl. Acad. Sci. **105**(3), 833–838 (2008)
Potts, C.: Conventional implicature and expressive content. In: von Heusinger, K., et al. (eds.) Semantics: An International Handbook of Natural Language Meaning, pp. 2516–2536. De Gruyter, Berlin (2012)
Quinn, F.: Law for Journalists. Pearson, Boston (2015)
Rattansi, A.: Racism. Oxford University Press, Oxford (2007)
Richard, M.: When Truth Gives Out. Oxford University Press, Oxford (2008)
Richard, M.: How do slurs mean? In: Sosa, D. (ed.) Bad Words: Philosophical Perspectives on Slurs, pp. 155–167. Oxford University Press, Oxford (2018)
Robertson, G., Nicol, A.: Media Law. Sweet & Maxwell, London (2002)
Schlenker, P.: Expressive presuppositions. Theor. Linguist. **33**(2), 237–245 (2007)
Simester, A. (ed.): Appraising Strict Liability. Oxford University Press, Oxford (2005)
Soames, S.: Drawing the line between meaning and implicature – and relating both to assertion. Noûs **42**(3), 440–465 (2008)
Sternau, M., Ariel, M., Giora, R., Fein, O.: A graded strength for privileged interactional interpretations. In: Allan, K., et al. (eds.) Pragmemes and Theories of Language Use. Perspectives in Pragmatics, Philosophy & Psychology, pp. 703–723. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-43491-9_35

Stokke, A.: Lying and Insincerity. Oxford University Press, Oxford (2018)
Theron, L.V.: Rustenburg Platinum Mine v SAEWA obo Bester and Others ZACC 13. Constitutional Court of South Africa, Johannesburg (2018)
Whiting, D.: It's not what you said, it's the way you said it: slurs and conventional implicatures. Anal. Philos. **54**(3), 364–377 (2013)
Williamson, T.: Reference, inference, and the semantics of pejoratives. In: Almog, J., Leonardi, P. (eds.) The Philosophy of David Kaplan, pp. 137–158. Oxford University Press, Oxford (2009)
Young, E.A.: Regulation of racist speech: in re welfare of R.A.V. Harvard J. Law Public Policy **14**(3), 903–913 (1991). 464 N.W.2d 507 (Minn. 1991)

Contextuality Across the Sciences: Bell-Type Theorems in Physics and Cognitive Science

Yoshihiro Maruyama(✉)

The Hakubi Centre for Advanced Research, Kyoto University, Kyoto, Japan
maruyama@i.h.kyoto-u.ac.jp

Abstract. Contextuality is a transdisciplinary phenomenon observed across the sciences, including in particular physics, linguistics, artificial intelligence, and cognitive science. In the present paper we shed new light on cognitive contextuality based upon recent developments of quantum cognitive science. We first discuss quantum cognitive science from the perspective of contextuality studies, and formulate the basic tenet of quantum cognitive science as the quantum cognition thesis or the structural quantum mind thesis as opposed to Penrose's (controversial) quantum brain thesis or material quantum mind thesis. We then discuss Bell-type contextuality results in physics and in cognitive science, and elucidate fundamental differences between cognitive contextuality and physical contextuality. Quantum cognitive science allows us to explicate similarities and dissimilarities between the laws of matter and the laws of mind, contributing to a deeper understanding of the Cartesian dualism.

1 Introduction

Contextuality is a phenomenon observed across different sciences, even though there is no single discipline devoted to contextuality studies yet. Let us give a bird's-eye view of contextuality in different sciences as follows (which however is not intended to be an exhaustive list).

- Contextuality of truth in epistemology: truth is a function of contexts; a single proposition may be true in one context and at the same time false in another; it may have different truth values in different contexts; epistemic contextualism has an origin in Wittgenstein's later philosophy [17].
- Contextuality of being in ontology: agents exist within contexts; being is inseparable from contexts such as environments; this sort of contextualism in ontology has an origin in Heidegger's philosophy and is related to the issues of situated AI, embedded-embodied AI, and Heideggerian AI [7].
- Contextuality of meaning in language: words get meaning within contexts; their meaning may be different in different contexts; the indispensability of contexts in the meaning determination process may lead to some weak form of Quinean semantic holism (no meaning without some wider context).

G. Bella and P. Bouquet (Eds.): CONTEXT 2019, LNAI 11939, pp. 147–160, 2019.
https://doi.org/10.1007/978-3-030-34974-5_13

- Contextuality of reality in quantum physics: measurement values and their statistics exist within contexts; there may be no global assignment of values and probabilities; this sort of contextuality in physics has an origin in Bell's and Kochen-Specker's No-Go theorems refuting classical (non-local) realism.
- Contextuality of reason in cognitive science: cognitive behavior is a function of contexts; a single question may have different answers in different contexts; contextual effects such as coexisting information and environmental noise in the real world may affect and change results of cognitive decision making.

In the following we shall discuss recent developments of contextuality studies across physics and cognitive science. In physics, Bell-type theorems prove contextuality of reality, a particular case of which is quantum non-locality, what Einstein called "spooky action at a distance"; Bell's theorem tells us that local hidden variable theories cannot account for quantum correlations, and we cannot keep the simple, classical picture of reality [1,2]. Although quantum theory is occasionally said to support indeterminism, there are actually deterministic interpretations of quantum theory, such as Bohmian mechanics and many-worlds interpretation. The non-classical feature of quantum theory is arguably best understood as contextuality and non-locality rather than indeterminism. Contextuality in quantum physics means that there is no consistent assignment of values or probability distributions on all variables in different contexts involved; put another way, measurement values or measurement statistics essentially depend upon measurement contexts. That is to say, we can measure each variable in each context, and yet the results thus obtained are not consistent as a whole when combined together. This is why contextuality is characterized as local consistency plus global inconsistency [1,2]. There is no problem in measuring each variable in different local contexts, and yet it impossible to have the global assignment of values or distributions to all variables at once that is consistent with the results in local contexts. Logically phrasing, all physical propositions cannot have bivalent truth values at once. Truth values in quantum physics are contextual, i.e., the truth values of physical propositions only exist within particular contexts. All this is about contextuality in physics. Cognitive science is involved in contextuality as well as quantum physics. Subjects in psychological experiments are readily influenced by different environments or contexts. In both quantum physics and cognitive science, observers affect systems observed; in psychology, for example, those who ask questions affect those who are asked them. Context sensitivity in psychology conceptually looks analogous to that in physics. This idea has led to recent developments of quantum cognitive science. In the following we shall review quantum cognitive science and differentiate the quantum mind thesis of quantum cognitive science from the quantum brain thesis of Penrose et al. (Sect. 2). And we then discuss contextuality in cognitive science and its relationships with contextuality in quantum physics (Sect. 3).

2 Quantum Cognitive Science and Contextuality Studies

What is quantum cognitive science? And what is it good for? Quantum cognitive science (aka. quantum cognition) is an emerging field; the literature on it has rapidly expanded in the last ten or fifteen years. Quantum cognitive science is highly interdisciplinary, involving psychology, linguistics, decision theory, behavioral economics, and so fourth. It is well known today that quantum systems exhibit super-classical correlations such as non-local correlations. Yet there is still an upper bound on the strength of quantum correlations; for example PR (Popescu-Rohrlich) box correlations are known to be super-quantum. Some researchers think that context sensitivity in psychology results in super-quantum correlations; for example, what is called the marginal selectivity condition is violated in certain psychological experiments while it always holds in quantum physics (these issues shall be discussed in more detail in the next section). Quantum cognition is a rapidly developing, vital field and yet its future is still uncertain. We do not yet know whether it is a "new kind of science" or a new kind of "fashionable nonsense." Still, quantum cognition is shedding new light on the fundamental nature of human reason, such as rationality and contextuality, as we shall see in the following. Major issues in quantum cognition include the following, all of which are actually related to contextuality in cognition.

- The order effect in psychology. It is non-commutativity in psychology. Let Q_1 be "Is Clinton honest and trustworthy?", and Q_2 "Is Al Gore honest and trustworthy?". Then, Q_1 and Q_2 are non-commutative. That is, it is experimentally verified [21] that if you ask Q_1 first then you get a less probability for Q_2 (i.e., less people answer yes for Q_2), and if you ask Q_2 first you get a higher probability for Q_1 (i.e., more people answer yes for Q_1). Cognition is non-commutative as well as quantum reality. This is a case of contextuality in psychology; past questions are contexts for present questions.
- The conjunction effect in cognitive biases. Succinctly saying, it shows that $Prob(\varphi \wedge \psi) \leq Prob(\psi)$ (i.e., the monotonicity of probabilities in terms of conjunction) does not hold in certain cognitive bias experiments such as the Linda problem by Tversky-Kahneman. In the Linda experiment [20], subjects are given a description about Linda that makes Linda look like a feminist, such as the following: "Linda is 31 years old, single, outspoken, and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in anti-nuclear demonstrations." And then subjects are asked to judge which of "Linda is a bank teller" and "Linda is a bank teller and active in the feminist movement" is more probable. A significant number of people robustly tend to choose the latter, regardless of their various backgrounds. Hence the violation of $Prob(\varphi \wedge \psi) \leq Prob(\psi)$. This is a case of contextuality, since the description about Linda provides a context for the question (which can actually be answered correctly without any description) and thus makes human judgements biased.
- The disjunction effect in the prisoners' dilemma. Rational decision theory tells us that a prisoner defects regardless of whether the other prisoner defects or not. But this is experimentally violated, and information about the other prisoner's decision significantly affects the prisoner's decision [15]. Mathematically, it is violation of the sure thing principle or the law of total probability:

$\text{Prob}(\varphi) = \text{Prob}(\psi)\text{Prob}(\varphi|\psi) + \text{Prob}(\neg\psi)\text{Prob}(\varphi|\neg\psi)$. Note that this is a case of contextuality, since the contextual information about the other prisoner's decision affects the prisoner's decision.

Quantum cognition exploits quantum mechanical models (or more general probabilistic models) to account for such non-classical features of cognition, in particular violations of the laws of classical probabilities; the statistics of non-classical phenomena in psychology are explained using statistical models in quantum physics [3]. There are some physical phenomena that classical physics cannot account for, and thus we need quantum physics to do that. Likewise, there are some cognitive phenomena that classical psychology or cognitive science cannot account for, and thus we need quantum psychology or quantum cognitive science. Logically, these effects may be interpreted in terms of substructural logic, which allows us to control properties of conjunction, disjunction, and ordering via what are called structural rules [10]. Quantum cognitive science is particularly interesting for the following reason: if there are structural mechanisms shared by both physics and cognition, it would pave the way for overcoming the Cartesian dualism of matter and mind, just as Chalmers' property dualism or double-aspect theory of information [5] aims at elucidating the higher structural laws of information that govern the actual laws of matter and of cognition. It may eventually lead to a scientifically sound embodiment of Chalmers' property dualism or double-aspect theory of information. As the case studies above show, quantum cognitive science may be seen as elucidating cognitive contextuality in different guises, including, inter alia, order contextuality, description contextuality, and judgement contextuality. Quantum cognition also tells that cognitive biases as in the Linda problem may not be about irrationality in human cognition, but rather about a non-classical (or quantum) sort of rationality. It may be that such non-classical traits were advantages in the evolution of the mind.

Penrose is known for his controversial argument for the quantum nature of the brain [12,13]. Let us compare and contrast Penrose's quantum brain argument with quantum cognitive science. Penrose [12,13] argues as follows:

- (i) AI or the computational theory of mind is misconceived in light of Gödelian incompleteness. This is the (notorious) Lucas-Penrose argument; the capacity of human cognition is not bound by computability. Although the Lucas-Penrose argument has been criticized so much, it actually has an origin in Gödel's 1951 Gibbs Lecture [9] (which was not criticized so much): "the human mind (even within the realm of pure mathematics) infinitely surpasses the power of any finite machine" or "there exist absolutely unsolvable diophantine problems"; Gödel's mathematical realism excludes the latter possibility (that is, the truth values of formally undecidable propositions are determined in mathematical reality).
- (ii) The mind is materially quantum; consciousness emerges via material quantum processes in microtubules in the brain. Let us call it the Material Quantum Mind Thesis (Material QMT for short) or the Quantum Brain Thesis.

Penrose has been criticized so much for these arguments. And approximately two hundred pages of his *Shadows of the Mind* are devoted to replies to different types of criticisms. Chalmers [4], for example, argues against Penrose as follows: "Why should quantum processes in microtubules give rise to consciousness, any more than computational processes should?"; "reader who is not convinced by Penrose's Gödelian arguments is left with little reason to accept his claims that physics is non-computable and that quantum processes are essential to cognition." Yet, many years later, Pothos-Busemeyer [15] say as follows:

> [T]he success of human cognition can be partly explained by its use of quantum principles.

This may imply that there is now some reason to accept Penrose's claim that "quantum processes are essential to cognition." Yet we have to be careful here; quantum effects may not necessarily arise from the material brain. Indeed, Tegmark [19] argues as follows:

> Based on a calculation of neural decoherence rates [...] the degrees of freedom of the human brain that relate to cognitive processes should be thought of as a classical rather than quantum system [...] This conclusion disagrees with suggestions by Penrose and others that the brain acts as a quantum computer, and that quantum coherence is related to consciousness in a fundamental way.

If this is correct, there is no quantum effect in the material brain, and there is no problem on classical realism about the brain. In the following I assume that Tegmark is correct in this respect: the cognitive mechanism of the human brain is classical; human beings are classical macroscopic objects within the scope of classical realism. We shall later use this classical brain assumption. Note that Penrose actually does not claim that the brain is a quantum computer (while Hameroff does). Indeed, Penrose says in his 1995 paper "Consciousness Involves Noncomputable Ingredients" [14] as follows:

> When I argue that the action of the conscious brain is noncomputational, I'm not talking about quantum computers. Quantum computers are perfectly well-defined concepts, which don't involve any change in physics; they don't even perform noncomputational actions. [...] I don't think it can explain the way the brain works. That's another misunderstanding of my views.

The Material QMT or Quantum Brain Thesis is highly controversial; yet quantum cognitive science allows us to devise a less controversial, modest version of the QMT, what we call the Structural QMT or the Quantum Cognition Thesis:

– The mind is structurally quantum; the structure of cognitive systems is homomorphic to the structure of quantum systems. The structure of economic systems is homomorphic to the structure of physical systems, and so we can apply physical models to explain economical phenomena. Yet this never means that the nature of economy is materially physical. Likewise, we can apply quantum physical models to explain cognitive phenomena, and yet it is not that quantum processes are materially going on in the macroscopic physical brain, which is entirely classical as Tegmark shows.

– This idea is also related to functionalism in philosophy of mind. It endorses the view that the mind is characterized by what it does rather than by what it is made of, that is to say, the essence of the mind is its cognitive function rather than its material substance (cf. Cassirer's *Substance and Function*). And different functions of the mind may be explained by different quantum-like models, i.e., structurally quantum models, rather than materially quantum models. The Material QMT is based upon cognitive substantivalism, and the Structural QMT upon cognitive functionalism.

Quantum cognition research does not claim anything like material quantum effects in the physical brain. In quantum cognitive science, mathematical models of quantum physics are applied to cognitive phenomena just in the same way as certain physical models are applied to economical phenomena (e.g., mathematical models of Brownian motion in physics are applied to financial market analysis). As shown in many case studies in quantum cognitive science, we can apply quantum models to explain (the statistics of) cognitive phenomena. Yet this never means that the nature of cognition is materially quantum; rather it is structurally quantum. If the universe is a materially quantum computer as quantum physicists such as Llyod [11] argue, the mind might be a structurally quantum computer.

This would be one way to make Penrose's Material QMT scientifically more acceptable. Let us finally think of how to make Penrose's Gödelian argument less controversial. One way to do so would be replacing computability by complexity (not necessarily in the technical sense), thus arguing that classical AI or the classical computational theory of mind is misconceived in light of the super-classical features and efficiency of human cognition, and the mind cannot be a classical computer due to the non-classical complexity of cognition. There is no clear evidence that the human mind can actually go beyond the limit of computability; rather the Tractable Cognition Thesis [16] asserts the opposite. Yet at the same time the human mind exhibits super-classical features as the aforementioned experimental studies show, and those non-classical features were not able to be explained by classical models of cognition and decision (this is why quantum cognitive science was born), considered to be anomalies or biases within the classical paradigm. The human mind also processes information in its environment in a highly effective manner, and human cognitive information processing is even today more effective in certain pattern recognition and classification tasks (especially, those requiring the recognition of contextual information within given environments as in the frame problem) than the classical computer equipped with statistical machine learning algorithms. In such a way, thus, we could possibly make Penrose's argument less controversial (though no decisive conclusion can be drawn at the present stage of human cognition research).

3 Contextuality Across Physics and Cognitive Science

Context-sensitivity is observed in both cognitive and quantum systems, and yet there are also differences between contextuality in cognitive systems and that in quantum systems. Contextuality in cognition also involves the fundamental problem of psychology in the following sense.

– Like quantum systems, cognitive systems are sensitive to contexts of measurement.
– Unlike them, cognition is so embedded in contexts that contextual side-effects cannot adequately be controlled.

Cognitive systems are "beings-in-context" or "beings-in-the-world" (cf. situated AI; embedded AI; Heideggerian AI [7]). The life scientist says, "Life is warm, wet, and noisy." Physical experiments, too, are subject to contextual noise, which can still mostly be controlled. There are both internal and external noise in psychological experiments, caused by uncontrollability of mental states and by uncontrollability of environments, respectively. Contextual effects make state preparation difficult in psychology, in which it is unclear what kind of cognitive states is to be measured in the first place. Due to such contextual noise, the statistics of human cognitive experiments often violates the so-called marginal selectivity condition [8]. The relationships between parts and wholes, therefore, are even more complex in psychology than in quantum physics (cf. Heisenberg's *Der Teil und das Ganze*):

– In quantum physics: wholes ($\otimes$) are not direct sums ($\times$) of parts. (Note that $H \times H'$ is isomorphic to $H \oplus H'$ for Hilbert spaces H, H' representing quantum systems, and that tensor product $H \otimes H'$ is strictly bigger than direct sum $H \times H'$). This is called quantum holism.
– In psychology: parts (marginals) are not direct restrictions of wholes (joint distributions). This amounts to the violation of marginal selectivity (which says that, given joint distributions, marginal distributions coincide with original non-joint distributions). It may be seen as the dual of the above holism.

Quantum holism is conceptually analogous to Gestalt psychology. Note that the difference between tensor product ($\otimes$) and direct product ($\times$) gives rise to quantum phenomena such as entanglement and non-locality.

When the marginal selectivity condition is violated as in the second case, Bell-type inequalities must be extended so as to take such additional effects into account. The violation of marginal selectivity is caused by contextual noise, which is called direct contextuality as opposed to genuinely quantum contextuality [8]. Bell-type inequalities can be extended so as to take direct contextuality into account and to detect proper quantum contextuality even in the coexistence of direct contextuality [8]. In terms of the characterization of quantum contextuality as local consistency plus global inconsisntecy, the dual holism above means that even local consistency may not hold in the statistics of human cognition. Local inconsistency and global inconsistency together imply that parts are independent of wholes; this property may be called strong holism as opposed to weak holism such as quantum holism, in which wholes are merely bigger than direct sums of parts.

It should be remarked that contextual noise is what threatens reproducibility in cognitive and life sciences while at the same time technically leading to the aforementioned statistical violation of marginal selectivity. The so-called reproducibility crisis has been a big concern in cognitive and life sciences. According to The Reproducibility Project: Psychology, only 35% of published experimental studies were reproducible [22]. There is a similar problem in medical science (e.g., the STAP affair at RIKEN). The internal and external noise as hidden contextual noise threatens the reproducibility of cognitive experiments. It is relatively rare in physical experiments; contextual noise is more controllable in physics (even though cosmology or the history of the universe is not really reproducible). Cognitive systems are so complex and sensitive that subtle contextual noise cannot be controlled on them.

What are called No-Go theorems in quantum foundations, such as Bell's and Kochen-Specker's, tell us that fundamental reality is inherently contextual (and non-local as a particular case); from a theoretical point of view [1,2], contextuality boils down to local consistency plus global inconsistency (of contextual information; e.g., probability distributions on measurement contexts). According to recent research in quantum cognitive science, Bell-type inequalities may be reformulated so as to be applicable in cognitive science, and they are actually violated in certain cognitive experiments [6]. Now several questions arise as follows. Do Bell-type results in cognitive science show that human reason is contextual as well as fundamental reality? If so, in what sense? Put another way, do cognitive systems exhibit the same kind of contextuality as quantum systems? And ultimately, do Bell-type results in cognitive science have such a massive impact on our understanding of the world as those in quantum physics indeed had? There are analogies and disanalogies between contextuality of reality and cotenxtuality of reason. In the following let us attempt to articulate the analogies and disnalogies between contextuality of reality and contextuality of reason, especially in light of the nature of probabilities involved. The point is that quantum and cognitive systems exhibit the same kind of contextuality at a mathematical level of statistical correlation (apart from the issue of violation of marginal selectivity), and yet physical contextuality differs from cognitive contextuality in terms of how relevant probabilities are interpreted therein. This disagreement about the nature of probabilities, arguably, makes the meaning of Bell-type theorems in cognitive science depart from that in quantum physics in a significant manner. This would also explicate the reason why science is not just about the analysis of statistical correlations. For these purposes we have to revisit some statistical details of Bell-type experiments in physics and cognitive science.

Bell's non-local statistics is as follows. As usual, a_i represents Alice's measurement and b_i Bob's; (x, y) represents a pair of resulting values of those measurements.

	$(0,0)$	$(0,1)$	$(1,0)$	$(1,1)$
(a_1, b_1)	$1/2$	0	0	$1/2$
(a_1, b_2)	$3/8$	$1/8$	$1/8$	$3/8$
(a_2, b_1)	$3/8$	$1/8$	$1/8$	$3/8$
(a_2, b_2)	$1/8$	$3/8$	$3/8$	$1/8$

It does not really matter how these probabilities obtained, but what does matter here is that it is actually possible to contrive a physical experiment yielding the above statistics. One can then compute what is called the CHSH inequality, one of the Bell-type inequalities, which must hold if the conditions of classical local realism hold, and it finally turns out that the CHSH inequality is violated in the above statistics, thus showing by reductio ad absurdum that there is no classical way to explain the above statistics, that is, it truly goes beyond classical physics. This is the basic storyline of Bell-type No-Go results, which usually show contextuality in the sense of the impossibility of globally assigning probability distributions to all the random variables at once in a consistent manner. In more informal terms, measurement statistics are essentially contextual in quantum physics. If one assumes non-contextual statistics, then one gets a Bell-type inequality, the violation of which shows contextuality. There is even a general theory of showing contextuality via Bell-type inequalities [2]. Contextuality here is of purely statistical nature and contextual statistics can in principle arise from cognitive experiments as well as physical ones. Indeed, the following statistics obtained from a cognitive decision making experiment [6] (note that this is simplified statistics and actual probabilities are more complex, but the difference does not affect the following discussion):

	$(0,0)$	$(0,1)$	$(1,0)$	$(1,1)$
(a_1,b_1)	9/10	0	0	1/10
(a_1,b_2)	8/10	0	0	2/10
(a_2,b_1)	8/10	0	0	2/10
(a_2,b_2)	0	6/10	4/10	0

The above statistics is similar to the following statistics of what is called the PR box in quantum foundations (see, e.g., [1,2]); the possibilistic versions (which focus on whether values are zero or not) of the two tables are exactly the same.

	$(0,0)$	$(0,1)$	$(1,0)$	$(1,1)$
(a_1,b_1)	1/2	0	0	1/2
(a_1,b_2)	1/2	0	0	1/2
(a_2,b_1)	1/2	0	0	1/2
(a_2,b_2)	0	1/2	1/2	0

The PR box statistics exhibits the maximal violation of the CHSH inequality. The above statistics of cognition does not exhibit the maximal violation, but still violates a CHSH inequality extended to take into account the violation of marginal selectivity as well. The extended CHSH inequality is as follows. Let a_i^j denote a_i in context b_j (contextuality-by-default). Let $\tau = \sum_{i\in\{1,2\}} |E[a_i^1]-E[a_i^2]|+\sum_{j\in\{1,2\}} |E[b_j^1]-E[b_j^2]|$. The extended CHSH inequality is then as follows:

$$\mathrm{CHSH} - \tau \leq 2$$

where $\mathrm{CHSH} = \max_{k,l\in\{1,2\}} |\sum_{i,j\in\{1,2\}} E[a_i^j b_j^i] - 2E[a_k^l b_l^k]|$ (for more details see [6]). The original CHSH inequality is simply: CHSH ≤ 2. Note that the asymmetry of probabilities in the above statistics of cognition seemingly comes from the asymmetry of question order, and in a symmetrized question experiment, we could obtain statistics closer to that of the PR box above. Note also that similar statistics may arise from experiments on animals as well as humans because the structure of the experiment can be implemented in quite a simple situation.

Formally, there is no difference between physical contextuality and cognitive contextuality: there is strange statistics in the first place; there is an inequality whose violation implies contextuality; and it is violated in the statistics. From the point of view of purely mathematical statistics, there is indeed no difference between them (except for the technical point that marginal selectivity violation must be taken into account in cognitive experiments). The difference rather lies in how statistics is produced or where probabilities come from. That is, the intrinsic difference between physical and cognitive contextuality is concerned with the origins of probabilities involved. Where are origins of probabilities in the above contextuality experiments in quantum physics and psychology?

– Physics: you get statistics by repeating an experiment on a single state (e.g., an entangled state such as the Bell state). Probabilities come from the state per se. That is, probabilities are state-intrinsic in physical contextuality.
– Psychology: you get statistics by repeating an experiment on different subjects in different mental states (e.g., someone on a university campus). Probabilities do not come from any specific state. That is, probabilities are state-extrinsic in cognitive contextuality.

States are not fixed in the above cognitive experiment while they are fixed in the Bell experiment. Why? There would be both positive and negative reasons. The negative reason is that it is difficult to fix mental states of human subjects in psychological experiments. The positive reason is that the results get averaged if the number of experiments repeated tends to infinity, and this is crucial because you want to apply the law that results from those experiments to arbitrary persons (psychology and medicine should apply to anyone who suddenly comes to the clinic). Put another way, if you strictly fix a cognitive state in experiments, the resulting law of cgnition will only apply to those who are in the same cognitive state, but there is no firm reason that different persons can have exactly the same cognitive state in the first place. In a nutshell, if cognitive science gets highly personalized, it loses its broad applicability, generality, and universality. There is surely a merit for not fixing cognitive states; by doing so, we can make cognitive laws applicable to arbitrary persons, not with absolute certainty, and yet with some statistical certainty. This is the positive reason.

What happens in the phenomenon of quantum contextuality or non-locality in particular is basically that you have a special state, and then you get special statistics (here we do not think about state-independent contextuality arguments). What happens in cognitive contextuality is that you have ordinary people in ordinary, yet different mental states, and then a special experiment allows you to get special statistics. To elucidate the difference between cognitive and quantum contextuality, let us think of an experimental version of Laplace's demon. The theoretical Laplacian demon has the infinitary power of exact computation. We think of the experimental Laplacian demon as follows:

- The experimental Laplacian demon has the infinitary power of precise experimentation. In particular, the demon can fix every hidden parameter of a (cognitive) experiment in any noisy environment. The demon repeats the experiment countless times. Yet he always gets the same result since he fixed every parameter (under the classical brain assumption). Hence no probabilities in the demon's psychology experiment.
- To the demon, therefore, there is no indeterminacy or contextuality in cognition (because he fixed every parameter himself). Note that there is no statistics to be analyzed in the demon's experiment because the experimental Laplacian demon can fix every hidden parameter in the environment.
- What if the demon performs contextuality (or non-locality) experiments on quantum systems? The demon cannot erase contextuality by manipulating parameters since quantum contextuality is caused by the intrinsic properties of states and measurements. Rather the demon would obtain a more precise detection of contextuality in the system. In a nutshell, quantum contextuality is immune to the demon.

This (Gedanken) experiment explicates the crucial difference between cognitive contextuality and quantum contextuality. And we can conclude that what causes cognitive contextuality is the nature of experimental set-up rather than the nature of reality. The brain is deterministic, and does not involve any probabilities. Still a special experimental-set up can yield contextuality by performing the experiment on different cognitive states, which are the source of probabilities. From another angle:

- Probabilities in quantum contextuality come from the probabilistic nature of reality per se, and thus a single state can yield contextual statistics.
- Probabilities in cognitive contextuality come from the probabilistic nature of collective agents, and thus only collective states can yield contextual statistics.
- The demon encounters contextuality in the first case, and yet can erase contextuality in the second case by fixing parameters involved.

At a mathematical level, cognitive and quantum contextuality basically have the same statistical structure. As to the nature of probabilities, quantum contextuality concerns intrinsic probabilities; probabilities come from the nature of a state measured. Cognitive contextuality concerns extrinsic probabilities; they come from the nature of different states collected together. The difference may thus be conceived of as the difference between intrinsic and extrinsic probabilities. For exact prediction in cognitive science, you have to take into account all relevant hidden contexts. And it may happen that it is practically impossible to consider all relevant contexts. In this case the demon can erase contextuality but the human cannot. Note that, if the demon conducts experiments on different states without fixing parameters, the demon can still observe contextuality.

Now let us wrap up the discussion. Our principal question was: what demarcates cognitive contextuality from physical contextuality? The answers may be summarized as follows:

- The human mind is both classical and quantum: its hardware, the material brain, is classical and deterministic; its software, the cognitive function, can be structurally quantum and indeterministic when measured collectively.
- Whilst physical Bell-type results refute classical realism about single state dynamics, cognitive Bell-type results refute classical realism about collective state dynamics, but not about single state dynamics.
- Physical No-Go results refute (non-contextual) hidden variable theories about single state dynamics. Cognitive No-Go results refute (non-contextual) hidden variable theories about collective state dynamics, but not about single state dynamics.
- Each single state is classical in psychology, and there is no indeterminism, contextuality, or anything quantum within the single state, as shown by the thought-experiment of the experimental Laplace's demon.

Since it is practically impossible to explicate all relevant contexts and to fix all relevant parameters, we obtain probabilities in cognitive experiments, and when the experimental set-up is specially prepared, we obtain contextuality as the violation of Bell-type inequalities. Yet this never means that the brain is indeterministic. Cognitive contextuality is caused by the following two factors: the statistical nature of collective state dynamics and the special structure of experimental set-up. By contrast, quantum contextuality is caused by the following two factors: the statistical nature of single state dynamics and the existence of special states or operations.

In a broader perspective, we can still find commonalities between Bell-type No-Go results in cognitive science and in quantum physics. Quantum cognition arguably shares the same spirit as original Bell-type theorems to the following extent:

- Quantum Physics: classical models of physics do not hold any more. Because there are contextual effects in fundamental reality.
- Quantum Cognitive Science: classical models of cognition and decision do not hold any more. Because there are contextual effects in human reason. We need new models to take into account non-classical features of cognition. And several quantum models have been shown to work for this purpose as we discussed above.

Sen [18] says as follows:

> [T]he puzzle from the point of view of rational behavior lies in the fact that in actual situations people often do not follow the selfish strategy. Real life examples of this type of behavior in complex circumstances are well known, but even in controlled experiments in laboratory conditions people playing the Prisoners' Dilemma frequently do the unselfish thing.

Classical rationality is selfish. Human rationality is unselfish, at least occasionally. Quantum rationality can be unselfish, or at least there is a quantum model of the non-classical statistics of the prisoners' dilemma experiment as discussed above. In a broader perspective, contextuality in cognition does matter for two reasons: it elucidates the nature of human rationality, and it also explicates the embeddedness and situatedness of being, which is relevant in AI as well. Human reason is inherently contextual in actual practice; subtle differences in contexts can make large differences in cognition. Cognition and intelligence do not exist in vacuum; it is embedded, embodied, and situated.

4 Concluding Remarks

In the present article, we have discussed cognitive contextuality phenomena in different forms, including order contextuality, description contextuality, and judgement contextuality in the second section, and probabilistic contextuality in the last section. Quantum cognitive science sheds new light on such contextuality phenomena in cognition. As discussed above, the Quantum Cognition Thesis or Structural QMT may be seen as a modest and less controversial version of Penrose's Quantum Brain Thesis or Material QMT. Quantum cognition research just exploits mathematical models of quantum physics to explain different statistics in different cognitive experiments, just as mathematical economics exploits mathematical models of physical phenomena such as Brownian motion to explain economical phenomena. It does not involve any material quantum effects in the physical brain; it only involves the mathematical structure of quantum theory and its applications to cognitive science.

Nevertheless, it may have substantial consequences, impacting our very conception of nature and ourselves therein. Indeed, it is relevant to the Cartesian dualism. In a sense, it allows us to overcome the Cartesian dualism. Quantum cognitive science tells us that there are commonalities between the laws of nature and the laws of reason. It thus suggests that the severe gap between matter and mind in the Cartesian dualism may be filled in to some extent (because they share certain laws). In another sense, it may account for the reason why the Cartesian dualism must be a dualism. Contextuality across physics and cognitive science tells us that there are substantial commonalities between the laws of nature and the laws of reason; yet, as we discussed above, it also tells us fundamental differences between the science of matter and the science of mind, or why the Cartesian dualism must be a dualism rather than a monism. Understanding the relationships between cognitive and physical contextuality contributes to this broader enterprise of elucidating the Cartesian dualism.

Finally there is a remark about the potential rôle of (phenomenal) free will in cognitive contextuality. Measuring a cognitive agent in different states (e.g., via questions) yields different results (e.g., answers); it looks as if there were some free will involved here, since the same agent voluntarily gives different answers (probably according to different states of it). This (not necessarily real but phenomenal) free will may be a partial source of cognitive contextuality as observed in the cognitive contextuality experiment above, in which there are always two correct answers to given questions, and subjects have to choose one of them by their (phenomenal) free will. Because if subjects always chose one of them then there would be no room for contextuality at all. Thus (phenomenal) free will plays some rôle in the cognitive contextuality phenomenon.

Acknowledgements. This work was supported by JST (grant code JPMJPR17G9).

References

1. Abramsky, S., Brandenburger, A.: The sheaf-theoretic structure of non-locality and contextuality. New J. Phys. **13**, 113036 (2011)
2. Abramsky, S., Hardy, L.: Logical Bell inequalities. Phys. Rev. A **85**, 062114 (2012)
3. Busemeyer, J., Bruza, P.: Quantum Models of Cognition and Decision. Cambridge University Press, Cambridge (2014)
4. Chalmers, D.: Minds, machines, and mathematics. Psyche **2**, 11–20 (1995)
5. Chalmers, D.: The Conscious Mind: In Search of a Fundamental Theory. Oxford University Press, Oxford (1996)
6. Cervantes, V., Dzhafarov, E.: Snow Queen is Evil and Beautiful: Experimental Evidence for Probabilistic Contextuality in Human Choices, Decision, vol. 5, 193–204 (2018)
7. Dreyfus, H.: Why Heideggerian AI failed and how fixing it would require making it more Heideggerian. Philos. Psychol. **20**, 247–268 (2007)
8. Dzhafarov, E., et al.: On contextuality in behavioural data. Philos. Trans. A Math. Phys. Eng. Sci. **374**, 20150234 (2016)
9. Gödel, K.: Collected Works, vol. III. Oxford University Press, New York (1995)
10. Galatos, N., et al.: Residuated Lattices: An Algebraic Glimpse at Substructural Logics. Elsevier Science, San Diego (2007)
11. Lloyd, S.: Programming the Universe. Alfred A. Knopf, New York (2006)
12. Penrose, R.: The Emperor's New Mind. Oxford University Press, Oxford (1989)
13. Penrose, R.: Shadows of the Mind. Oxford University Press, Oxford (1994)
14. Penrose, R.: Consciousness involves noncomputable ingredients. In: The Third Culture: Beyond the Scientific Revolution. Simin & Schuster, New York (1996)
15. Pothos, E., Busemeyer, J.: A quantum probability explanation for violations of 'rational' decision theory, vol. 276, no. 1665 (2009)
16. van Rooij, I.: The tractable cognition thesis. Cogn. Sci. **32**, 939–984 (2008)
17. Rysiew, P.: Epistemic Contextualism, Stanford Encyclopedia of Philosophy. Stanford University, Stanford (2016)
18. Sen, A.: Rational fools: a critique of the behavioral foundations of economic theory. Philos. Public Aff. **6**, 317–344 (1977)
19. Tegmark, M.: Importance of quantum decoherence in brain processes. Phys. Rev. E **61**, 4194 (2000)
20. Tversky, A., Kahneman, D.: Judgments of and by representativeness. In: Judgment under uncertainty: heuristics and biases. Cambridge University Press, Cambridge (1982)
21. Wang, Z., Solloway, T., Shiffrin, R., Busemeyer, J.: Context effects produced by question orders reveal quantum nature of human judgments. Proc. Nat. Acad. Sci. U.S.A. **111**, 9431–9436 (2014)
22. Weir, K.: A reproducibility crisis?, American Psychological Association (2016). Accessed 23 June 2019

Compositionality and Contextuality: The Symbolic and Statistical Theories of Meaning

Yoshihiro Maruyama(✉)

The Hakubi Centre for Advanced Research, Kyoto University, Kyoto, Japan
maruyama@i.h.kyoto-u.ac.jp

Abstract. Compositionality and contextuality give two fundamental principles of linguistic analysis, and yet there is a conflict between them as Burge, Dummett, and others find. Here we aim at elucidating conceptual views underlying their tension in light of both symbolic and statistical paradigms of semantics, arguing, inter alia, that: (i) the conflict is a case of vicious circle analogous to hermeneutic circularity, and may be understood as a tension between symbolic and statistical semantics; (ii) the productivity, systematicity, and learnability of language can be accounted for in accordance with the principle of contextuality as well as compositionality; and (iii) the Chomsky versus Norvig debate on the (symbolic versus statistical) nature of language may be considered a broader manifestation of the tension in the form of the traditional conflict in philosophy between rationalist and empiricist worldviews. We conclude the paper with an outlook for the Kantian synthesis of them, especially the categorical integration of symbolic and statistical AI.

1 Introduction: Two Fregean Principles

Gottlob Frege, one of the founders of symbolic logic and analytic philosophy (especially, philosophy of language), is known for two principles, i.e., the principle of compositionality and the principle of contextuality (even though historical studies [17,25] suggest that both are not strictly rooted in Frege; they then would better be called Fregean principles rather than exactly Frege's). Pelletier [24] summarizes the principle of compositionality as follows: "The Principle of Semantic Compositionality (sometimes called 'Frege's Principle') is the principle that the meaning of a (syntactically complex) whole is a function only of the meanings of its (syntactic) parts together with the manner in which these parts were combined." Another principle of syntactic compositionality may be grounded upon the recursive structure of the syntax of language, which is a character most of the formal and natural languages indeed have (obviously, a text is composed of sentences, which, in turn, are composed of words). Compositionality is considered to be a source of the productivity, systematicity, and learnability of language. Frege [15] says as follows himself:

> I do not believe that we can dispense with the sense of a name in logic; for a proposition must have a sense if it is to be useful. But a proposition consists

G. Bella and P. Bouquet (Eds.): CONTEXT 2019, LNAI 11939, pp. 161–174, 2019.
https://doi.org/10.1007/978-3-030-34974-5_14

> of parts which must somehow contribute to the expression of the sense of the proposition, so they themselves must somehow have a sense. Take the proposition 'Etna is higher than Vesuvius.' This contains the name 'Etna', which occurs also in other propositions, e.g., in the proposition 'Etna is in Sicily'. The possibility of our understanding propositions which we have never heard before rests evidently on this, that we construct the sense of a proposition out of parts that correspond to the words. If we find the same word in two propositions, e.g., 'Etna', then we also recognize something common to the corresponding thoughts, something corresponding to this word. Without this, language in the proper sense would be impossible.

This would suggest that Frege himself grounded the so-called productivity, systematicity, and learnability of language upon the compositional nature of language; the reason why we can understand new sentences we have never heard of before may be accounted for by the compositional nature of language. Concerning these distinctive characteristics of language, Frege [14] also says as follows:

> It is astonishing what language can do. With a few syllables it can express an incalculable number of thoughts, so that even a thought grasped by a terrestrial being for the very first time can be put into a form of words which will be understood by someone to whom the thought is entirely new.

From a historical point of view, however, Janssen [17] argues that Frege did not really endorse the principle of compositionality, but rather endorsed the principle of contextuality (in contrast, Pelletier [25] argues that Frege adopted neither of them). Concerning contextuality, Frege [13] says as follows: "Never ask for the meaning of a word in isolation, but only in the context of a sentence". Wittgenstein [28] also asserts: "Only the proposition has sense; only in the context of a proposition has a name meaning." Notice that this comes from Wittgenstein's early philosophy in *Tractatus*. His later thesis of meaning as use also has some contextualist flavor; our everyday use of language is highly contextual in practice, for example, in that just the same expression can mean different things in different situations or contexts. Wittgenstein's thesis of meaning as use actually inspired what is called statistical distributional semantics, which is probably the most successful contextual semantics in artificial intelligence, especially natural language processing and information retrieval. Yet Chomsky harshly criticizes the nature of contextual statistical semantics in favor of compositional symbolic semantics. Norvig, Google's research director, counters Chomsky in light of the later Wittgensteinian, contextual nature of language. In the following, we fist elucidate the rôles of compositionality and contextuality in our understanding of language, articulating a tension between them, and then shed light on the Chomsky versus Norvig debate from a broader, conceptual perspective.

2 Compositionality Versus Contextuality

Is the nature of meaning in language contextual or compositional? Let us summarize the basic tenets of contextuality and compositionality principles as follows:

- The principle of compositionality: the meaning of a whole (expression) is a function of, and completely determined by, the meaning of its parts (and the syntactical way they are combined together). This is basically atomism; the meaning of atomic expressions recursively generate the meaning of more complex expressions. Compositionality is considered to be a source of the productivity, systematicity, and learnability of language; thanks to compositionality, we can systematically create new expressions.
- The principle of contextuality: the meaning of a word (or more complex expression) is a function of, and can only be determined within, contexts; the meaning of parts depends upon larger wholes surrounding them. This is basically holism about meaning; Quine's holism may be considered a strong form of contextualism. In atomism, parts are prior to wholes, which are secondary; in holism, parts only exist as parts of wholes, which are primary.

The principle of contextuality is often seen as being in conflict with the principle of compositionality (if so Frege cannot endorse both). If meaning is compositional, the meaning of a whole must be determined with reference to the meaning of its parts only, i.e., without reference to anything larger, such as contexts. In contrast, contextuality is a holistic principle. Holism says that a whole cannot be reduced to the mere combination of its parts, whereas the central tenet of compositionality is that this is indeed possible, i.e., the meaning of a whole is composed of that of its parts. Burge [3] is one of the earliest commentators who was clearly aware of the tension in the Fregean philosophy of language:

> It is worth noting that Frege's reasoning here is prima facie *incompatible* with the idea that the notion of the denotation of a term has no other content than that provided by an analysis of the contribution of the term in fixing the denotation (or truth value) of a sentence. The argument presupposes [...] that the notion of term-denotation is more familiar than that of sentence denotation [...]

Dummett [11] also says as follows:

> It was meant to epitomize the way I hoped to reconcile that principle, taken as one relating to sense, with the thesis that the sense of a sentence is built up out of the senses of the words. This is a difficulty which faces most readers of Frege [...] The thesis that a thought is compounded out of parts comes into apparent conflict, not only with the context principle, but also with the priority thesis [...]

Note that Dummett here regards the compositionality of thought as deriving from that of sense (cf. the language of thought hypothesis). The essence of the tension between compositionality and contextuality may be understood in the following manner. The point is whether wholes have to refer to parts or parts have to refer to wholes in oder to determine meaning. Suppose that the two principles are both indispensable in meaning determination. Then, in order to determine meaning, wholes refer to parts, and parts refer to wholes (and mutual reference continues ad infinitum). There is a vicious circle here, and this is essentially an analogue of what is called hermeneutic circularity in the continental tradition of philosophy. We could speculate that, in the analytic tradition, the two principles came to be understood separately in order to keep the theory of meaning immune to vicious circles, and yet in the continental tradition, both of them were taken at face value at the same time, thus leading to the idea of hermeneutic circularity.

Contemporary developments of semantics in artificial intelligence and machine learning allow us to shed new light on the tension between compositionality and contextuality. Formal semantics today are mostly compositional, whether in linguistics, symbolic logic, or the theory of programming languages. In general, giving semantics is understood to be giving a homomorphism from the algebra of grammar to the algebra of meaning while preserving the compositional structure of language, or in terms of category theory, giving a structure-preserving functor from the category of grammar to the category of meaning. In such developments, both syntax and semantics are compositional, and the issue of contextuality in language is only given a marginal place, or considered to be within the realm of pragmatics rather than proper semantics. Nonetheless, what is dominant in artificial intelligence and machine learning is contextual semantics, called distributional semantics, a conceptual (not strictly historical) origin of which is in Wittgenstein's later philosophy, which puts a strong emphasis on the contextual nature of language in practical use. The Wittgenstein's earlier conception of meaning as correspondence with reality has led to developments of logical semantics. The Wittgenstein's later conception of meaning as use (or meaning in the context of use) has led to developments of statistical semantics today. The relevance of Wittgenstein's later philosophy in contextual statistical semantics in artificial intelligence is much less known than the relevance of his earlier philosophy in compositional symbolic semantics in logic and formal linguistics.

Distributional semantics is based upon what is called the distributional hypothesis, which allow us to generate meaning vectors; the following account builds upon Turney-Pantel [27] (a survey paper cited thousands of times):

– The distributional hypothesis: words that occur in similar contexts have similar meanings. The more contexts, the less possibilities of meaning; this is duality between meaning and context.
– In distributional semantics or the vector space model of meaning, words are represented by vectors, the values of which are determined according to the distributional hypothesis. There are various ways to do this. In the simplest implementation, each value represents how many times the word concerned occur in a given context such as document, sentence, and word co-occurring with it (cf. co-occurrence matrices with each column representing a word and each row a context).
– Similarity between words is given by the inner product of the corresponding meaning vectors, or the relative angle between them. If meaning vectors are parallel, for example, the similarity value is one, which means that they have the same meaning.

Distributional semantics thus gives the linear geometry of meaning. From another angle, we may fix a basis of space, i.e., a set of basic meaning vectors, and then the weighted sums of them give all meaning vectors, and the weights are given by the distribution of words in different contexts. The distributional hypothesis may be implemented in different ways to compute the weights in practice. Distributional semantics or the VSM (Vector Space Model) based on the distributional hypothesis has made a great success in natural language processing and information retrieval. Turney-Pantel [27], for example, say as follows:

> The success of the VSM for information retrieval has inspired researchers to extend the VSM to other semantic tasks in natural language processing, with impressive results. For instance, Rapp (2003) used a vector-based representation of word meaning to achieve a score of 92.5% on multiple-choice synonym questions from the Test of English as a Foreign Language (TOEFL), whereas the average human score was 64.5%.

The conflict between compositionality and contextuality exhibits a tension between symbolic and statistical semantics. Symbolic semantics is based upon compositionalism: the meaning of a whole is the sum of the meanings of its parts (and the way they compose). Distributional semantics is based upon holism: the meaning of parts is determined with reference to wholes. Put another way, it is based upon contextualism: the meaning of the parts is determined with reference to different contexts surrounding them. Frege has established the welll-known distinction between sense and reference. "The evening star" and "the morning star" have the same reference and yet their senses are different. Reference is about what words denote. Symbolic semantics is basically concerned with reference. It is compositional, accounting for meaning while respecting the structure of language such as grammar. Sense is about the mode of presentation, i.e., how words are presented in expressions concerned. Words may have sense without reference (e.g., names of fictional characters). Distributional semantics is more concerned with sense (e.g., meaning vectors for the evening and morning stars are different), accounting for meaning via statistical distribution while ignoring the underlying, generative structure. Let us elaborate the last point in the following. Concerning problems of distributional semantics, Turney-Pantel [27] remark as follows:

> Most of the criticism stems from the fact that term-document and word-context matrices typically ignore word order. In LSA, for instance, a phrase is commonly represented by the sum of the vectors for the individual words in the phrase; hence the phrases house boat and boat house will be represented by the same vector, although they have different meanings.

Note that LSA means Latent Semantic Analysis. Natural language processing, including distributional semantics, is mostly based on so-called "bag of words" models, in which expressions larger than words are seen as multisets of words, thus ignoring word order. It should be emphasized here that such bag-of-words models have achieved great successes in artificial intelligence and natural language processing. Interestingly, Landauer [20] argues that 80% of the meaning of English text is due to word choice and the remaining 20% is due to word order. Another problem of distributional semantics or contextual semantics in general is that it is difficult to find right representations of non-contextual terms, such as logical connectives. There is thus a trade-off between symbolic logical semantics and statistical semantics. The philosophical tension between compositionality and contextuality manifests as the technical conflict between symbolic logical semantics and statistical distributional semantics. Yet it should not be impossible to overcome the conflict as we shall discuss later.

Davidson [9] also argues that compositionality is indispensable in order to account for the productivity, systematicity, and learnability of language:

> It is conceded by most philosophers of language, and recently by some linguists, that a satisfactory theory of meaning must give an account of how the meanings of sentences depend on the meanings of words. Unless such an account could be supplied for a particular language, it is argued, there would be no explaining the fact that we can learn the language: no explaining the fact that, on mastering a finite vocabulary and a finitely stated set of rules, we are prepared to produce and to understand any of a potential infinitude of sentences. I do not dispute these vague claims, in which I sense more than a kernel of truth.

It is often argued that the productivity, systematicity, and learnability of language could not be accounted for without compositionality (see, e.g., Szabó [26]). Yet we could argue that contextuality is actually able to account for them as well as compositionality. Contextuality could account for the productivity and creativity of language for the following reason. If there were no contextual use of language allowed in our linguistic practice, then the productivity and creativity of language would be lost or weakened. If we could only mean one thing with one expression, our use of language would be more restricted than it actually is. Put another way, we can creatively use just one expression in different situations to mean different things, and this is possible because language is contextual. Contextuality is a source of productivity and creativity. Contextuality never means that the contextual use of language is random, but it rather means that the laws of language are context-dependent, and the contextual laws of language could account for the systematicity of language. And finally our leaning of language is actually highly contextual in practice. Our symbol grounding in the process of language learning is made by associating linguistic expressions with contexts in the world. Above all, the success of machine learning in natural language processing via contextual semantics shows that the principle of contextuality can account for natural language learning. In terms of empirical power so far, systems based upon statistical contextual semantics outperform those based upon logical contextual semantics. It would thus be fair to say that the productivity, systematicity, and learnability of language, in general, can be accounted for by contextuality as well as compositionality, and in certain particular domains, contextuality-based statistical learning systems can even beat compositionality-based symbolic reasoning systems. In the final section we shall think of possible integrations of symbolic AI and statistical AI in natural language processing; before that, we have a look at an interesting debate on the nature of language.

3 Chomsky Versus Norvig on the Nature of Language

There is an insigitful debate between Noam Chomsky, the father of modern linguistics, and Peter Norvig, Google's research director, concerning the nature of language. It is a battle between the symbolic compositional approach and the statistical contextual approach to natural language. From an even broader perspective, it involves a deeper understanding of the nature of science in general, especially the ultimate purpose of the enterprise of science as we shall see below.

Katz [18] interviews Chomsky at MIT, and Chomsky expresses his scepticism about the statistical approach to natural language as follows:

> [I]f you get more and more data, and better and better statistics, you can get a better and better approximation to some immense corpus of text [...] but you learn nothing about the language.

What Chomsky expects to linguistics or science in general is the systematic account of mechanisms or inner workings that gives us a fundamental understanding of phenomena concerned. A statistical approximation to data, by itself, does not give us any understanding. It may be an excellent simulation, but cannot be a scientific explanation, according to Chomsky. Meaning vectors in distributional semantics are constructed based solely upon statistical information about the way linguistic expressions are used in different contexts, in particular how often they co-occur with other expressions. Representing language via meaning vectors, the computer can then solve different problems, actually in a fairly successful and efficient manner. From Chomsky's point of view, however, this is more like semantic engineering rather than semantic science. He nevertheless argues in the interview that statistical data analytics would even be superior to physics in terms of future prediction. Still, what he prefers is a systematic account and understanding rather than purely predictive power.

Gold [16] summarizes the Norvig versus Chomsky debate as follows:

> Recently, Peter Norvig, Google's Director of Research and co-author of the most popular artificial intelligence textbook in the world, wrote a webpage extensively criticizing Noam Chomsky, arguably the most influential linguist in the world. Their disagreement points to a revolution in artificial intelligence that, like many revolutions, threatens to destroy as much as it improves. Chomsky, one of the old guard, wishes for an elegant theory of intelligence and language that looks past human fallibility to try to see simple structure underneath. Norvig, meanwhile, represents the new philosophy: truth by statistics, and simplicity be damned.

To Norvig, what matters is the proper simulation of actual linguistic practice rather than any idealized theoretical analysis. The emphasis on actual linguistic practice is reminiscent of Wittgenstein's later philosophy, which was strongly against any theorizing tendency, and espoused a nuanced, down-to-earth analysis of our linguistic practice in real world situations. Gold [16] seems to be more sympathetic with the Norvig side than with the Chomsky side:

> Norvig is now arguing for an extreme pendulum swing in the other direction, one which is in some ways simpler, and in others, ridiculously more complex. Current speech recognition, machine translation, and other modern AI technologies typically use a model of language that would make Chomskyan linguists cry: for any sequence of words, there is some probability that it will occur in the English language, which we can measure by counting how often its parts appear on the internet. Forget nouns and verbs, rules of conjugation, and so on: deep parsing and logic are the failed techs of yesteryear.

Statistical models do not much take into account the structure of language such as the grammar of language; nonetheless, a purely statistical analysis of co-occurrence of words yields a surprisingly effective model of natural language. Chomsky [7] himself says at an MIT symposium as follows:

> There's a lot work which tries to do sophisticated statistical analysis, you know bayesian and so on and so forth, without any concern for the actual structure of language, as far as I'm aware that only achieves success in a very odd sense of success. There is a notion of success which has developed in computational cognitive science in recent years which I think is novel in the history of science. It interprets success as approximating unanalyzed data.

He admits that statistical models are superior, in terms of predictive power, to other models including symbolic ones. Chomsky [7] further proceeds:

> You would get some kind of prediction of what's likely to happen next, certainly way better than anybody in the physics department could do. Well that's a notion of success which is I think novel, I don't know of anything like it in the history of science. In those terms you get some kind of successes, and if you look at the literature in the field, a lot of these papers are listed as successes. And when you look at them carefully, they're successes in this particular sense, and not the sense that science has ever been interested in. But it does give you ways of approximating unanalyzed data, you know analysis of a corpus and so on and so forth.

To Chomsky, data science may look like a new kind of post-truth science. But science is not just about understanding; it has played crucial rôles in different aspects of human civilization, for example, in wars. If we argue for Norvig, suppose that a war happens between two countries, Chomsky's and Norvig's nations. Which would win the war in the end? Better predictive power would yield better weapons of destruction. The profound understanding Chomsky aims at may be useless to defend his nation. Yet Chomsky could still argue that a deeper understanding of nature results in revolutionary weapons, just as fundamental physics yielded atomic bombs (or as Turing broke the German code) in the WWII.

There is a well-known argument by Chomsky to show that statistical models cannot capture grammaticalness, which he thinks is of purely symbolic nature. To illustrate his point, Chomsky [4] think of the following two sentences, which look similar in form and yet very different in meaning:

- (1) Colorless green ideas sleep furiously.
- (2) Furiously sleep ideas green colorless.

Chomsky [4] then argues as follows:

> It is fair to assume that neither sentence (1) nor (2) (nor indeed any part of these sentences) has ever occurred in an English discourse. Hence, in any statistical model for grammaticalness, these sentences will be ruled out on identical grounds as equally "remote" from English. Yet (1), though nonsensical, is grammatical, while (2) is not grammatical.

Markie [21] explains Chomsky's idea as follows: "Chomsky argues that the experiences available to language learners are far too sparse to account for their knowledge of their language." Chosmky thinks that, without the innate capacity of reason to generate language, we could not even learn language in the first place; this may be called the innate knowledge thesis. Machine learning could be an alleged counterexample to the thesis, but it is not so decisive at the moment. More than fifty years later, Norvig [23] counters Chomsky [4] as follows:

> I'm not sure what he meant by "any of their parts," but certainly every two-word part had occurred, for example: [...] Pereira (2001) showed that such a model, augmented with word categories and trained by expectation maximization on newspaper text, computes that (1) is 200,000 times more probable than (2). I repeated the experiment, using a much cruder model [...] and found that (1) is about 10,000 times more probable.

With respect to this particular point, Norvig can probably win against Chomsky. Yet at the same time, certain deficiencies of present statistical models of language are going to be remedied and overcome with the help of symbolic methods (see, e.g., Coecke et al. [8]). So the complete theory of language and meaning, if any, may require some symbolic methods as well. There is also a huge cost to pay for statistical models, that is, the interpretability of models. Norvig [23] himself says as follows: "I agree that it can be difficult to make sense of a model containing billions of parameters. Certainly a human can't understand such a model by inspecting the values of each parameter individually". Kuhn-Johnson [19] also assert as follows: "Unfortunately, the predictive models that are most powerful are usually the least interpretable." The present situation in statistical machine learning may be comparable to that in quantum mechanics, which comes with strong predictive power and yet with miserably poor interpretation as Feynman-Mermin say "Shut up and calculate!". Kuhn-Johnson [19] also argue:

> If a medical diagnostic is used for such important determinations, patients desire the most accurate prediction possible. As long as complex models are properly validated, it may be improper to use a model that is built for interpretation rather than predictive performance.

There is surely a trade-off between predictability and interpretability. If the primary concern of science lies in empirical prediction, interpretability ought to be compromised in favor of predictive power. If science is about a rational understanding of nature, on the other hand, interpretability must be maintained at the cost of predictive power. What is at stake here is the very conception of science and its primary aim as an intellectual endeavor of humanity. It should be remarked here that Chomsky and those against statistical data science tend to simplify the actual practice of data science to some extent. They contrast the empirical data-driven approach with the theoretical knowledge-driven approach. Yet data science is not entirely about approximating empirical data. It, for example, attempts to prevent overfitting by introducing domain knowledge in the form of what are called regularization terms or specific priors. It does care about the interpretability of models, and hence the recent trend of interpretable machine learning. The methodology of data science is not purely empiricist but partly rationalist, even if the resulting knowledge is purely empiricist with no explication of underlying mechanisms. A more practical aspect of the cost that complex models have to pay is as follows. Bourguignat [2] says:

> In real organizations, people need dead simple story-telling — Which features are you using? How your algorithms work? What is your strategy? etc. ... If your models are not parsimonious enough, you risk to the audience confidence. Convincing stakeholders is a key driver for success, and people trust what they understand. What's more, at the end of the day, the ultimate goal of the data science work is to put a model in production. If your model is too complicated, this will turn out to be impossible or, at least, very difficult.

Norvig [23] summarizes the points of Chomsky's critique of statistical semantics:

- "Statistical language models have had engineering success, but that is irrelevant to science."
- "Accurately modeling linguistic facts is just butterfly collecting; what matters in science (and specifically linguistics) is the underlying principles."
- "Statistical models are incomprehensible; they provide no insight."

Actually, the metaphor of butterfly collecting as non-science is Chomsky's favorite, and has an older origin. In 1979 Chomsky [6] says:

> You can also collect butterflies and make many observations. If you like butterflies, that's fine; but such work must not be confounded with research, which is concerned to discover explanatory principles of some depth and fails if it does not do so.

Chomsky's point may be clarified by having a look at a related case in the history of science, that is, the shift from Ptolemy's predictive model of the heavens to Copernicus's. In terms of predictive power or approximation of data, Copernicus's model did not really outperform Ptolemy's, and Copernicus's nonetheless won the race of science. This is well known among historians of science (see, e.g., Evans [12]). Machine learning papers often claim their methods outperform state-of-the-art methods, in the spirit of what Chomsky calls a new notion of success in science. If the aim of science is at approximating empirical data, the shift from Ptolemy's to Copernicus's model was unreasonable in light of the very aim of science. In reality, there is some other parameter in the aim of science. Science is concerned with a rational understanding of nature as well as approximating and predicting empirical data. Chomsky aims at the former while Norvig at the latter. The Chosmky versus Norvig debate, therefore, may be seen as a revival of the classic debate between rationalism and empiricism in the context of the contemporary science of language and intelligence. Indeed, Chomsky [5] asserts that he has a "rationalist conception of the nature of language", and contrasts it with an empiricist conception as follows:

> Furthermore, I employed it [...] to support what might fairly be called a rationalist conception of the nature of language [...] In sharp contrast to the rationalist view, we have the classical empiricist assumption that what is innate is (1) certain elementary mechanisms of peripheral processing (a receptor system), and (2) certain analytical machanisms or inductive principles or mechanisms of association.

As the above quote shows, Chosmky is actually aware that what is at stake here is the rationalist versus empiricist conception of language. In the history of philosophy, Kant aimed at reconciling the two camps; Kantian AI (rather than so-called Heideggerian AI) may be a solution to the discrepancy between the two camps. There are recent developments to integrate symbolic AI and statistical AI via different methods; the integrations of symbolic AI and statistical AI may be able to embody Kantian AI, thus allowing us to make predictive power compatible with interpretability. We shall touch upon them in the next section.

Norvig [23] concludes the discussion with the following remark on the contingent nature of language:

> [L]anguages are complex, random, contingent biological processes that are subject to the whims of evolution and cultural change. What constitutes a language is not an eternal ideal form, represented by the settings of a small number of parameters, but rather is the contingent outcome of complex processes. Since they are contingent, it seems they can only be analyzed with probabilistic models.

In his view, language is contingent by nature, and thus statistical models are (not just useful but also) necessary for the analysis of language. By contrast, Chomsky wants to explicate the "eternal ideal form" of language, which is universal rather than contingent. Is the ultimate nature of language contingent or universal? A possible route to do justice to both would be to argue that the surface structure of language is contingent, and yet the core structure of it is universal. Compositional symbolic semantics focuses upon the universal, core structure of language, and contextual distributional semantics upon the contingent, surface structure of language. In this way we could peacefully reconcile the two opposing views.

The tension between Chomsky's and Norvig's views, or between the compositional symbolic approach and the contextual distributional approach, can be understood in a broader context of philosophy of language. Wittgenstein's early philosophy may be seen as based on compositionalism, and Wittgenstein's late philosophy as based on contextualism. The latter sees meaning as rooted in the "natural history" of linguistic practice (and so dynamic and contingent), whereas the former sees meaning as arising from the picture-theoretical correspondence between language and reality (and thus static and universal). The tension between Chomsky's and Norvig's views is also comparable to the conflict between the realist and antirealist conceptions of meaning in Dummett's terms. Compositional symbolic semantics is usually referential, and presupposes reality outside language, which accommodates denotations to interpret language, and it may be seen as realism. Contextual distributional semantics does not presuppose anything outside language, deriving meaning vectors just from contexts within language. It may thus be seen as antirealism. The autonomy of language as independent of reality is emphasized in Wittgenstein's later philosophy [29]: "The words are not a translation of something else that was there before they were." Meaning comes not from reality, but from the internal structure of language.

Concerning the nature of science rather than language in particular, the tension between Chomsky's and Norvig's views is basically the same as the tension between empiricism and rationalism. As a historical remark, Maxwell had an interesting idea of the relationships between empiricism and rationalism. On one hand, Maxwell [22] says as follows (in his 1850 letter to Lewis Campbell): "the true Logic for this world is the Calculus of Probabilities." Maxwell [22] thinks, perhaps on the basis of the British tradition of empiricism, that human knowledge as a whole rests upon the faculty of sensibility as well as the faculties of reason and understanding, and so it is probabilistic by nature:

> [A]s human knowledge comes by the senses in such a way that the existence of things external is only inferred from the harmonious (not similar) testimony of the different senses, understanding, acting by the laws of right reason, will assign to different truths (or facts, or testimonies, or what shall I call them) different degrees of probability.

Maxwell [22], on the other hand, admits certainty, immutability, and universality, not in empirical experiments, but in things themselves, which can go beyond the world of statistical correlations (though it is not clear whether he thought there was any finitary, human way to access such a world of certainty):

> [O]ur experiments can never give us anything more than statistical information [...] But when we pass from the contemplation of our experiments to that of the molecules themselves, we leave a world of chance and change, and enter a region where everything is certain and immutable.

In such a manner, Maxwell attempted to reconcile the empiricist and rationalist views of science, and it could also allow us to reconcile Chomsky's universalist view and Norvig's probabilist view by regarding the former as concerned with reality per se and the latter as concerned with our surface knowledge of it.

4 Concluding Remarks: Towards the Kantian Synthesis

The principle of compositionality gives the virtually dominant paradigm of semantics in formal linguistics, symbolic logic, and the theory of programming languages. And the principle of contextuality looks marginal from this perspective. Janssen [17] says: "it is not possible to accept both principles at the same time: there is a conflict between them. This is underscored by the fact that the only modern theory which obeys contextuality [...] is proud of being non-compositional and uses this feature to defeat other theories." But what he has in mind is Hintikka's game semantics only. There is another highly successful contextual semantics in artificial intelligence, that is, statistical distributional semantics, which is actually the dominant paradigm of natural language processing, widely used in practical applications. The principle of contextuality gives a mainstream approach to language in artificial intelligence, allowing computer systems to actually learn different features of meaning in natural language. They can even solve some TOEFL problems better than humans. Let us finally discuss whether the two ideas can be combined and integrated into a coherent whole. From a conceptual point of view we can argue that, if language is contextual and contingent in its surface and yet compositional and universal in its core, they are just concerned with different sides of the same coin, and compatible with each other (put another way, neither Chosmky nor Norvig is wrong, and we can do justice to both). There are indeed developments to embody such an idea. One of them relies upon category theory to integrate the compositional and contextual theories of language and meaning (Coecke et al. [8]; one of the earliest integrations). It may be extended to a global paradigm of AI aiming at the categorical integration of symbolic and statistical AI in general, which can, in turn, be regarded as part of an even more global paradigm of categorical unified science qua pluralistic unified science. A similar approach is given by Baroni et al. [1], who adopt "the idea that word meaning can be approximated by the patterns of co-occurrence of words in corpora from statistical semantics, and the idea that compositionality can be captured in terms of a syntax-driven calculus of function application from formal semantics." Domingos et al. [10] also aim to unify logical and statistical AI in a general setting beyond language. All this may be seen as an attempt to achieve the ultimate goal of making empirical performance (Norvig) compatible with systematic understanding (Chomsky). Yet no one (and no machine) can really tell the future. We may possibly end up with the situation that theoretically inclined researchers like Chomsky stick to compositional symbolic semantics, and empirically inclined researchers like Norvig to contextual distributional semantics. In that case it would be justified to conclude that compositionality and contextuality are in true conflict with each other.

References

1. Baroni, M., et al.: Frege in space: a program of compositional distributional semantics. Linguist. Issues Lang. Technol. **9**, 5–110 (2014)

2. Bourguignat, C.: Interpretable VS Powerful Predictive Models: Why We Need Them Both, Medium, 17 September 2014. Accessed on 16 July 2019
3. Burge, T.: Truth, Thought, Reason: Essays on Frege. Clarendon, Oxford (2005)
4. Chomsky, N.: Syntactic Structures. Mouton & Co, Berlin (1957)
5. Chomsky, N.: Recent contributions to the theory of innate ideas. In: Stitch, S. (ed.) Innate Ideas. California University Press, Berkeley (1975)
6. Chomsky, N., Ronat, M.: Language and Responsibility: Based on Conversations with Mitsou Ronat. Pantheon Books, Paris (1979)
7. Chomsky, N.: Keynote Panel: The Golden Age - A Look at the Original Roots of Artificial Intelligence, Cognitive Science, and Neuroscience, MIT Symposium on Brains, Minds, and Machines. Accessed on 23 June 2019
8. Coecke, B., et al.: Mathematical foundations for a compositional distributional model of meaning. Linguist. Anal. **36**, 345–384 (2010)
9. Davidson, D.: Truth and Meaning. Synthese **17**, 304–323 (1967)
10. Domingos, P., et al.: Unifying logical and statistical AI. In: Proceedings of the 31st Annual ACM/IEEE Symposium on Logic in Computer Science, pp. 1–11 (2016)
11. Dummett, M.: The Interpretation of Frege's Philosophy. Duckworth, London (1981)
12. Evans, J.: The History and Practice of Ancient Astronomy. Oxford University Press, Oxford (1998)
13. Frege, G.: The Foundations of Arithmetic. Basil Blackwell, Oxford (1953). (originally 1884)
14. Frege, G.: Compound thoughts. Mind **72**, 1–17 (1963). (originally 1923)
15. Frege, G.: The Philosophical and Mathematical Correspondence. University of Chicago Press, Chicago (1980)
16. Gold, K.: Norvig vs. Chomsky and the Fight for the Future of AI, TOR.COM, 21 June 2011. Accessed on 26 June 2019
17. Janssen, T.: Frege, contextuality and compositionality. J. Logic Lang. Inf. **10**, 87–114 (2001)
18. Katz, Y.: Noam Chomsky on Where Artificial Intelligence Went Wrong, The Atlantic, 1 November 2012. Accessed on 23 June 2019
19. Kuhn, M., Johnson, K.: Applied Predictive Modeling. Springer, New York (2013). https://doi.org/10.1007/978-1-4614-6849-3
20. Landauer, T.: On the computational basis of learning and cognition: arguments from LSA. Psychol. Learn. Motiv. **41**, 43–84 (2002)
21. Markie, P.: Rationalism vs. Empiricism, Stanford Encyclopedia of Philosophy (2017)
22. Maxwell, J.C.: The Scientific Letters and Papers of James Clerk Maxwell: 1846–1862. Cambridge University Press, Cambridge (1990)
23. Norvig, P.: On chomsky and the two cultures of statistical learning. Berechenbarkeit der Welt?, pp. 61–83. Springer, Wiesbaden (2017). https://doi.org/10.1007/978-3-658-12153-2_3
24. Pelletier, F.J.: The principle of semantic compositionality. Topoi **13**, 11–24 (1994)
25. Pelletier, F.J.: Did frege believe frege's principle? J. Logic Lang. Inf. **10**, 87–114 (2001)
26. Szabó, Z.G.: Compositionality, Stanford Encyclopedia of Philosophy (2017)
27. Turney, P., Pantel, P.: From frequency to meaning: vector space models of semantics. J. Artif. Intell. Res. **37**, 141–188 (2010)
28. Wittgenstein, L.: Tractatus Logico-Philosophicus. Humanities Press, New York (1961). (originally 1922)
29. Wittgenstein, L.: Zettel. University of California Press, California (1967)

Towards a Logic of Epistemic Theory of Measurement

Claudio Masolo and Daniele Porello(✉)

Laboratory for Applied Ontology, ISTC -CNR, Trento, Italy
daniele.porello@cnr.it

Abstract. We propose a logic to reason about data collected by a number of measurement systems. The semantic of this logic is grounded on the epistemic theory of measurement that gives a central role to measurement devices and calibration. In this perspective, the lack of evidences (in the available data) for the truth or falsehood of a proposition requires the introduction of a third truth-value (the undetermined). Moreover, the data collected by a given source are here represented by means of a possible world, which provide a contextual view on the objects in the domain. We approach (possibly) conflicting data coming from different sources in a social choice theoretic fashion: we investigate viable operators to aggregate data and we represent them in our logic by means of suitable (minimal) modal operators.

Keywords: Measurement theory · Social-choice theory · Three-valued logic · Logic of evidence · Epistemic logic

1 Introduction

The need for grounding rational beliefs, for understanding what supports the epistemic states of agents, is vastly acknowledged. Grounding is particularly relevant for scientific claims that are usually justified in terms of observations and of empirical data. In this context, data cannot be private, they must be shared and trusted by different subjects at different times and in different places. According to [15], the *objectivity* and the *inter-subjectivity* of the scientific results distinguish *measurement* from *evaluation*. Objectivity concerns the independence of the measurements of a given property (of a measurand) of other properties, measuring devices, and environmental conditions. Inter-subjectivity regards the sharing of measurements and it is achieved by establishing measurement standards and calibration procedures for devices. Measurement theories play then a central role for the collection and the sharing of trusted data, a pre-requisite for grounding empirical science. Our aim is to develop a logic that explicitly represents how propositions connect to empirical data by exploiting the *epistemic measurement theory* (EMT) introduced in [9,11,15,16].

Epistemic logics [26] and evidence logics [3] have already considered the grounding of epistemic states. Evidences for beliefs are encoded as sets of (or, in neighborhood semantics, as families of sets of) *possible worlds* that, in these logics, reduce in fact to plain and unstructured indexes.

G. Bella and P. Bouquet (Eds.): CONTEXT 2019, LNAI 11939, pp. 175–188, 2019.
https://doi.org/10.1007/978-3-030-34974-5_15

By contrast, in our proposal, possible worlds are structured entities that explain the contextual nature of epistemic measurements. Possible worlds, called here *states*, are then characterised in terms of EMT. A world provides a group of *datasets*, i.e., a set of data (about the objects in the domain) collected by batteries of measurement systems. Intuitively, a world gathers the (possibly *partial*) information about the objects in the domain supplied by a given source. The truth of a proposition at a certain state is then *contextual* because it depends on the data collected by the source, i.e., on the measurement systems available at that source and on the performed measurements. As we shall see, the partiality of data forces the use of a three-valued semantics. A given dataset may not contain enough information to establish neither the truth nor the falsity of a proposition. A third truth-value, the *undetermined* (or unknown) is required. The interpretation of the undetermined value that we endorse here is close to the one introduced by Kleene [13] to represent the situation where an algorithm does not terminate yielding a 'true' or 'false' output.

Ideally, objectivity and inter-subjectivity guarantee that all the data collected by calibrated devices are consistent and sharable without errors. However, in a realistic scenario, measurement devices may be used in unsuitable environmental conditions or following wrong procedures. They may also malfunction or lose calibration during their life (typically, devices are not re-calibrated at every use). These issues may especially be present in (*i*) large-scale and distributed collaborative science, that often relies on (user-generated) data which are collected, for instance, through sensors in mobile and ubiquitous devices; and (*ii*) scientific endeavour that relies on tests, e.g. neuropsychological, clinical or behavioural, where the scores of the tests are the result of very complex procedures that aggregate heterogeneous measurements. Because of these complications, conflicting datasets may exist, i.e., the sources of data may disagree.

To approach conflicting sources, we extend our logic with modal operators inspired by social choice theory and judgment aggregation [14]. These operators represent different strategies to aggregate the heterogeneous data collected by several (possibly conflicting) sources. We shall see that a careful analysis of such procedures is required, as some procedures do not guarantee the consistency of the aggregated data. Moreover, in a scientific scenario, the data collected by the sources are often *sparse*, i.e., typically, only few sources have information about the same objects. This scenario is different from the one of judgement aggregation, where 'abstainers' are usually not the majority. The standard aggregation procedures need then to be adapted to take into account what are the sources that have relevant information about a given proposition.

By aggregating the information coming from heterogeneous epistemic contexts, these modal operators introduce a *de-contextualisation*, as intended in [19], of the truth of propositions. Aggregated statements are cross-contextual, they integrate (following a given strategy) the perspectives of several epistemic contexts (see Sect. 3.4 for more details).

The paper is organised as follows. Section 2 introduces EMT and precisely defines states as datasets collected by batteries of measurement systems.

Section 3 introduces the logical framework for representing and reasoning about data collected by a single source and the modal aggregation operators. Moreover, it discusses the contextual nature of the measurement statements. Section 4 concludes the paper.

2 Measurement Systems and Datasets

According to the *Representational Measurement Theory* (RMT) [25], measurement consists in building a mapping from an *empirical* relational structure to a *numerical* relational structure such that the relations among numbers represent the empirical relations among objects. Despite the precise and deep mathematical treatment, RMT seems too abstract to be used in empirical contexts [9]. One first problem concerns the fact that RMT focuses on quantitative measurement (interval or ratio properties). Secondly, RMT considers the empirical relational structure (and the axioms governing it) as given. The problems of founding measurement on empirical methods and of data sharing are not addressed.

The perspective of *qualitative* classification (nominal or ordinal properties), that plays a fundamental role in disciplines like psychology, medicine, or sociology, has been addressed by the theory of *weak measurement* introduced by Finkelstein [8] and further elaborated by Mari [15]. In this weak perspective, measurement becomes "uncorrelated with quantification: the measurability of a property is a feature derived from experiment, not algebraic constraints" [15, p. 2894]. Thus, not all the empirical structures are mapped into numerical structures, they can also be mapped into symbolic classification systems, where symbols may have a weak organization, not necessarily an algebraic one. Moreover, Frigerio and colleagues [9] follows this weak perspective by presenting a formal model that grounds measurement on *measurement systems* (MSs). Roughly speaking, an MS is a (physical) device that is able to interact with the system under measurement (SUM)[1] and that is characterized by a set of empirically discernible states and relations to which symbols are conventionally associated. The output of the interaction between an MS and a SUM is a piece of (symbolic) information. While weak measurement (as well as RMT) assumes SUMs to be states or individual properties (tropes) of objects, following [16], we do not commit to these kinds of entities and consider SUMs to be objects, i.e., MSs are mediators between (external) objects and measurements, sorts of physical embodiments of the classification systems ([17] provides additional details).

Definition 1 (Measurement system). *A measurement system is a tuple* $\mathcal{M} = \langle d, O, \mathcal{E}, \kappa, \mathcal{S}, \lambda \rangle$ *where:*

- *d is a device (usually a physical object);*
- *O is the set of objects the device d is able to interact with;*

[1] MSs are "provided with instructions specifying how such interaction must be performed and interpreted" [9].

– $\mathcal{E} = \langle U, R_1, \ldots, R_n \rangle$ *is an empirical structure, i.e., a structure where* U *is the set of empirically discernible states of all possible complex systems resulting from the interaction of any object* $o \in O$ *with* d *(noted by* $d \bullet o$*) and* R_i *are empirically discernible relations among the states in* U*;*[2]
– $\kappa : O \rightarrow U$ *is the interaction function that associates to an object* $o \in O$ *the state of the complex system* $d \bullet o$*;*
– $\mathcal{S} = \langle S, RS_1, \ldots, RS_n \rangle$ *is a symbolic structure, i.e., a structure where* S *is a set of symbols and* RS_i *are relations defined on* S*;*[3]
– $\lambda : U \rightarrow S$ *is the symbolization function, a one-to-one function between* U *and* S*, such that* $R_i(u_1, \ldots, u_n)$ *iff* $RS_i(\lambda(u_1), \ldots, \lambda(u_n))$*.*

The states U (through the interaction function κ) induce a partition on the set of objects O: $o \approx o'$ iff $\kappa(o) = \kappa(o')$ (U establishes the *resolution* of d). Similarly, each R_i (and each RS_i) induces a relation on objects: $\bar{R}_i(o_1, \ldots, o_n)$ iff $R_i(\kappa(o_1), \ldots, \kappa(o_n))$. The empirical structure is here determined by the MS that *induces* a structure on objects (by interacting with them), i.e., an MS (and the measurement procedure) provides an empirical access point to the world. The symbolization function and the symbolic structure allow to abstract from the empirical structure, they provide a symbolic encoding, i.e., $\mathcal{S}$ contains the whole information in $\mathcal{E}$ but in a *communicable* and *manipulable* form. Different measurement systems can then share the same symbolic structure allowing for alternative ways to measure the same kind of properties.

As we discussed in the introduction, the objectivity and inter-subjectivity of data is obtained via measurement standards and calibration. A measurement standard establishes a set of physical (or theoretical) objects that is isomorphic to the symbols in the classification system, i.e., they are the perfect realization of the properties represented by the symbols. Calibration determines a one-to-one correspondence between the (relations between the) positions of the pointers of an MS and the (relations between the) properties in the classification system, i.e., the positions of the pointers and the output symbols stand for properties, they have a meaning.[4] Thus, a measurement standard determines a classification system while calibration individuates all the MSs that can be (interchangeably) used to classify objects in this system.

As we anticipated in the introduction, we do not commit to perfect calibration. The MSs have been calibrated, but nothing guarantees that, at every time data are collected, the MSs are correctly used and still calibrated. The outputs of a single MS have a shared and precise meaning and are consistent, but conflicting data collected by different MSs may exist.

A *dataset* groups all the measurements collected by a single MS.

Definition 2 (Dataset). *A dataset is a couple* $\mathcal{D} = \langle \mathcal{M}, D \rangle$ *where:*

– $\mathcal{M} = \langle d, O, \mathcal{E}, \kappa, \mathcal{S}, \lambda \rangle$ *is a measurement system;*

[2] Notice that $\mathcal{E}$ refers to potential interactions with objects, i.e., by abstracting from specific objects, it depends only on the (physical) structure of d.
[3] Differently from RMT, $\mathcal{S}$ is not necessarily a *numerical* structure.
[4] See [16] for the formal details.

– *D is the set of data collected by $\mathcal{M}$, i.e., the (possibly empty) set of pairs $\langle o, m\rangle$ such that $\lambda(\kappa(o)) = m$.*

Note that D is consistent by construction, it is not possible to have $\langle o, m_1\rangle$, $\langle o, m_2\rangle \in D$ with $m_1 \neq m_2$.

An MS is able to classify objects along a single classification system. However, one can have data that concern different properties of the same object.[5] The notion of *measurement battery* (MB) extends the one of MS by considering sets of MSs able to classify objects along several classification systems (e.g., a thermometer together with a scale, a ruler, etc.).[6]

Given a set $\{\mathcal{M}_1, \ldots, \mathcal{M}_n\}$ of MSs, we denote by S_i the set of symbols of $\mathcal{M}_i$ and by $\mathbb{RS}_i^h$ the set of h-ary relations on S_i-symbols in $\mathcal{M}_i$.

Definition 3 (Measurement battery). *A measurement battery is a finite set of MSs $\mathfrak{M} = \{\mathcal{M}_1, \ldots, \mathcal{M}_n\}$ such that, for all $\mathcal{M}_i = \langle d_i, X_i, \mathcal{E}_i, \kappa_i, \mathcal{S}_i, \lambda_i\rangle$ and $\mathcal{M}_j = \langle d_j, X_j, \mathcal{E}_j, \kappa_j, \mathcal{S}_j, \lambda_j\rangle \in \mathfrak{M}$ with $i \neq j$, we have that:*

1. *$X_i = X_j = O$, i.e., every $\mathcal{M}_i \in \mathfrak{M}$ is about the same set of objects O; and*
2. *$S_i \cap S_j = \emptyset$, i.e., the symbols of the MSs in $\mathfrak{M}$ are disjoint.*

A *state* collects all the datasets provided by the MSs in an MB. MBs and states are the multidimensional counterparts of, respectively, MSs and datasets.

Definition 4 (State). *A state s is a set of datasets s.t. their respective MSs form a MB, i.e., $s = \{\langle \mathcal{M}_1, D_1\rangle, \ldots, \langle \mathcal{M}_n, D_n\rangle\}$ where $\{\mathcal{M}_1, \ldots, \mathcal{M}_n\}$ is a MB.*

Note that, since each MS is consistent, the condition 2 in Definition 3 guarantees the consistency of states (i.e. it is not possible to have $\langle o, m_1\rangle$, $\langle o, m_2\rangle$ such that $m_1 \neq m_2$ and m_1, m_2 are in the same symbolic structure).

Finally, we introduce a finite set of states $\mathfrak{S}$ to model data coming from distinct measurement batteries. Single states do not contain any contradictory measurement, while different states can disagree. This disagreement is due to the use of different MSs that classify objects along the same system of properties.

3 A Logic for Measurement

We present a predicative modal logic to represent and reason about the data provided by a number of MBs. We shall see that a single state provides sufficient information to define the semantics of logical connective and quantifiers. We start by defining the predicative structure, then we shall discuss several modal operators that may be used to represent aggregations of MBs.

5 In terms of the theory of *conceptual spaces* [10], single classification systems correspond to the *domains* of a conceptual space (e.g., color, taste, shape, temperature, etc.), while the whole space requires the composition of several systems.

6 It is possible to extend the notion of MB to allow to have different MSs relative to the same classification system, e.g., different scales, different thermometers, etc.

3.1 Syntax and Semantics

The vocabulary of our predicative language $\mathcal{L}$ contains: a set of individual constants $\mathbf{C} = \{c_1, c_2, \dots\}$, a set of individual variables $\mathbf{V} = \{x_1, x_2, \dots\}$, a set of n-ary ($n \geq 1$) predicates $\mathbf{R} = \{R^1_1, R^1_2, \dots, R^2_1, R^2_2, \dots, R^j_i, \dots\}$, the set $\{\neg, \wedge, \vee, \rightarrow\}$ of connectives, and the set $\{\forall, \exists\}$ of quantifiers. The set of atomic formula *Atom* of $\mathcal{L}$ is defined as follows: $Q^j_i(a_1, \dots, a_j) \in Atom$ iff $Q^j_i \in \mathbf{R}$ and $a_1, \dots, a_j \in \mathbf{C}$. This definition inductively extends to the full predicative language as usual.

Given a state $s = \{\langle \mathcal{M}_1, D_1 \rangle, \dots, \langle \mathcal{M}_l, D_l \rangle\}$, we denote by $\delta(s)$ the set of all measurements that are present in some dataset of s, i.e., $\delta(s) = D_1 \cup \dots \cup D_l$.

Definition 5 (Measurement model). *A* measurement model *for $\mathcal{L}$ is a tuple $M = \langle s, \varepsilon, \iota \rangle$ where:*

- *s is a state concerning the set of objects O, i.e. a set $\{\langle \mathcal{M}_1, D_1 \rangle, \dots, \langle \mathcal{M}_l, D_l \rangle\}$;*
- *ε is a function that maps individual constants into objects, $\varepsilon : \mathbf{C} \rightarrow O$;*
- *ι is a function that maps:*
  - *unary predicates into symbols of the MSs in the MB in s: $\iota : \mathbf{R^{(1)}} \rightarrow S_1 \cup \dots \cup S_l$;*
  - *n-ary ($n \geq 2$) predicates into n-ary relations of the MSs in the MB in s: $\iota : \mathbf{R^{(n)}} \rightarrow \mathbb{RS}^n_1 \cup \dots \cup \mathbb{RS}^n_l$.*

The domain of the interpretation is then given by the set of objects of the state s, i.e., by O, the interpretation of the individual constants is provided by ε, and the interpretation of predicates is provided by ι. We shall introduce the interpretation for the variables when discussing the semantics of quantifiers.

The valuation function $||\cdot||_M$ maps formulas to a suitable set of truth-values.

Since unary predicate and n-ary ($n \geq 2$) relations have slightly different interpretations, we present their semantics separately. Moreover, to reflect a verificationist perspective on truth-making, we assume three truth values $\{t, f, u\}$. Intuitively, *true* means that there exists a verifier of ϕ in $\delta(s)$, *false* means that there exists a falsifier of ϕ in $\delta(s)$, and *undetermined* means that there is neither a verifier nor a falsifier of ϕ in $\delta(s)$.

The semantics for atomic formulas involving unary predicates is defined as:

- $||P(a)||_M = t$ iff $\langle \varepsilon(a), \iota(P) \rangle \in \delta(s)$;
- $||P(a)||_M = f$ iff there exists $\langle \varepsilon(a), m \rangle \in \delta(s)$ with m and $\iota(P) \in S_i$, for some i, and $m \neq \iota(P)$;
- $||P(a)||_M = u$ iff otherwise.

A falsifier of $P(a)$ is then a measurement of the object $\varepsilon(a)$ along the same system of properties of $\iota(P)$. For instance, to falsify *1KG*(a), among the data available in s, one needs to find a *weight*-measurement of $\varepsilon(a)$ with a result different from $\iota($*1KG*$)$. We can follow this idea because the symbols in the S_i are considered as mutually exclusive, i.e., in principle, the measurements of a single object along a given classification system cannot result in different outputs.

The case of n-ary relations, for $n \geq 2$, is captured by the following definition:

- $||R(a_1, \ldots, a_n)||_M = t$ iff for $1 \leq i \leq n$, there exist $\langle \varepsilon(a_i), m_i \rangle \in \delta(s)$ such that $\langle m_1, \ldots, m_n \rangle \in \iota(R)$;
- $||R(a_1, \ldots, a_n)||_M = f$ iff for $1 \leq i \leq n$ there exist $\langle \varepsilon(a_i), m_i \rangle \in \delta(s)$ such that $m_i \in S_l$, $\iota(R) \in \mathbb{RS}_l^n$, and $\langle m_1, \ldots, m_n \rangle \notin \iota(R)$;
- $||R(a_1, \ldots, a_n)||_M = u$ iff otherwise.

Negation, conjunction and disjunction are defined according to Kleene three-valued semantics, see Table 1.a–c. Intuitively, $\neg A$ is true (false) only when there exist data that support the falsity (truth) of A. When A is undetermined also $\neg A$ is undetermined, i.e., when we lack support for the falsity or truth of A, we also lack support for the falsity or truth of $\neg A$. The data that falsify one conjunct are enough to falsify the whole conjunction, while when one conjunct is undetermined. Dual considerations hold for the disjunction. Implication is more problematic. In Kleene logic, the implication is defined, as usual, by $\neg A \vee B$. In this case, when both A and B are undetermined, according to Table 1. a&c, $A \rightarrow B$ is also undetermined. This seems empirically plausible but it clashes with the idea that the logical principle $A \rightarrow A$ holds even when A is undetermined. Moreover, the refusal of $A \rightarrow A$ results in a very weak logic. Thus, to obtain a well-behaved logical implication, three-valued logics usually add the Lukasiewicz implication that has the truth-table in Table 1. d, cf. [2]. With respect to the classical definition of the implication, the only difference is that when both A and B are undetermined, $A \rightarrow B$ is true rather than undetermined.

Table 1. Truth-tables for connectives

(a)

$\neg$	t	u	f
	f	u	t

(b)

$\wedge$	t	u	f
t	t	u	f
u	u	u	f
f	f	f	f

(c)

$\vee$	t	u	f
t	t	t	t
u	t	u	u
f	t	u	f

(d)

$\rightarrow$	t	u	f
t	t	u	f
u	t	t	u
f	t	t	t

Let $A(x)$ be a formula with x among its free variables and let $\sigma : \mathbf{V} \rightarrow O$ an assignment of the variables to the elements of O.

- $||\forall x A||_{M,\sigma} = t$ iff for every $d \in O$, $||A||_{M,\sigma(x/d)} = t$;
- $||\forall x A||_{M,\sigma} = f$ iff there is a $d \in O$ such that $||A||_{M,\sigma(x/d)} = f$;
- $||\forall x A||_{M,\sigma} = u$ iff otherwise.

The existential quantifier is defined by $\exists x A(x) \leftrightarrow \neg \forall x \neg A(x)$. We say that a formula ϕ is satisfiable if there exists a model M such that $||\phi||_M = t$. A formula ϕ is *valid* iff for every model M, $||\phi||_M = t$.

The Hilbert system for propositional first-order Lukasiewicz three-valued logic is proposed in [2,12].

3.2 Dataset Aggregation and Modal Operators

A single state provides sufficient information to express and to reason about the formulas that are made true by a single MB. In this section, we aim at addressing possible disagreements about the data provided by distinct MBs by defining modalities that aggregate datasets. Each MB (its associated state) is then viewed as a source of data to be submitted to an aggregation procedure that has the task of integrating datasets and solving possible inconsistencies.

We assume a finite set of N states $\mathfrak{S}$ and we extend $\mathcal{L}$ by adding a number of modal operators $\Box_F$ that depend on a certain aggregation function F. An *aggregation function* is a function F that maps N-tuples of truth-values associated to formulas to a collective/aggregated assignment of truth-values to that formula, i.e., $F : \{t, u, f\}^N \rightarrow \{t, u, f\}$. By defining aggregators by means of F, we are assuming that the method for aggregation is the same for every statement (a property called *neutrality* in judgment aggregation) and that the method is the same for any tuple of truth values (*independence*), cf. [7]. Moreover, we are defining aggregators on three possible truth-values, thus the standard definitions of the theory of judgment aggregation have to be adapted, cf. [6,21,23].[7]

The $\Box_F$ operators aggregate the truth-values of the formulas that hold in the various states, thus no new formula can be introduced in the aggregated outcome. We follow here a *coarse*, rather than *fine-grained*, aggregation of formulas (cf. [17,23]), where in fact each collectively accepted formula must be accepted by at least one state. Coarse aggregations often fail to elect an aggregated formula that is a good trade off between the individual sources. For instance, suppose that state 1 makes true $\mathit{1KG}(a)$ and state 2 makes true $\mathit{3KG}(a)$. A fine grained aggregation allows to introduce a formula that expresses the *mean* of the weights, i.e., $\mathit{2KG}(a)$, whereas a coarse aggregation cannot. A model of a fine-grained aggregation in the context of measurement is left for future work. We refer to [24] for an approach to fine-grained aggregation that can be applied to the logic of measurement.

For the sake of example, we introduce a few aggregation functions. The first example is the unanimous aggregator that associates a certain truth value only if every state (MB) in $\mathfrak{S}$ agrees on that truth value.

$$un(x_1, \ldots, x_N) = \begin{cases} x_i, & \text{if for all } i, j \text{ we have } x_i = x_j; \\ u, & \text{otherwise.} \end{cases}$$

For the simple majority rule we assume that *maj* returns true (false) if the majority of states accept the truth (falsity), and it returns u in any other case.[8]

$$maj(x_1, \ldots, x_N) = \begin{cases} t, & \text{if } |\{x_i \mid x_i = t\}| > N/2; \\ f, & \text{if } |\{x_i \mid x_i = f\}| > N/2; \\ u, & \text{otherwise.} \end{cases}$$

[7] A treatment for a larger class of aggregators in social choice is presented in [23]. The motivation for the present treatment is that it easily allows for viewing aggregators as modalities. An overview of functions used to aggregated data is discussed in [4].

[8] The majority rule is generalized by *quota rules* that specify a threshold for acceptance of a certain truth-value. In this case, to define F as a function, we have to separately define quota rules for true, false, and undetermined.

The majority rule can be adapted to select only informative votes, that is, MBs that return true or false. We label this aggregator *determined majority.*

$$dmaj(x_1, \dots, x_N) = \begin{cases} t, \text{ if } |\{x_i \mid x_i = t\}| > (N - |\{x_i \mid x_i = u\}|)/2; \\ t, \text{ if } |\{x_i \mid x_i = f\}| > (N - |\{x_i \mid x_i = u\}|)/2; \\ u, \text{ otherwise.} \end{cases}$$

The previous aggregators are *anonymous*, namely any permutation of the MBs provides the same value, i.e., the reliability of the MBs is not considered. However, aggregators that use information about the reliability of MBs, when available, can be designed. Suppose to have a reliability partial order $\preceq$ defined on the states $\mathfrak{S}$. It is possible, for instance, to define a family of aggregators that associate truth-value x if the most reliable n sources wrt. $\preceq$ agree on x and u otherwise. Moreover, to handle disagreement among the most reliable source, one can use an auxiliary aggregation procedure, e.g. the majority rule. A detailed analysis of the properties of these aggregators is left for future work. However, we want to highlight that the reliability structure of the states, of the contextual information provided by the MBs, allows to define more refined aggregations. Additional meta-information could clearly be taken into account. For instance, one could consider the W3C PROV-ontology[9] to explicitly represent some characteristics of the MBs and of the measurement processes.

The language $\mathcal{L}$ can then be extended by adding a number of modal operators $\Box_F$ that depend on the aggregator F:

$$\mathcal{L}_{\Box_F} ::= \phi \in \mathcal{L} \mid \Box_F \phi$$

where the possible nesting of modalities is excluded, cf. [18].

A *modal structure* is a couple $\langle \mathfrak{S}, F \rangle$, where $\mathfrak{S}$ is a set (with cardinality N) of states all about the same set of objects O and F is an aggregation function. A *model* M for our modal logic is then obtained by adding for each state $s \in \mathfrak{S}$, the interpretation ε_s for the individual constants and the interpretation ι_s for the predicates.

The semantics of the non-modal formulas of $\mathcal{L}$ is the one provided in Sect. 3.1, now relative to a state $s \in \mathfrak{S}$, i.e., $||\phi||_{M,s} = ||\phi||_{\langle s, \varepsilon_s, \iota_s \rangle}$.

The semantics of modal formulas relies on the function F:

$$||\Box_F \phi||_{M,s} = F(||\phi||_{\langle s_1, \varepsilon_{s_1}, \iota_{s_1} \rangle}, \dots, ||\phi||_{\langle s_N, \varepsilon_{s_N}, \iota_{s_N} \rangle}).$$

Note that the truth-value of any modal formula is the same in all the states in $\mathfrak{S}$. We can construe the modal formulas as assessed wrt. the whole set of states $\mathfrak{S}$, rather than wrt. a single state. Moreover, by our definition of aggregators, every F is *systematic* [18], i.e., if $\models \phi \leftrightarrow \psi$, then $\models \Box_F \phi \leftrightarrow \Box_F \psi$. The modalities $\Box_F$ are then well-defined and they validate the *rule of equivalents* (RE) of minimal modal logic [5,18]. An axiomatisation of the minimal modal extension of three-valued logic can then be given by adding (RE) to the propositional axioms.

$$(\textbf{RE}) \quad \vdash \phi \leftrightarrow \psi, \text{ then } \vdash \Box_F \phi \leftrightarrow \Box_F \psi$$

[9] See https://www.w3.org/TR/prov-overview.

To characterise the aggregators, even in the bivalent case, a number of additional axioms are required, see for instance [18] for the case of the majority aggregator. We leave this aspect to a future work.

The condition (RE) does not constrain the way ε_s and ι_s can vary across the different states $s \in \mathfrak{S}$. In empirical terms, it is plausible to assume that the interpretation of the individual constants is fixed for every $s \in \mathfrak{S}$, i.e., for every $s, s' \in \mathfrak{S}$ and $c \in \mathbf{C}$ we have $\varepsilon_s(c) = \varepsilon_{s'}(c)$. The Barcan formula and its converse (BC), cf. [1], allows to axiomatise this property (i.e., a fixed domain assumption).

$$(\mathbf{BC}) \;\; \Box_F \forall x A(x) \leftrightarrow \forall x \Box_F A(x)$$

We also assume that the interpretation of predicates is stable across states, i.e., for every $s, s' \in \mathfrak{S}$ and $P \in \mathbf{R}$ we have $\iota_s(P) = \iota_{s'}(P)$. On the one hand this is empirically plausible: by means of measurement standards and calibration, measurement theories aim at guaranteeing the sharing of data collected by different MBs, i.e., they individuate a set of reference systems. A predicate needs then to have a stable intension, to always refer to the same symbol. On the other hand, this is in line with standard modal logic where the intension of a predicate P is represented by a unique function that provides, for each world, the extension of P in such world. In our framework, the extension of a predicate P in a state s may be defined as the set of objects $o \in O$ such that $\langle o, \iota(P)\rangle \in \delta(s)$, which of course can vary in different states.

3.3 Reasoning About Aggregated Data: Possible Inconsistency

We informally discuss a few issues in reasoning about aggregated data beyond the minimal principle assumed by (RE). Consider the following example.

Example 1. Suppose $\mathfrak{S} = \{s_1, s_2, s_3\}$, where the datasets provide information about weights, lengths, and colours. Suppose we assess the atomic propositions $\mathit{1KG}(a)$, $\mathit{1MT}(a)$, $\mathit{RED}(a)$, and $\neg \mathit{RED}(a)$, which are grounded on their respective datasets. The profile of truth-values for each state is reported in Table 2. Consider now the formula $\lambda = \forall x((\mathit{1KG}(x) \wedge \mathit{1MT}(x)) \leftrightarrow \mathit{RED}(x))$ representing a law that relates weights and lengths with colours. According to the semantics of the connectives previously introduced, each state in $\mathfrak{S}$ validates such law.

In this scenario, the aggregation by majority of the data exhibits a case of *discursive dilemma*, [14]. In empirical terms, the law is consistent with all the single sources, but not with the aggregated data, preventing, in this case, an inductive generalisation. This means that aggregators may in principle provide inconsistent information even if every input is consistent.

In order to infer the inconsistent outcome in our modal setting, three principles of reasoning are required.[10]

$$\begin{array}{rl}(\mathbf{RM}) & \text{if } \vdash \phi \rightarrow \psi, \text{then} \vdash \Box_F \phi \rightarrow \Box_F \psi \\ (\mathbf{C}) & \Box_F \phi \wedge \Box_F \psi \rightarrow \Box_F(\phi \wedge \psi) \\ & (\bot) \neg \Box_F \bot\end{array}$$

[10] Note that an analogous argument applies to the determined majority rule.

Table 2. Truth-values profile of the example 1

	$1KG(a)$	$1MT(a)$	$RED(a)$	$\neg RED(a)$	$\forall x((1KG(x) \wedge 1MT(x)) \leftrightarrow RED(x))$
s_1	t	t	t	f	t
s_2	t	f	f	t	t
s_3	f	t	f	t	t
maj	t	t	f	t	t

(RM) is the monotonicity principle, the principle (C) allows for combining aggregated information, and ($\bot$) excludes possibly inconsistent aggregated data By assuming (RM) (C) and ($\bot$), together with the axioms for the propositional logic, the calculus becomes inconsistent for the majoritarian aggregation.

In Example 1, $\Box_{maj} 1KG(a)$ and $\Box_{maj} 1MT(a)$ are true, therefore by (C) we infer $\Box_{maj}(1KG(a) \wedge 1MT(a))$. Since the law λ is true in every state, $\Box_{maj}\lambda$ is true. From $\Box_{maj}(1KG(a) \wedge 1MT(a))$ and $\Box_{maj}\lambda$, by (C) and (RM), we infer $\Box_{maj} RED(a)$. However, we also have $\Box_{maj}\neg RED(a)$, since a majority of states makes $RED(a)$ false. By (C) we obtain $\Box_{maj}(RED(a) \wedge \neg RED(a))$ and by (RE), since every contradiction is logically equivalent, we obtain $\Box_{maj}\bot$, against ($\bot$).

The principle (RM) legitimates the use of a logical inference at the level of aggregated data. E.g., it justifies to infer $\Box_{maj} RED(a)$ from $\Box_{maj} 1KG(a)$ and $\Box_{maj} 1MT(a)$ via $\Box_{maj}\lambda$. Notice that (RM) applies regardless of the majority that supports those data, the actual set of states that produces them.

While the principle (C) appears a reasonable principle for combining aggregated data, in fact it is also insensitive to the fact that possibly distinct, although overlapping, sets of MBs can be the source of the data. In the example, s_1 and s_2 agree on $1KG(a)$, whereas s_1 and s_3 agree on $1MT(a)$.

(RM) and (C) seem to identify two types of reasoning: an *intra-state* reasoning, where each state reasons about the data by means of the law, and an *inter-state* reasoning, where reasoning is performed at the level of aggregated data, by means of the law. It is in fact possible to separate the two forms of reasoning; for instance, by distinguishing two types of combinations of data (i.e., conjunction), one that applies to the case where the same states support a number of data, the other that combines data produced by distinct sets of states. This move is capable of restoring consistency, although it requires to enter the realm of substructural logics for modelling reasoning about aggregated data [20–22]. In fact, the possible inconsistency of the aggregated sets depends only on the meaning of logical connectives, not on the atomic formula produced by the MBs. If our language only contains atomic proposition, e.g., we prevent talking about $1KG(a) \wedge 1MT(a)$ and we content with $1KG(a)$ and $1MT(a)$ or we exclude laws to connect data, the majority is indeed consistent. Hence, it is worthy to investigate logical operators that preserve consistency under the majority rule and suitably represent the rules of reasoning about aggregated data.

By contrast, a simple solution, is to accept that there might be cases of inconsistent data aggregation and give up the axiom ($\perp$). Note that the non-anonymous procedures that we defined do indeed preserve consistency, however they rest on the demanding assumption of knowing in advance the most reliable MBs. More sophisticated aggregators require dropping the systematicity assumption that we embraced here and they are therefore left for future work.

3.4 The Contextual Nature of Measurement

We discuss now the contextual nature of measurement statements in the model that we proposed. We have introduced two types of statements expressing measurement: non-modal statements, that are assessed with respect to a single state, and modal statements that are assessed with respect to a plurality of states, by aggregating the information therein.

It useful here to apply the distinction between two interpretations of context proposed in [19]. A context can be intended in an *objective* (or *ontological*) way, i.e., basically as a metaphysical state of affairs, or in a *subjective* (or *epistemic*) way, i.e., basically as a cognitive representation of the world. According to this dichotomy, the non-modal statements of our logic are close to the epistemic view of context. The holding of a formula (e.g., $1KG(a)$) at a given state depends on the considered measurement battery, on its representation systems (the symbolic classification systems), on the resolution of the devices in the battery, and on the actual measurement processes performed. As observed in the introduction, it is however important to note that the degree of subjectivity of measurements is lower than the one of personal evaluations, opinions, perceptions, etc.

By contrast, the modal statements of our logic aggregate the information coming from several epistemic contexts. Thus, on the one hand, the aggregated statements are not immediately objective, in the above sense, as they are always mediated by the measurement systems, they are not directly reducible to real states of affairs, to sets of features of the world. On the other hand they are not merely subjective, as they balance between the viewpoints of different epistemic contexts. The aggregated statements seem then to constitute a further type of context, which we may term an *inter-subjective* context, which results from the aggregation of a number of subjective (epistemic) contexts.

We suggest an analogy between the aggregation of different epistemic context with the operation of *de-contextualisation* used in [19] to dismiss the demanding idea of an ontological context, while preserving the possiblity of an objective context, as resulting from intersubjective agreement (viz. “Objectivity is always a result of our interaction, not a datum”, [19], p. 283.)

In this sense, a theory of the aggregation of heterogeneous (epistemic) contexts may serve as the formal backbone of a theory of de-contextualisation, viewed as a theory of multiagent interaction. We leave the development of this suggestion to a dedicated work.

4 Conclusion

This work has three main contributions. Firstly, we introduced an explicit definition of states in terms of the epistemic theory of measurement. States are not simple indexes for possible worlds, they are sets of measurements collected by MBs. The datasets associated to each MB depend on the nature of the MB, therefore the information that we may assess at a state has indeed a contextual nature explained and justified in terms of the epistemic measurement theory.

The second contribution concerns the characterisation of the meaning, or more precisely the intension, of the properties represented by the predicates in $\mathcal{L}$. The theory of measurement allows us to interpret (unary) predicates into symbols that, by means of measurement standards, are conventionally assigned to perfect realisations of properties. While standard modal logic encapsulates the intension of a predicate into a function from worlds to sets of objects, our approach is more descriptive and operative, it associates a computational 'recipe' to a predicate: to calculate the extension of a predicate in a given state, one needs to look for the measurements (in such state) that have as output the symbol associated to the predicate.

Thirdly, we introduced modal operators to model aggregators of (possibly conflicting) data and we discussed the contextual nature of measurement statements distinguishing the device-based measurement and the aggregated measurement.

References

1. Arló-Costa, H., Pacuit, E.: First-order classical modal logic. Stud. Logica **84**(2), 171–210 (2006)
2. Avron, A.: Natural 3-valued logics characterization and proof theory. J. Symbolic Logic **56**(01), 276–294 (1991)
3. van Benthem, J., Pacuit, E.: Dynamic logics of evidence-based beliefs. Stud. Logica **99**(1), 61–92 (2011)
4. Calvo, T., Kolesárová, A., Komorníková, M., Mesiar, R.: Aggregation operators: properties, classes and construction methods. In: Calvo, T., Mayor, G., Mesiar, R. (eds.) Aggregation Operators. Studies in Fuzziness and Soft Computing, vol. 97, pp. 3–104. Springer, Heidelberg (2002). https://doi.org/10.1007/978-3-7908-1787-4_1
5. Chellas, B.: Modal Logic: An Introduction. Cambridge University Press, Cambridge (1980)
6. Duddy, C., Piggins, A.: Many-valued judgment aggregation: characterizing the possibility/impossibility boundary. J. Econ. Theory **148**(2), 793–805 (2013)
7. Endriss, U., Grandi, U., Porello, D.: Complexity of judgment aggregation. J. Artif. Intell. Res. **45**, 481–514 (2012)
8. Finkelstein, L.: Widely, strongly and weakly defined measurement. Measurement **34**, 39–48 (2003)
9. Frigerio, A., Giordani, A., Mari, L.: Outline of a general model of measurement. Synthese (Published online: 28 February 2009)
10. Gärdenfors, P.: Conceptual Spaces: The Geometry of Thought. MIT Press, Cambridge (2000)

11. Giordani, A., Mari, L.: Property evaluation types. Measurement **45**, 437–452 (2012)
12. Hájek, P.: Metamathematics of Fuzzy Logic. Trends in Logic, vol. 4. Springer, Netherlands (1998)
13. Kleene, S.: Introduction to Metamathematics. Bibliotheca Mathematica, North-Holland (1952)
14. List, C., Puppe, C.: Judgment aggregation: a survey. In: Handbook of Rational and Social Choice. Oxford University Press, Oxford (2009)
15. Mari, L.: A quest for the definition of measurement. Measurement **46**(8), 2889–2895 (2013)
16. Masolo, C.: Founding properties on measurement. In: Galton, A., Mizoguchi, R. (eds.) Proceedings of the Sixth International Conference on Formal Ontology and Information Systems (FOIS 2010), pp. 89–102. IOS Press (2010)
17. Masolo, C., Benevides, A.B., Porello, D.: The interplay between models and observations. Appl. Ontology **13**(1), 41–71 (2018)
18. Pauly, M.: Axiomatizing collective judgment sets in a minimal logical language. Synthese **158**(2), 233–250 (2007)
19. Penco, C.: Objective and cognitive context. In: Modeling and Using Context, Second International and Interdisciplinary Conference, CONTEXT 1999, Trento, Italy, September 1999, Proceedings, pp. 270–283 (1999)
20. Porello, D.: A proof-theoretical view of collective rationality. In: IJCAI 2013, Proceedings of the 23rd International Joint Conference on Artificial Intelligence, Beijing, China, 3–9 August 2013 (2013)
21. Porello, D.: Judgement aggregation in non-classical logics. J. Appl. Non-Classical Logics **27**(1–2), 106–139 (2017)
22. Porello, D.: Logics for modelling collective attitudes. Fundamenta Informaticae **158**(1–3), 239–275 (2018)
23. Porello, D., Endriss, U.: Ontology merging as social choice: judgment aggregation under the open world assumption. J. Logic Comput. **24**(6), 1229–1249 (2014)
24. Porello, D., Troquard, N., Peñaloza, R., Confalonieri, R., Galliani, P., Kutz, O.: Two approaches to ontology aggregation based on axiom weakening. In: Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, IJCAI 2018, July 13–19, 2018, Stockholm, Sweden, pp. 1942–1948 (2018)
25. Suppes, P., Krantz, D.M., Luce, R.D., Tversky, A.: Foundations of Measurement. Additive and Polynomial Representations, vol. I. Academic Press, Cambridge (1971)
26. Van Ditmarsch, H., van Der Hoek, W., Kooi, B.: Dynamic Epistemic Logic. Synthese Library, vol. 337. Springer, Netherlands (2007). https://doi.org/10.1007/978-1-4020-5839-4

Belief Puzzles as Paradoxes of Identity

Ramón García Moya(✉)

BIAP/LOGOS, Department de Filosofia, Universitat de Barcelona, Barcelona, Spain
ramongarciamoya@gmail.com

Abstract. This paper approaches traditional puzzles about belief and belief attributions as if they involved instances of paradoxes of identity. I shall argue that the solution to these puzzles comes with a proper understanding of the way we identify individuals in situations where their *persistence conditions* allow for a "split" (through time or possible worlds) and the way context constraints how we talk about them. My aim in this work is to outline the basics of such a solution and show how well-motivated it is compared to more conventional alternatives.

Keywords: Frege's puzzle · Belief reports · Paradoxes of identity · Derive context

1 Introduction

The central puzzle I shall be concerned with in this paper is *Frege's puzzle*, also known as *the problem of substitution between co-referential names in belief contexts*. The problem is, in a nutshell, to explain how pairs of reports like (1) and (2) differ in truth-value even though standard assumptions about the semantics for belief and proper names disallow any difference in the propositions they semantically express:

(1) The Greeks believed that Hesperus shines in the evening
(2) The Greeks believed that Phosphorus shines in the evening

Apparently, 'Hesperus' and 'Phosphorus' were the names that the ancient Greeks gave to (the personifications of) the brightest heavenly body visible in the evening and the brightest heavenly body visible in the morning, respectively. They initially thought that they were two distinct objects. It was only after they accepted the Babylonian theory that they came to believe that these objects were one and the same and gave it the name 'Aphrodite'. Standard intuitions say that an utterance of (1) is true whereas an utterance of (2) is false, or so when produced to answer the question <<What did the ancient Greeks believe before accepting the Babylonian theory?>>.

Now, a subject's beliefs are true (or false) and thus have representational content. On the simplest hypothesis, the content of a belief is individuated by the set of circumstances supporting its truth. On the assumption that truth-supporting circumstances are possible worlds, belief contents become *propositions* as (the characteristic function of) possible world sets. The content of a subject's belief system will therefore be given by the class

G. Bella and P. Bouquet (Eds.): CONTEXT 2019, LNAI 11939, pp. 189–201, 2019.
https://doi.org/10.1007/978-3-030-34974-5_16

of possible worlds supporting the truth of each of the subject's beliefs. This class is the subject's "doxastic alternatives", and is identified with the set of worlds W such that the subject believes nothing that rules out the hypothesis that w is the actual world, for all $w \in W$ [1, p. 27].

Given the identification of belief contents with propositions, it becomes natural to think of (1) and (2) as characterizations of the Greek's belief system by means of the propositions expressed by their respective prejacents. Thus, if W is the set of worlds modeling the content of the Greek's actual beliefs, an utterance of (1) is true if, and only if, each world in W supports the truth of the proposition which 'Hesperus shines in the evening' expresses. Likewise, an utterance of (2) is false if, and only if, at least some world in W does not support the truth of the proposition which 'Phosphorus shines in the evening' expresses.

However, the occurrences of the names 'Hesperus' and 'Phosphorus' that appear in (1) and (2) co-refer to Venus, and it is customary to think of *referential* uses of names as contributing their referent, if any, to the proposition expressed by sentences in which they occur[1]. As a result, the propositions expressed by the prejacents of (1) and (2) are the same, and so it remains a mystery how the reports can differ in truth-value when they are characterizations of the Greeks' beliefs by means of one and the same proposition. In other words, we expect to preserve the truth-value of (1) when substituting 'Phosphorus' for 'Hesperus', given that they make the same contribution to "what the reports say" that the Greeks believed. This is the substitution failure between co-referentials that Frege's puzzle exhibits and which we seek to account for.

Standard solutions to the substitution failure come with the rejection of some aspect or other of the overall picture about belief and/or names sketched. Thus, solutions within the Fregean camp construe 'believe' as an operator on contents individuated more finely than propositions (e.g. senses, characters, open propositions, etc.) so that we can distinguish between the beliefs ascribed by (1) and (2), respectively [3–5]. On their part, solutions within the Russellian camp claim that our truth-differing intuitions track contents that are only pragmatically associated with the reports: contents whose complete specification mentions the particular "notion", or "guise", that the speaker tacitly refers to and via which the subject is said to "agree" with the proposition *that Venus shines in the evening* [6–8]. But there are issues with the motivation and psychological plausibility of these solutions, issues which suggest that departures from the original picture should be avoided if possible. What I propose is to frame an account of our truth-differing intuitions regarding pairs of reports like (1) and (2) within a general approach to paradoxes of identity while leaving the original picture about belief and names intact.

1.1 Paradoxes from the Split of Individuals Through Time

The kind of account I propose can be motivated as follows. Our standard semantic theory carries a minimal metaphysics about the entities that make our statements true. This semantic theory says, again, that the contribution that a proper name makes to the

[1] This "referentialist intuition" is stated by the claim that proper names are *directly referential* expressions, in Kaplan's sense: i.e. expressions *"[...] whose referent, once determined [...] is taken as being the propositional component"* [2, p. 493].

proposition expressed is its bearer, or so in their apparently referential, literal uses. It also says that temporal adverbs (e.g. 'tomorrow') and attitude verbs (e.g. 'believe') are operators on propositions, and so that 'Aristotle will come to dinner tomorrow' is true if, and only if, Aristotle comes to dinner the day after the utterance, and that 'John believes that Aristotle was a philosopher' is true if, and only if, Aristotle was a philosopher in each of John's doxastic alternatives. Thus, our semantic theory says that the truth of these sentences depends on how things *will* be, or *might* be with Aristotle. So if these sentences have any chance to come out true, Aristotle must be the kind of entity that can feature in the domain of future times and doxastic alternatives, i.e. the kind of entity that can *persist* through time and doxastic alternatives.

Now, persistence conditions *through time* give rise to familiar paradoxes about individual identity. The Paradox of the Ship of Theseus is an example of how these conditions allow for cases where we seem to have *two* ships that are nevertheless equal candidates to be *the* ship of Theseus. The story goes as follows. Imagine that the ship of Theseus, called 'Victory', is on display in a museum at time t and that, as the years go by, its worn planks are gradually replaced by new ones. Let t' be the time in which the ship has been fully restored and call that restored ship 'Restoration'. Since ships persist through the gradual replacement of their parts, it seems clear that 'Restoration' is a new name for Victory, and so that Victory is identical to Restoration. Therefore, (3)

(3) At t', Victory is in the museum

is true in that story. Now imagine an alternative story where Victory has never had its planks replaced, but instead was taken apart and stored in a warehouse. After many years, its parts were put back together in just the way they were originally ensemble. Let t' be the time in which the ship has been fully reassembled in the warehouse and call the reassembled ship 'Reconstruction'. Since ships persist through their disassembling and reassembling, it seems clear that 'Reconstruction' is a new name for Victory, and so that Victory is identical to Reconstruction. Therefore, (3) is false in that story.

But now consider a "combined story" where Victory is subject to the two processes at once. Its planks have been replaced gradually in the museum and the replaced planks have been used to build an exact replica of the original ship in the warehouse. And suppose that at time t' the restoration and the reconstruction concluded. As a result, we have two ships at t': the one we called 'Restoration' in the first simple story, which is in the museum, and the one we called 'Reconstruction' in the second simple story, which is in the warehouse. So while (4) is true, (5) is false:

(4) At t', Restoration is in the museum
(5) At t', Reconstruction is in the museum

But this is paradoxical: how can Restoration and Reconstruction be two distinct ships when they are both the result of processes on Victory that it could (temporally) survive? If, by contrary, they are the same ship, why does substitution of 'Reconstruction' for 'Restoration' fail to preserve truth-value? And what is the truth-value of (3) in this combined story?

1.2 Paradoxes from the Split of Individuals Through Possible Worlds

Not only persistence conditions through time give rise to paradoxes of identity, though. So do persistence conditions *through possible worlds* taken as doxastic alternatives. The famous Hesperus/Phosphorus-case is, I think, an example of this. Imagine that the Greeks see Venus in the evening and come to believe that there is a heavenly body *x* which *only* shines in the evening. Let us call that body inhabiting their doxastic alternatives 'Hesperus', as they themselves did. Let us follow Stalnaker [9, p. 140, 164] in that we can identify an object o in a subject's doxastic alternative with an object o' in the actual world whenever o' plays a causal role in the generation of the subject's beliefs about o' and properties of o tend to co-vary with properties of o'. Since Venus played a causal role in the generation of the Greek's beliefs about *x* and properties of *x* tend to co-vary with properties of Venus (e.g. had Venus been bigger and square, *x* would have been bigger and square in each of the Greek's doxastic alternatives), we may identify *x* with Venus and so conclude that Venus "persists" in the Greek's doxastic alternatives. As a result, (6)

(6) The Greeks believed that Venus shines in the evening

is true in our first simple story. Now imagine instead that the Greeks see Venus in the morning and come to believe that there is heavenly body *y* which *only* shines in the morning. Let's call that body inhabiting their doxastic alternatives 'Phosphorus', as they themselves did. Since Venus played a causal role in the generation of the Greek's beliefs about *y* and properties of *y* tend to co-vary with properties of Venus, we may identify *y* with Venus and so conclude that Venus "persists" in the Greek's doxastic alternatives. As a result, (6) is false in this second simple story.

But now let's combine the two simple stories. The Greeks see Venus in the evening and come to believe that there is a heavenly body *x* which only shines in the evening. They also see it in the morning and come to believe that there is a heavenly body *y* which only shines in the morning. As a result, we have two heavenly bodies in the Greek's doxastic alternatives: the one we called 'Hesperus' in the first simple story, which shines in the evening, and the one we called 'Phosphorus' in the second simple story, which shines in the morning. So while (1) is true, (2) is false. But this is paradoxical: how can Hesperus and Phosphorus be two distinct heavenly bodies when they bear the same sort of causal dependency with one and the same entity? If, by contrary, they are identical, why does substitution of 'Phosphorus' for 'Hesperus' fail to preserve truth-value? And what is the truth-value of (6) in the combined story?

It is, I think, issues concerning the (re-)identification of individuals in situations where their persistence conditions allow for a "split" that give rise to substitution failures between co-referential terms within attitude reports and our mixed intuitions about problematic reports like (6). My hypothesis is that once we get clear on how an individual can become two in a given doxastic alternative we will be able to account for how two co-referential names can become names for two distinct individuals and so fail to be interchangeable.

2 Substitution Failures in Quantifying-in Reports

The present approach to Frege's puzzle is not based simply on commonalities between substitution failures in attitude reports and substitution failures in paradoxes of identity. The traditional takes on the puzzle have problems on their own that suggest the kind of approach I am advocating. For instance, Fregeans accommodate our truth-differing intuitions regarding (1) and (2) by claiming that belief contexts are "opaque", meaning that names occurring in these contexts are used to more than simply denote their referents. A "mode of presentation" of Venus is being referred to as well. This semantic thesis is accompanied by another claim within the philosophy of mind: namely, that a belief about Venus is not fully specified without the mentioning of the way in which Venus is presented to or thought about by the subject.

Despite the role that modes of presentation are assigned by a Fregean semantics for belief, it was always kept in mind that we seem to report beliefs by quantifying into variables occupying a purely objectual position: i.e. we seem to report beliefs "*de re*". As Quine [10, p. 178] pointed out, (7)

(7) Ralph believes that someone is a spy

has two readings, one apparently equivalent to (7_{narrow}), which expresses Ralph's humdrum belief that there are spies, and another apparently equivalent to (7_{wide}), the one which belief-relates Ralph with a particular (although unspecified) individual:

(7_{narrow}) Ralph believes that $\exists$x. x is a spy
(7_{wide}) $\exists$x. Ralph believes that x is a spy

But the kind of quantification that (7_{wide}) exhibits makes the mode in which the unspecified individual is presented irrelevant to the question whether Ralph has the ascribed belief. Thus, if the Fregean account of the substitution failure exhibited in (1) and (2) in terms of a difference in the modes of presentation contributed by 'Hesperus' and 'Phosphorus' to the contents of the beliefs ascribed is correct, we should not find substitution failures in reports with quantification in, since modes of presentation do not feature in the contents of the beliefs they ascribe.

However, there are truth-differing intuitions regarding reports that, like (8) and (9)

(8) There is a heavenly $body_1$ which the Greeks called 'Hesperus'.
The Greeks believed that it_1 shines in the evening.
(9) There is a heavenly $body_2$ which the Greeks called 'Phosphorus'.
The Greeks believed that it_2 shines in the evening.

exhibit quantification in. This might come as a surprise, as it is commonly held that substitution between names occupying "transparent contexts" is unproblematic [10, p. 181], [11, p. 242], [12, p. 329]. But the intuitions, I believe, are there, and as Pickel [13, p. 348] claims, *"[i]f one takes Frege's puzzle cases seriously for proper names, then one should also take them seriously in cases of quantifying in"*.

2.1 Fallacies of Intelligibility

Of course, Fregeans may react to this by claiming that on their analysis of (8) and (9) modes of presentation *do* feature in the contents of the beliefs ascribed. After all, Fregeans charged purely objectual quantification into belief contexts of *unintelligibility*. The unintelligibility claim was given an intuitive flavor by stories like Ralph's. According to the story, *"[t]here is a certain man in a brown hat whom Ralph has glimpsed several times under questionable circumstances [...] Ralph suspects he is a spy. Also there is a gray-haired man, vaguely known to Ralph as rather a pillar of the community, whom Ralph is not aware of having seen once at the beach. Now Ralph does not know it, but the men are one and the same. Can we say of this* man *(Bernard J. Ortcutt, to give him a name) that Ralph believes him to be a spy?"* [10, p. 179] The story was meant to stress a certain conceptual defectiveness in saying that Ortcutt satisfies (or fails to satisfy) the open formula λx. Ralph believes that x is a spy *independently of any way of specifying him*, and so Fregeans proposed to analyze quantifying-in reports as indirect characterizations of "notional", or "sentential" beliefs meeting an external condition [14]. On an sententialist version of the analysis, (7) would become equivalent to (7_{not*}),

(7_{not*}) ∃x. ∃n. n refers to x & Ralph believes-true the sentence 'n is a spy'

and come out true of Ortcutt and the expression 'that man on a brown hat' (as used to demonstrate Ortcutt). If now we allow that the contexts in which (8) and (9) are uttered restrict the domain quantification over modes of presentation to the class of modes of presentation of men on a brown hat, Fregeans can account for the difference in the truth-value of (8) and (9) while preserving a central role to modes of presentation in the characterization of the beliefs ascribed.

Stories like Ralph's succeed in bringing to light that something is faulty with *de re* reports like 'Ralph believes *of* Ortcutt that he is a spy' traditionally construed. But the Fregean analysis of reports with quantification in is motivated only if that faultiness is due to the fact that modes of presenting Ortcutt are missing in the specification of the belief ascribed. And there is reason to think that this is not the source of what is faulty with the report. Shaw [15] convincingly argues that parallel arguments for the unintelligibility of objectual quantification into temporal (and modal) constructions are fallacious. Just as it happened with (7), (10) exhibits what we would initially treat as scope ambiguity (where **F** is an operator shifting the utterance time to any future time),

(10) The oldest senator will greet you on the Capitol steps
(10_n) **F [**∃x. x is the oldest senator & x greets you on the Capitol steps]
(10_w) ∃x. x = the oldest senator & **F** [x greets you on the Capitol step]

and so there is quantification into temporal contexts. And just as it happened in belief contexts, substituting 'the winner of the 2006 NJ senatorial election' for 'the oldest senator from NJ' for in (11)

(11) **F** [The oldest senator from NJ will greet you in the Capitol steps]

can affect its truth-value: although both descriptions are satisfied by Menendez at the present time, it might be that at no future time the oldest senator is the winner of the 2006 senatorial election and that only the winner will great you. So the fact in (10_w) there is no mode of presenting Menendez denoted in the statement of how things will be with him should make (10_w) somewhat faulty. But, as Shaw observes, no-one would take issue with the intelligibility of (10_w) on these grounds. We would deem it unqualifiedly true whenever we can uniquely identify Menendez in some future time and see that he greats you in the Capitol steps then.

Interestingly, Shaw [15, p. 54] observes that *"[w]hat helps us make sense of the* de re *reading of the description in* [(10)]*, despite requiring quantification across a temporal operator, is that we seem to have [...] a good sense of what it is for some future person on the Capitol steps to be identical with Menendez, the man we presently identify as the oldest senator from New Jersey."* This observation, I think, helps us see what underlies the faultiness in saying of Ortcutt that Ralph's believes him to be a spy. Arguably, Ralph has an inconsistent conception of Ortcutt: he believes that he spies and that he is not a spy. So it is reasonable to model these beliefs by means of doxastic alternatives with *two* existential instantiations of Ortcutt: one that is a spy and another that isn't. But since there are two Ortcutts in of Raph's doxastic alternatives, we lack a good sense of what it is for something to be (uniquely) Ortcutt in these doxastic alternatives. Arguably, this affects the intelligibility of our reporting Ralph's beliefs about Ortcutt. So although Fregeans are right that stories like Ralph's put pressure on a conception of belief that somewhat allows for a single individual to satisfy both λx. Ralph believes that x is a spy and λx. Ralph believes that x is not a spy, the pressure vanishes once Ralph is regarded as a partially inconsistent believer, and the unintelligibility of problematic *de re* reports can be seen as simply the result of the ambiguity that some descriptions of inconsistent beliefs fall prey to.

To sum up, substitution failures in reports with quantification in cast doubts on the Fregean claim that these failures are produced by a truth-conditional import of modes of presentation. In turn, the Fregean claim that quantifying-in reports are equivalent to quantificational statements over contextually restricted classes of modes of presentation is based on a suspect diagnosis of the faultiness of some *de re* reports standardly construed. That faultiness, I have argued, has more to do with our incapacity to identify the relevant individuals across the subject's doxastic alternatives, a capacity which is hindered when the subject has an inconsistent conception of them. All this, I think, raises a question whether we really need to introduce modes of presentation into the contents of (1) and (2) so as to account for our truth-differing intuitions, or it is other features of the reports that make us distinguish between the beliefs ascribed.

3 Referring to Particular Notions

The other traditional approach to Frege's puzzle is the Russellian approach. Their advocates hold that belief should be analyzed in terms of its (Russellian, structured) proposition and its *vehicles of representation*, whether the "notions" in virtue of which the belief has the content that it has [8], the "guises" via which subjects grasp the relevant propositions [6], or sentence-like entities in one's "belief box" [16]. On this characterization of beliefs *qua* concrete cognitive structures, it is tempting to think that the Greeks had a belief with the proposition *that Venus shines in the evening* and a 'Hesperus'-notion yet lack a belief with that very content and a 'Phosphorus'-notion. Thus, *if* belief reports characterized a subject's beliefs by specifying both their contents and the notions they involved we could explain our truth-differing intuitions regarding (1) and (2).

This is, in fact, what most Russellians claim happens at a pragmatic level. Thus, advocates of the "hidden-indexical" version of the view state that *"[i]n reporting beliefs, we* *quite often are talking about such notions, although our belief reports do not explicitly mention them. The general solution to the puzzles is to allow a condition on particular beliefs, over and above a content condition, to be part of the claim made. [...] a specification of the notions that are supposed to be involved."* [9, p. 697] According to this version of the view, the notions the report is about need not be denoted by any expression in the report (as it happens in notionally explicit reports of the form 'A believes that p via the notion N'), but may be provided by the speaker's "tacitly referring" to them. Furthermore, the view takes it that the use of a particular expression usually gives the audience clues about which notion the speaker is providing. This way, the fact that the speaker uses 'Hesperus' when uttering (1) and 'Phosphorus' when uttering (2) suggests that she is referring to different notions each time.

3.1 Psychological Plausibility

So the Russellian hypothesis about what causes the impression that substitution between co-referential names fails is that when we deem (1) true and (2) false we are not judging the notionless proposition they semantically express (to wit, the proposition *that there is some (type of) notion/guise X such that A believes that p via X*), but rather the notion-involving enrichments explicit in reports like (12) and (13):

(12) The Greeks believed that Venus shines in the evening via a 'H_{esperus}'-notion.
(13) The Greek believed that Venus shines in the evening via a '$P_{\text{hosphorus}}$'-notion

However, several authors, otherwise Russellian-friendly, have taken issue with the idea that notions are inserted into the contents of our reports by tacit reference. Thus, Braun [17, p. 560] finds it questionable *"[t]that ordinary speakers have such sophisticated thoughts and intentions about mental representations when they utter belief sentences"*. In a similar vein, Schiffer [18, p. 512] claims that *"[o]ne may reasonably doubt that belief ascribers mean what the hidden indexical theory requires them to mean when they ascribe beliefs"*. I think there is indeed a worry about the psychological plausibility of the Russellian hypothesis. Typically, ascribers are not concerned with notions when reporting beliefs. They do not have to: a speaker may truly say that so-and-so believes that Fido is a dog *"[e]ven though she is not in a position to refer to any particular mode of presentation so-and-so has for either Fido or doghood"* [19, p. 100, fn. 11].[2] But even if ascribers did refer to notions occasionally, there are substitution failures in reports with quantification in, i.e. reports with no specification of notions or ways of grasping propositions. And this, I think, undermines much of the motivation for the claim that our truth-differing intuitions regarding (1) and (2) track notion-involving enrichments of their otherwise *conceptually* complete contents.

4 Back to the Hypothesis: The Paradox of Theseus Revisited

The discussion so far suggests that an account of our truth-differing intuitions regarding (1) and (2) should not assume that the speaker makes reference to anything other than particular heavenly bodies. Frege's puzzle is a puzzle involving reports that describe an agent's beliefs by means of a *singular proposition* about Venus. But now the question is

[2] See also [12, p. 328–329], [20, p. 230] or [21, p. 33] for related claims.

how (1) and (2) can differ in truth-value given that 'Hesperus' and 'Phosphorus' co-refer? My proposal is to address this question in the same way we would address the question about how (4) and (5) differ in the truth-value when evaluated at Victory's combined story. So let me briefly sketch my preferred answer to this latter question and then show how a suitably related answer helps with Frege's puzzle.

4.1 Illusions of Identity

We know that in the first simple story about the ship of Theseus, s_1, a ship has been built and named 'Victory', that it has been fully restored gradually, left in a museum, and that ships persist through the gradual replacement of their planks. If someone comes and names the restored ship 'Restoration', she will have introduced a new name for Victory. Similarly, we know that in the second simple story, s_2, Victory has been built and named 'Victory', that it has been disassembled and reassembled in a warehouse while keeping their planks as they were originally arranged, and that ships persist through their disassembling and reassembling. If someone comes and names the reassembled ship 'Reconstruction', she will have introduced a new name for Victory. In the combined story, s_3, though, Victory has been built and named 'Victory'. However, it is fully restored gradually and left in the museum, and the planks taken off have been reassembled in a warehouse so as to make a replica of the original ship. The restored ship is named 'Restoration' and the reconstructed ship 'Reconstruction'.

Now, we know that the restored ship in s_3 is not identical to the reconstructed ship in s_3, if only because the former is in a museum and the latter in a warehouse. Therefore, *one of the two ships is not to be identified with Victory*. There are two options at this point: either (i) ships do not persist through gradual replacement of their planks if at the same time they are subject to their disassembling and reassembling, or (ii) ships do not persist through their disassembling and reassembling if at the same time they are subject to a gradual replacement of their planks. If (i) then (3) is true at s_3, but if (ii) then (3) is false at s_3. It is reasonable to think that our mixed intuitions about (3) when evaluated at s_3 come from doubts concerning which option is right.

But let us suppose that (ii) holds: i.e. that preservation of overall form takes priority over preservation of matter. On this assumption, the names 'Venus' and 'Restoration' that originate in s_3 are names for Victory, just as their respective orthographical counterparts in s_1 and s_2. However, the name 'Reconstruction' that originated in s_3 is a name for another ship, call it 'New', the one that has been generated in the process of reassembling the planks that used to be Victory's.[3] This way we can explain the difference in truth-value at s_3 between (4) and (5) in terms of a difference in the propositions that they express

[3] We can represent the different contents that each of these orthographically identical names have on each of these stories with the following two-dimensional concepts,

	'Victory'			'Restoration'			'Reconstruction'		
	s_1	s_2	s_3	s_1	s_2	s_3	s_1	s_2	s_3
s_1	Victory	Victory	Victory	Victory	Victory	Victory	----	----	---
s_2	Victory	Victory	Victory	----	----	----	Victory	Victory	Victory
s_3	Victory	Victory	Victory	Victory	Victory	Victory	New	New	New

where the stories featuring in the vertical axis play the role of utterance situations and the stories on the horizontal axis the role of truth-supporting circumstances.

when used in s_3, i.e. by simply interpreting (4) and (5) as containing occurrences of the very names *that were originated in s_3*. After all, when we considered s_3 we looked at it as if it were the actual utterance situation, and so it was very natural to evaluate (4) and (5) as if the names that occurred in them had their semantic properties determined by facts in s_3. Any failure of substitution between co-referential names as well as any appearance that Victory has two "existential instantiations" in s_3, as Pickup [22] has recently claimed, is an illusion. If (ii) is the right option, the illusion is induced by the following plausible yet ultimately false conditional: (if Victory has its planks disassembled and reassembled at s_3, then name 'Reconstruction' that originates in s_3 is a name of Victory) → (if Victory has its planks disassembled and reassembled at s_3 *and its planks replaced in s_3*, the name 'Reconstruction' that originates in s_3 is a name of Victory). Its falsehood is not surprising, though. In the end, the facts that obtain in s_3 are not simply the sum of those obtained in s_1 and s_2: some are new (e.g. New is generated), and some of the facts that obtained in s_2 are lost (e.g. Victory has been reassembled or disassembled). So we may expect that the referential facts holding in the simple stories do not "transfer" to a story that combines them.[4]

5 Extending the Approach to Frege's Puzzle

Let us now address Frege's puzzle as if it involved just another paradox of identity. We know that the evening body in the Greek's doxastic alternatives is not identical to the morning body, if only because the former shines only in the evening and the latter shines only the morning. Therefore, one of the heavenly bodies inhabiting these doxastic alternatives is not identical to Venus. There are two exclusive options at this point: either (i) the evening body is Venus, or (ii) the morning body is Venus. If (i) then (6) is true; if (ii) then (6) is false. It is reasonable to think that our mixed intuitions about (6) come from doubts concerning which option is right.

Unlike the case of Victory, though, we may assume that the world is itself indeterminate as to which of the two bodies is Venus. With Victory we had two distinct criteria for identification, that of preservation of overall form and that of preservation of matter, and it was reasonable to assume that the world settles the matter about which criterion should take priority over the other. And although there is metaphysical question about how the world settles the matter, any appearance of indeterminacy about which ship is Victory in s_3 is merely epistemic. But in the present case we only have one criterion of identity: that which says, roughly, that if an entity o in a possible world tends to carry information about o' by having properties of o', o can be identified with o'. And we have that both the evening and the morning bodies in the Greek's doxastic alternatives tend to carry information about Venus by having properties of Venus (and, let's assume, in the same degree). So here we are in a situation where two entities have equal right to be identified with Venus according to one and the same criteria, and so where the world itself does not seem the settle the matter.

[4] If instead it is (i) that holds, there would be a difference in how we should resolve the paradox at the metaphysical level, but not in the way we would account for the difference in truth-value between (4) and (5), which is my main purpose.

5.1 Representing Indeterminate Belief States

The moral of this is that when we represent the Greek's belief state in a possible world framework we need a representation of *indeterminate* belief states. Such a representation, I think, requires that the Greek's doxastic alternatives be modeled by two disjoint world sets, W_1 and W_2. For all worlds in W_1, Venus is the evening body, whereas for all worlds in W_2, it is the morning body. Intuitively, W_1 is the set of worlds that would be compatible with the Greeks' beliefs had the world settled that Venus is the evening body, and W_2 the set of worlds that would be compatible with the Greeks' beliefs had the world settled that Venus is the morning body.

Aside from that, we need to stipulate that for each world w in W_1 there is one (and only one) world in W_2 which is *qualitatively indiscernible* from w: i.e. that it cannot be told apart from w by the way it looks. This is so because the indeterminacy hypothesis as to the identity of Venus in these doxastic alternatives translates into an irresolvable matter as to which of the two world sets in fact represents the Greek's belief state, *not* into an irresolvable matter as to whether the Greeks believe, e.g. that the heavenly body that looks such-and-such, whichever that is, has such-and-such property. So let f be a function from worlds in W_1 (and W_2) to their indiscernible worlds in W_2 (and W_1). And let w_1 be a world in W_1. By assumption, Venus is the evening body in w_1. Now suppose that Venus looks red in w_1. It follows that the evening body in $f(w_1)$ (which is not Venus) will look red in $f(w_1)$ just like Venus does in w_1.

5.2 Patterns of Update on the Context

Our representation of the indeterminacy of the Greek's belief state imposes two patterns on the updating of *our* beliefs about what the Greek's believe. Let W_1* and W_2* be the two disjoint world sets that, for all we (ascribers) mutually presuppose, might be the Greek's actual doxastic alternatives.[5] If we are totally right in our presuppositions about what the Greeks believe, $W_1* = W_1$ and $W_2* = W_2$. This is, however, unlikely, and the goal of attributing beliefs in a conversation is, we may assume, to turn $W*_1$ into W_1 and $W*_2$ into W_2. The first of these patterns is that a belief attribution updates W_1* and W_2* by ruling pairs of world sets related by our indiscernibility function f. Thus, if a belief attribution like 'The Greeks believed that Venus doesn't look red' *directly* updates W_1* (to wit, by ruling out those worlds in W_1* where Venus looks red), it *indirectly* updates W_2* by ruling their respective qualitatively indiscernible worlds where the evening body inhabiting them looks red.

[5] The union set $W_1* \bigcup W_2*$ is the "derive context" set for the Greek's beliefs, as defined by Stalnaker [9, p. 157]. The derived context (for the Greeks) is determined by the basic context (i.e. the set of possible ways the world might be compatible with what all conversational partners mutually presuppose) in the following way: for each possible world in the basic context, the Greeks are in a belief state defined by the set of possible worlds compatible with what they believe in that world. The union of all these possible belief states is the set of all possible worlds that might, for all it is presupposed in the conversation, be compatible with the Greek's beliefs. On another note, this set will be typically updated when explicit belief attributions are made, and the update will proceed by the usual intersection operation: e.g. an utterance of 'A believes that S' will intersect the derive context for A's beliefs with the proposition expressed by S (as used in the utterance context).

The second of the patterns is this. A belief attribution either directly updates W_1* or directly updates W_2*, yet never both at once. The rationale for this pattern is the following. The names 'Venus', 'Hesperus' and 'Phosphorus' are all names for Venus, since their semantic properties are determined by facts obtaining in the actual world, *not* by those that obtain in the worlds as the Greeks took them to be.[6] Suppose that our utterance of (6) had the following effect on our derive context: it rules out from W_1* *and* W_2* those worlds where Venus does not shine in the evening. The utterance would produce two indirect updates: one that rules out those worlds in W_2* where their evening body does not shine in the evening and another that rules out those worlds in W_1* where their morning body does not shine in the evening. As a consequence, my utterance of (6) will count as an assertion about *both* the evening and the morning bodies in each of the doxastic alternatives. But this is intuitively incorrect, as the utterance was intended to describe how things are with only one of them in each of the doxastic alternatives. This result is blocked if our utterance of (6) is interpreted as directly updating only one of the world sets in the derive context.

Now, the second pattern of update raises an important question. Suppose that someone utters (6). Is she intending to directly update W_1* or W_2*? In other words, is she making a claim about the Greek's beliefs about the evening or the morning body? This is important, as the utterance will be true only if it directly updates W_1*, since only then it counts as a proposal to drop all worlds in W_1* and W_2* where their evening bodies do not shine in the evening *and both W_1 and W_2 are sets of worlds where these bodies do shine in the evening*. Our earlier concern about how to decide between options (i) and (ii) shows up again and stays on the way of an unambiguous interpretation of the report. This contrasts the case of (1) or (2), reports which manage to suggest which set is being intended to directly update, and so how we should resolve the identification of Venus for the purpose of the conversation. The speaker who uses 'Hesperus' as in (1) suggests that we identify Venus with the evening body in each of the doxastic alternatives, most likely by exploiting the fact that this is the body which the Greeks called 'Hesperus', and so that W_1 be taken as the set to be directly updated. If the suggestion is accepted, the utterance will drop from W_1* those worlds where Venus does not shine in the evening, and indirectly drop from W_2* those worlds where their evening body (not Venus) does not shine in the evening (nothing has been said about the morning bodies in W_1 or Venus in W_2). This accounts for our ordinary intuitions about the truth of (1). Likewise, the use of 'Phosphorus' in (2) suggests that we identify Venus with the individual that the Greeks called 'Phosphorus', and so that one takes W_2* to be the set to be directly updated. If the suggestion is accepted, the utterance will drop from W_2* those worlds where Venus does not shine in the evening, and indirectly drop from W_1* those where the morning body does not shine in the evening. This accounts for our ordinary intuitions about the falsity of (2).

The choice of the name 'Venus' when reporting as in (6) does not carry a suggestion about how to identify its referent as readily as 'Hesperus' or 'Phosphorus' do, however. So unless alternative conventions are in place the report won't proceed in accordance

[6] I am here departing from those views like Stalnaker's [9], Cumming's [3] or Pickel's [13], to mention a few, according to which 'believe' shifts the contextual parameters relevant for the interpretation of the names that occur within its scope.

with the second of the patterns and so will have an unambiguous interpretation. So this is just another case where contextual resolution is required to further specify what is said (in this case, about what others believe) even after superficial context-dependency has been fully resolved.

6 Conclusion

I have explained how utterances of (1) and (2) differ in truth-value by exploiting the different ways we have to identify Venus across doxastic alternatives, and so by appealing to the different worlds sets that can be identified with the (coarse-grained) proposition *that Venus shines in the evening*. This is a merit of the account, as it is congenial with a simple view about quantification into belief contexts. Moreover, it is economical, as it does not resort to claims about speakers referring to representational entities (e.g. modes of presentation, notions, sentences in the mind, etc.) when reporting beliefs in order to vindicate our truth-differing intuitions. Furthermore, it is fit to account for substitution failures exhibited in reports with quantification in.

References

1. Lewis, D.: On the Plurality of Worlds. Wiley-Blackwell, Hoboken (1986)
2. Kaplan, D.: Demonstratives. In: Almog, J., Perry, J., Wettstein, H. (eds.) Themes from Kaplan, pp. 481–563. Oxford University Press, Oxford (1989)
3. Frege, G.: Über Sinn und Bedeutung. Philos. Rev. **1**, 572 (1892)
4. Cumming, S.: Variabilism. Philos. Rev. **117**(4), 525–554 (2008)
5. Chalmers, D.: Propositions and attitude ascriptions: a Fregean account. Noûs **45**(4), 595–639 (2011)
6. Salmon, N.: Frege's Puzzle. MIT Press, Cambridge (1986)
7. Soames, S.: Direct reference, propositional attitudes, and semantic content. Philos. Top. **15**(1), 47–87 (1987)
8. Crimmins, M., Perry, J.: The prince and the phone booth: reporting belief reports. J. Philos. **86**(12), 685–711 (1989)
9. Stalnaker, R.: Context and Content, 1st edn. Oxford University Press, Oxford (1999)
10. Quine, W.: Quantifiers and propositional attitudes. J. Philos. **53**, 177–187 (1956)
11. Kripke, S.: A puzzle about belief. In: Margalit, A. (ed.) Meaning and Use, pp. 239–283. Reidel (1979)
12. Recanati, F.: Direct Reference. From Language to Thought. Blackwell, Hoboken (1993)
13. Pickel, B.: Variable and attitudes. Noûs **49**(2), 333–356 (2015)
14. Kaplan, D.: Quantifying in. Synthese **19**, 178–214 (1969)
15. Shaw, J.: De Re belief and cumming's puzzle. Anal. Philos. **56**(1), 45–74 (2015)
16. Richard, M.: How I say what you think. Midwest Stud. Philos. **14**, 317–337 (1989)
17. Braun, D.: Understanding belief reports. Philos. Rev. **107**(4), 555–595 (1998)
18. Schiffer, S.: Belief ascription. J. Philos. **89**(10), 499–521 (1992)
19. Schiffer, S.: Belief ascription and a paradox of meaning. Philos. Issues **3**, 89–121 (1993)
20. Bach, K.: Do belief reports report beliefs? Pac. Philos. Q. **78**, 215–241 (1997)
21. Hanks, P.: Structured propositions as types. Mind **120**(477), 11–52 (2011)
22. Pickup, M.: A situationalist solution to the ship of theseus puzzle. Erkenntnis **81**, 973–992 (2015)

Combining Probabilistic Contexts in Multi-Agent Systems

Livia Predoiu(✉)

Free University of Bozen-Bolzano, Bolzano, Italy
predoiu@unibz.it

Abstract. We propose an approach for modelling, integrating, and querying distributed probabilistic contexts in multi-agent systems. We assume each agent to be equipped with an independently acquired uncertain context. By taking advantage of established database technologies, we represent the uncertain context of each agent as a set of probabilistic facts conveniently stored in a probabilistic database. Members of a multi-agent system act autonomously and interact with each other. The interaction between agents consists of sharing access to their contexts with each other and allowing queries over the combined shared contexts. This amounts to the challenge of combining and querying distributed probabilistic databases. To combine probabilistic contexts, we define a context-matching operator that creates a joint probability distribution with given marginal probabilities. Furthermore, we propose a query answering method over combinations of probabilistic contexts.

Keywords: Joint Probability Distributions · Copulas · Contexts · Data Integration · Probabilistic Databases · Query Answering

1 Introduction

Multi-agent systems (MASs) consist of a set of intelligent agents where each agent is supposed to carry out a set of goals by engaging in both, acting autonomously and interacting with other agents. The purpose of a MAS is to solve complex problems by dividing them into smaller tasks, which then are assigned to agents of the MAS. Each agent is autonomous and equipped with a set of actions that it can choose from in order to create a strategy to successfully complete its assigned tasks. The agent decides which actions to use, in which order, using different kinds of inputs, such as a history of already performed actions and outcomes obtained, or background information collected from large data sets such as the Web. In addition, each agent is allowed to interact with other agents of the MAS and query their history of actions, outcomes or background information in order to help them to make an informed decision about their next actions.

The applicability of a MAS is wide, and there is a large body of literature on them; see e.g., [10,11] for recent surveys of MASs. Coordination between cooperating agents, structuring their interaction and integrating their shared data or

G. Bella and P. Bouquet (Eds.): CONTEXT 2019, LNAI 11939, pp. 202–215, 2019.
https://doi.org/10.1007/978-3-030-34974-5_17

knowledge while they interact remains to be a challenge. In this paper, we study a variant of this problem where the goal of the MAS is to learn probabilistic data from a big data or real-world environment such as the Web. More specifically, each agent collects (or learns, respectively) probabilistic information, which is stored in a probabilistic database, and which is shared, combined, and queried in an interaction between agents.

MASs that summarize and process large data sets in formats such as text, pdf, and images via e.g. the web, apps, and interfaces of content providers support humans in their information needs, e.g., for informed decision making. Agents that process and summarize large data sets given certain goals and user preferences in the context of information gathering tasks or user interests can exploit the fact that they can act at the same time in parallel and explore data either in different data sets and/or in the same data set but according to different preferences or information needs of the users or the information needed for the tasks and the goals to be achieved. We consider the collected, summarized, and processed data of each agent to be its *context*.

We assume a finite set $\mathcal{A}$ of agents to be given. Each of these agents, $a_i \in \mathcal{A}$, is equipped with its own context c_i, which we assume to be a probabilistic structured data set, which has been collected automatically by the respective agent according to its strategy. For example, given appropriate goals and strategies, it may have crawled the Web and/or studied the behavior of users and gathered structured data. We associate the uncertainty inherently present in the automatic collection of contextual (structured) data with a probabilistic model, and we assume it to be stored in a probabilistic database allowing us to draw on well-established database technology and its probabilistic extension, probabilistic databases (see [2] and [7] for surveys). This setting is inspired by applications crawling the Web and storing data in databases such as DIADEM [12] and DeepDive [14,15]. Observe that DIADEM is storing data in a deterministic database, but DeepDive uses a probabilistic database. Our setting is also inspired by the system for context-based concurrent experience sharing proposed in [9]. Recent related work on integrating contexts in MASs is, amongst others, [8], where ontological context is communicated between agents and integrated.

Our contribution in this paper is, first, a proposal for modelling probabilistic contexts in an information gathering MAS by means of probabilistic databases (see Sect. 3), which allows each agent to also query its own uncertain data, and second, to integrate different probabilistic contexts by creating joint probability distributions over the combined contexts as copulas, which are known in statistics as multivariate joint distributions with given marginals. In Sect. 4, we propose a context-based matching operator for creating those joint distributions and focus on some special shapes of the joint distributions that can be created. Finally, in Sect. 5, we propose a query answering method to query the combined contexts. To our knowledge, using probabilistic databases to store contexts and combining them with copula constructions has not yet been studied in the context of MASs. Our approach of distributed query answering is also novel in the area of probabilistic databases. In statistics, the copulas considered are usually continuous and are based on continuous marginal distributions (see also [6]), and no databases are used for storage and querying.

2 Preliminaries

This section briefly introduces some background on databases, queries and probability theory. We define schemas, databases, and queries in Sect. 2.1. Subsequently, in Sect. 2.2, we define probability spaces, and in Sect. 2.3, we define joint probability spaces.

2.1 Schemas, Databases, and Queries

We assume fixed countably infinite sets of constants Δ and *nulls* Δ_N, such that $\Delta \cap \Delta_N = \emptyset$ as elementary ingredients of *databases*. We also assume a countable infinite set of variables $\mathcal{X}$ as additional elementary ingredients for *queries*. We assume the *unique name assumption* for constants, i.e., different constants represent different values. Different nulls may represent the same value.

A *schema* is a finite sequence $\mathbf{R} = \langle R_1, \ldots, R_k \rangle$ of different relation symbols, each with a fixed arity $r_i > 0$. A database instance d is defined over $\mathbf{R}$ as a sequence $R_1^d, \ldots, R_k^d$ such that each R_i^d is a finite relation (i.e., a finite set of atoms with relation or predicate symbol R_i^d) of arity r_i over $\Delta \cup \Delta_N$ (i.e., the atoms of the database consist of constants and nulls).

A *conjunctive query (CQ)* over $\mathbf{R}$ has the form $Q(\mathbf{X}) = \exists \mathbf{Y} \Psi(\mathbf{X}, \mathbf{Y})$, where $\Psi(\mathbf{X}, \mathbf{Y})$ is a conjunction of atoms with the variables $\mathbf{X}$ and $\mathbf{Y}$, and possibly constants, but without nulls. We call $\mathbf{X}$ the *query variables*. A *Boolean CQ (BCQ)* over $\mathbf{R}$ is a CQ of the form $Q() = \exists \mathbf{Y} \Psi(\mathbf{Y})$, i.e., it contains no query variables.

Answers to CQs and BCQs are defined as *homomorphisms*, i.e., mappings $h : \Delta \cup \Delta_N \cup \mathcal{X} \rightarrow \Delta \cup \Delta_N \cup \mathcal{X}$ such that (i) $c \in \Delta$ implies $h(c) = c$, (ii) $c \in \Delta_N$ implies $h(c) \in \Delta \cup \Delta_N$, and (iii) h is naturally extended to atoms, sets of atoms, and conjunctions of atoms. The set of all answers to a CQ $Q(\mathbf{X}) = \exists \mathbf{Y} \Psi(\mathbf{X}, \mathbf{Y})$ over a database d, denoted $Q(d)$, is the set of all tuples $\mathbf{t}$ over Δ for which there exists a homomorphism $h_Q : \mathbf{X} \cup \mathbf{Y} \rightarrow \Delta \cup \Delta_N$ such that $h_Q(\Psi(\mathbf{X}, \mathbf{Y})) \subseteq d$ and $h_Q(\mathbf{X}) = \mathbf{t}$. The answer to a BCQ $Q() = \exists \mathbf{Y} \Psi(\mathbf{Y})$ over a database d is *Yes*, denoted $d \models Q$, iff $Q(d) \neq \emptyset$, i.e., there exists a homomorphism $h_Q : \mathbf{Y} \rightarrow \Delta \cup \Delta_N$ such that $h_Q(\Psi(\mathbf{Y})) \subseteq d$.

2.2 Probability Spaces

We consider probability spaces over sets with countable many elements (i.e., finite or countably infinite many elements). A probability space is a pair $\tilde{\mathcal{U}} = (\Omega(\tilde{\mathcal{U}}), p_{\tilde{\mathcal{U}}})$, where $\Omega(\tilde{\mathcal{U}})$ is a countable set, and $p_{\tilde{\mathcal{U}}}$ is a *probability function* defined over the countable set $\Omega(\tilde{\mathcal{U}})$ with the following two conditions: (i) $p_{\tilde{\mathcal{U}}} : \Omega(\tilde{\mathcal{U}}) \rightarrow [0, 1]$ and (ii) $\sum_{u \in \Omega(\tilde{\mathcal{U}})} p_{\tilde{\mathcal{U}}}(u) = 1$.

$\Omega(\tilde{\mathcal{U}})$ is a *sample space* with each member of the set $\Omega(\tilde{\mathcal{U}})$ being a *sample*. The set of all samples $u \in \Omega(\tilde{\mathcal{U}})$ such that $p_{\tilde{\mathcal{U}}}(u) > 0$ is called the *support* of $\tilde{\mathcal{U}}$, denoted $\Omega_+(\tilde{\mathcal{U}})$. A probability space $\tilde{\mathcal{U}}$ is *finite* if its support $\Omega_+(\tilde{\mathcal{U}})$ is finite, i.e., if the number of samples with a positive probability is finite.

A subset $X \subseteq \Omega(\tilde{\mathcal{U}})$ is an *event*. Each member of the set $\Omega(\tilde{\mathcal{U}})$ then is not only a sample, but also an event of size one, also called *elementary event*. The probability of an event $X \subseteq \Omega(\tilde{\mathcal{U}})$, denoted $Pr_{\tilde{\mathcal{U}}}(X)$, is defined as $\sum_{u \in X} p_{\tilde{\mathcal{U}}}(u)$.

A *random variable* is a function defined on the sample space or event space, respectively, like discussed in [1]. With $\mathcal{U}$ we denote the *random variable* that represents a sample or event, respectively, of $\tilde{\mathcal{U}}$. Observe that a sample or an event, respectively, over $\mathcal{U}$ can be represented by a logical formula over the different consistent classes of pairs of observations (see, e.g., [1]). Then, given a logical formula $\varphi(\mathcal{U})$, $Pr(\varphi(\mathcal{U}))$ corresponds to $Pr_{\tilde{\mathcal{U}}}(\{u \in \Omega(\tilde{\mathcal{U}}) | \varphi(\mathcal{U})\})$.

2.3 Joint Probability Spaces

We consider joint probability spaces as follows. We assume $U_1, \ldots U_n$ with $n \geq 1$ to be countable sets. W.l.o.g., let $\tilde{\mathcal{P}}$ be a probability space over $U_1 \times U_2$. Then, $\tilde{\mathcal{P}} = (\Omega(\tilde{\mathcal{P}}), p_{\tilde{\mathcal{P}}})$ is the joint probability space over the set $\Omega(\tilde{\mathcal{P}}) = U_1 \times U_2$, which is the joint sample space. The *left marginal* of the probability space $\tilde{\mathcal{P}}$ is then defined to be the probability space $\tilde{\mathcal{U}_1} = (\Omega(\tilde{\mathcal{U}_1}), p_{\tilde{\mathcal{U}_1}})$ such that (i) $\Omega(\tilde{\mathcal{U}_1}) = U_1$ and (ii) for all $u \in U_1$, it holds that $p_{\tilde{\mathcal{U}_1}}(u) = \sum_{v \in U_2} p_{\tilde{\mathcal{P}}}(u, v)$. In an analogous way, the *right marginal* of the probability space $\tilde{\mathcal{P}}$ is defined to be the probability space $\tilde{\mathcal{U}_2} = (\Omega(\tilde{\mathcal{U}_2}), p_{\tilde{\mathcal{U}_2}})$ such that (i) $\Omega(\tilde{\mathcal{U}_2}) = U_2$ and (ii) for all $v \in U_2$, it holds that $p_{\tilde{\mathcal{U}_2}}(v) = \sum_{u \in U_1} p_{\tilde{\mathcal{P}}}(u, v)$.

3 Modelling Probabilistic Contexts

In the following, we assume a MAS with a finite set $\mathcal{A}$ of agents to be given. Each of these agents, $a_i \in \mathcal{A}$, is equipped with its own context c_i, which we assume to be a structured data set. This data set is supposed to have been collected automatically by the respective agent according to its set of goals and strategies. For example, given appropriate goals and strategies, it may have crawled the Web or large data and/or document sets. A system that inspired our MAS is, e.g., DIADEM [12], which is a system gathering structured data from the Web and storing it in a database. Another system extracting structured SQL-like databases from unstructured text and tables (Dark Data, see also [13]) is DeepDive [14,15]. Observe that both do not use MAS technology, but could be part of MAS technology. Also, while DIADEM stores the gathered data in a deterministic database, DeepDive works with a probabilistic database.

Since the structured data set or context has been collected automatically, it is uncertain. For further processing, we keep the uncertainty with the context information, since we consider it essential additional information about the reliability of the context data. With relational databases being the most established and optimized means to store and retrieve data, we assume each agent to maintain a probabilistic relational database for storing their gathered context data.

It follows that a probabilistic model for a MAS involving a set of agents $\mathcal{A}$ consists of a set of pairs (a_i, c_i), where each agent $a_i \in \mathcal{A}$ maintains its own probabilistic context $c_i \in C$ with C being the set of probabilistic contexts of all agents in $\mathcal{A}$. Since we model the contextual data space c_i as a probabilistic database, each c_i is the ith probabilistic database in $\mathcal{D}$, with $\mathcal{D}$ being the set of all probabilistic databases in the MAS. W.l.o.g., we assume that the schemas of the probabilistic databases in $\mathcal{D}$ are pairwise disjoint, and so are the constants and the labelled nulls.

We now provide a formal definition for a probabilistic (contextual) database. For a comprehensive overview on probabilistic databases, see the textbook [2] and the more recent survey [7]. A probabilistic database adheres to the possible world semantics and consists of a set of possible facts F, which is a finite set of atoms with relation or predicate symbols $R_i^d \in \mathbf{R}$ of arity r_i over $\Delta \cup \Delta_N$. D is the set of the possible states of the database, also called possible worlds, $\{d_1, \ldots, d_l\}$, and is a set of subsets of the set of possible facts, i.e., $d_i \subseteq F$, and each subset is associated with a probability. Intuitively, the probabilistic database corresponds to a random variable where each of its finitely many states corresponds to a database and is associated with a probability, the probabilities of all its states adding up to 1. A more formal definition of a probabilistic context corresponding to a probabilistic database is given in the following definition.

Definition 1 ***(Probabilistic Context).*** A *probabilistic context* c is a *probabilistic database* over a schema $\mathbf{R} = \langle R_1, \ldots, R_k \rangle$ and a probability space $Pr = (D, \mu)$ such that D is the set of all (finitely many) databases over $\mathbf{R}$, and $\mu\colon D \to [0, 1]$ is a function that satisfies $\sum_{d \in D} \mu(d) = 1$.

Next, we consider how an agent $a_i \in \mathcal{A}$ may use its own context c_i. Typically, it queries it to decide what action to perform based on the most probable answers.

We now define answers to conjunctive queries (CQ) over a probabilistic context (or a probabilistic contextual database). As defined in the preliminaries, a *conjunctive query (or CQ)* is a first-order formula of the form $Q(\mathbf{X}) \leftarrow \exists \mathbf{y} \Psi(\mathbf{X}, \mathbf{Y})$ with $\Psi(\mathbf{X}, \mathbf{Y})$ being a conjunction of atomic formulas with the free variables $\mathbf{X}$. Answering a CQ Q over a probabilistic database Pr yields a set of pairs, each pair consisting of a possible answer to the query in the set of the possible databases $\{d_1, \ldots, d_l\}$ along with the probability that it is part of the answer to the query. The following provides a formal definition of answering CQs over a probabilistic context (or probabilistic contextual database).

Definition 2 ***(CQ-Answering over a Probabilistic Context).*** The *answer to a CQ Q* over a probabilistic context, which corresponds to a probabilistic database $Pr = (D, \mu)$, denoted $Q_{Pr}(D)$, is a set of pairs $\{(t_i, p_i)\}$ such that t_i is a tuple from Δ^r with r being the arity of Q, and p_i being the *marginal probability* of t_i with respect to the query Q over the probabilistic database, which is computed as $p_i = \sum_{d \in D : d \models Q(t_i)} \mu(d)$.

Example 1 presents how an agent that has collected information from Web pages of universities may utilize its context to decide whether to further collect information.

Example 1. Figure 1 shows the context c of an agent a, which has crawled the Web pages of universities and has collected information about researchers and publications. The context of the agent corresponds to a probabilistic (contextual) database $Pr = (D, \mu)$ storing information on publications of researchers of universities. The probabilistic database that the agent maintains at the moment is shown on the right of Fig. 1, and consists of five possible worlds $D = \{d_1, d_2, d_3, d_4, d_5\}$, each containing a subset of the set of possible facts shown on the left of Fig. 1. Each of the five worlds has a probability μ_i, and their sum is 1, i.e., $\sum_{d_i \in D} \mu(d_i) = 1$.

The database contains information about two researchers, Alice and Paul, both affiliated with the University of Bolzano, both having published on databases in the TODS journal. Alice has additionally published on machine learning in the Journal on Machine Learning Research; Paul has additionally published on artificial intelligence in the Artificial Intelligence Journal. The agent may utilize its probabilistic contextual database to ask for the probability of a researcher having published in the area of artificial intelligence by asking the following query $q_{Pr}(X) = \exists Y \text{Publication}(Y, \text{AI}, X)$. As can be easily verified, the answer is (AIJ, 0.5). The agent may find that a probability of 0.5 together with a single journal AIJ in the answers of the query over its context is too uninformative and/or too unreliable for further processing and may want to add more researchers and journals (and universities if it also queries for the universities $q_{Pr}(X) = \exists Y \text{Researcher}(Y, X)$, yielding (UniBZ, 1.0)) publishing about the area of artificial intelligence to its context.

Possible facts	
r_a	*Researcher*(Alice, UniBZ)
r_p	*Researcher*(Paul, UniBZ)
p_{aml}	*Publication*(Alice, ML, JMLR)
p_{adb}	*Publication*(Alice, DB, TODS)
p_{pdb}	*Publication*(Paul, DB, TODS)
p_{pai}	*Publication*(Paul, AI, AIJ)

A probabilistic database $Pr = (D, \mu)$	
$d_1 = \{\mathrm{r}_a, \mathrm{r}_p, \mathrm{p}_{aml}, \mathrm{p}_{pdb}\}$	0.3
$d_2 = \{\mathrm{r}_a, \mathrm{r}_p, \mathrm{p}_{aml}, \mathrm{p}_{pai}\}$	0.3
$d_3 = \{\mathrm{r}_a, \mathrm{r}_p, \mathrm{p}_{adb}, \mathrm{p}_{pai}\}$	0.2
$d_4 = \{\mathrm{r}_a, \mathrm{r}_p, \mathrm{p}_{adb}, \mathrm{p}_{pdb}\}$	0.1
$d_5 = \{\mathrm{r}_a, \mathrm{p}_{adb}\}$	0.1

Fig. 1. Possible contextual facts (left) of an agent making up its probabilistic contextual database (right).

4 Matching Probabilistic Contexts

In this section, we define a context matching operator for probabilistic context databases that creates a joint distribution with given marginals. In Sect. 4.1, we define the matching operator, and in Sect. 4.2, we concentrate on some special shapes that these joint distributions can take.

4.1 Interaction Matching Operators

Recall that a MAS $\mathcal{M}$ corresponds to a finite set of pairs of agents and their probabilistic contexts $\mathcal{M} = \{(a_1, c_1), \ldots, (a_n, c_n)\}$ with the probabilistic context c_i being stored in and, hence, corresponding to the probabilistic database $\mathrm{Pr}_i = (D_i, \mu_i)$ over a schema $\mathbf{R}_i$. Note that each agent maintains its own, unique context, i.e., for $a_i \neq a_j$, it also holds that $c_i \neq c_j$. In addition, the schemas of the probabilistic context databases are disjoint ($R_i \cap R_j = \emptyset$) as are the constant symbols ($\Delta_i \cap \Delta_j = \emptyset$) and the null symbols ($\Delta_{N_i} \cap \Delta_{N_j} = \emptyset$).

We are now ready to look at the interaction between the autonomous agents and define a matching operator between contexts. An interaction consists of sharing access to contexts between agents, i.e., an agent shares its context database and its schema (Pr_i,$\mathbf{R}_i$) with other agents. A receiving agent needs to combine the shared context with its own contextual database. For this purpose, we propose a matching operator for creating a joint probability distribution over the shared contexts (or probabilistic contextual databases, respectively). In general, arbitrary many agents may be allowed to interact or share access to their contexts with each other at the same time, but for simplicity and ease of presentation, we define a matching operator for creating such a joint probability between only two or three agents or contexts, respectively, as defined below. Observe that these definitions can easily be generalized to an interaction between n agents (in the MAS).

Definition 3 ***(2-Interaction Matching Operator).*** Let c_V and c_W be two different contexts; let $\mathrm{Pr}_V = (V, \mu_V)$ and $\mathrm{Pr}_W = (W, \mu_W)$ be the two corresponding probabilistic context databases, and let $R_{V\times W} \subseteq \Omega(\mathrm{Pr}_V) \times \Omega(\mathrm{Pr}_W)$ be a binary relation, the matching relation. A probabilistic context match of V and W with respect to $R_{V\times W}$ is a joint probability space $\mathrm{Pr}_{V\times W}$ over $\Omega(\mathrm{Pr}_V)\times\Omega(\mathrm{Pr}_W)$ that satisfies the following conditions:

1. The V- and W-marginals of $\mathrm{Pr}_{V\times W}$ are Pr_V and Pr_W, respectively. That is,
   - $\sum_{v\in V} \mu_{V\times W}(v, w) = \mu_W(w)$ for all $w \in W$, and
   - $\sum_{w\in W} \mu_{V\times W}(v, w) = \mu_V(v)$ for all $v \in V$.
2. The support of $\Omega(\mathrm{Pr}_{V\times W})$ is contained in $R_{V\times W}$ (i.e., $\mathrm{Pr}(\mathrm{Pr}_{V\times W} \in R_{V\times W})$ $= 1$).

The second condition of Definition 3 above allows only joint probability spaces where no elements that do not belong to $R_{V\times W}$ have a positive probability. The study of discrete bivariate probability distributions with given marginals as defined above in Definition 3 has been proposed already in [3,4]. There, however, the construction for joint bivariate probability distributions with given marginals is different from ours. In particular, we additionally define a relation $R_{V\times W}$ that does not necessarily contain the whole joint space, but only a subset thereof, and we also restrict the support to be contained in it (see condition two in Definition 3). In addition, our univariate distributions are over sets of possible worlds of probabilistic databases that can be queried in the context of a MAS. In [20,22], we and other authors have employed such bivariate probability distributions in the context of exchanging data from a source database to a target database. Here, we are generalizing this construction to 3- to n-variate probability distributions with n being the number of member agents of the MAS. In addition, we are not exchanging data, but we are querying distributed probabilistic contextual databases.

The term *copulas* for multivariate probability distributions with given marginals has been coined in [5]. A lot of research effort has been devoted to continuous copulas with continuous given marginals, see, e.g., the textbook [16]. Our approach differs from copula models in (continuous) statistics in that we consider countable (and finite due to the finite support required) marginals and copulas. Observe that for discrete multivariate distributions with given marginals, many results do not carry over from the continuous case, see, e.g., [6].

Example 2 shows two different 2-interaction matches between two pairs of probabilistic context databases, each belonging to another agent.

Example 2. Figure 2 shows context matches between two pairs of contextual probabilistic databases.

The first pair is a match between the two probabilistic databases $Pr_D = (D, \mu_D)$ and $Pr_E = (E, \mu_E)$. Pr_D is defined over the databases $D = \{d_1, d_2, d_3, d_4, d_5\}$ with the marginal probabilities $\{\mu_D(d_1) = 0.5, \mu_D(d_2) = 0.2, \mu_D(d_3) = 0.15, \mu_D(d_4) = 0.075, \mu_D(d_5) = 0.075\}$. Pr_E is defined over the databases $E = \{e_1, e_2, e_3\}$ with marginal probabilities $\{\mu_E(e_1) = 0.55, \mu_E(e_2) = 0.1, \mu_E(e_3) = 0.35\}$. A 2-interaction match creating a joint probability with Pr_D and Pr_E is shown on the left of Fig. 2.

On right of Fig. 2, a context match between Pr_D and Pr_F is shown. Pr_D is defined as before. $Pr_F = (F, \mu_F)$ is defined over the databases $F = \{f_1, f_2, f_3, f_4\}$ with marginal probabilities $\{\mu_F(f_1) = 0.5, \mu_F(f_2) = 0.2, \mu_F(f_3) = 0.15, \mu_F(f_4) = 0.15\}$.

In both cases, a single arc between two databases shows the joint probability of these two databases. It can easily be verified that both matchings are proper copulas, i.e., the marginals of the joint distribution coincide with the databases that are combined.

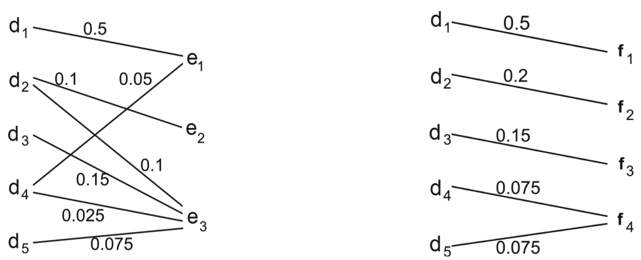

Fig. 2. Examples of results of a 2-interaction match between probabilistic contextual databases D and E (left) and D and F (right), respectively. See also Example 2 for more explanations.

We now generalize the 2-interaction matching operator defined for two interacting agents to three interacting agents below. When three agents interact, a trivariate probability distribution is generated.

Definition 4 ***(3-Interaction Matching Operator).*** Let c_U, c_V, and c_W be three different contexts; let $\mathrm{Pr}_U = (U, \mu_U)$, $\mathrm{Pr}_V = (V, \mu_V)$, and $\mathrm{Pr}_W = (W, \mu_W)$ be the three corresponding probabilistic context databases, and let $R_{U\times V\times W} \subseteq \Omega(\mathrm{Pr}_U) \times \Omega(\mathrm{Pr}_V) \times \Omega(\mathrm{Pr}_W)$ be a ternary relation, the matching relation. A probabilistic context match of U, V, and W with respect to $R_{U\times V\times W}$ is a joint probability space $\mathrm{Pr}_{U\times V\times W}$ over $\Omega(\mathrm{Pr}_U) \times \Omega(\mathrm{Pr}_V) \times \Omega(\mathrm{Pr}_W)$ that satisfies the following conditions:

1. The U-, V- and W-marginals of $\mathrm{Pr}_{U\times V\times W}$ are Pr_U, Pr_V, and Pr_W, respectively. That is,
   - $\sum_{v\in V}\sum_{w\in W} \mu_{U\times V\times W}(u,v,w) = \mu_U(u)$ for all $u \in U$, and
   - $\sum_{u\in U}\sum_{w\in W} \mu_{U\times V\times W}(u,v,w) = \mu_V(v)$ for all $v \in V$, and
   - $\sum_{u\in U}\sum_{v\in V} \mu_{U\times V\times W}(u,v,w) = \mu_W(w)$ for all $w \in W$.
2. The support of $\Omega(\mathrm{Pr}_{U\times V\times W})$ is contained in $R_{U\times V\times W}$ (i.e., $\mathrm{Pr}(\mathrm{Pr}_{U\times V\times W} \in R_{U\times V\times W}) = 1$).

An extension from three interacting agents to n interacting agents yielding an n-interaction matching operator is straight-forward.

In the terminology of copulas, we say that for an n-variate distribution $F \in \mathcal{F}(F_1, \ldots, F_m)$, with j-th univariate margin F_j, the copula associated with F is a distribution function $C : \mathbb{Q}^m_{[0,1]} \rightarrow \mathbb{Q}_{[0,1]}$ that satisfies $F(\mathbf{x}) = C(F_1(x_1), \ldots, F_m(x_m))$, with $\mathbf{x} \in \mathbb{Q}^m_{[0,1]}$, and $\mathbb{Q}_{[0,1]}$ being the interval of the rational numbers between 0 and 1.

4.2 Shapes of Joint Matching Distributions

The particular form of the joint distribution that will be created by the n-interaction context matching operators for the given (marginal) probabilistic contexts is shaped by the dependencies between the random variables that they join, i.e., between the probabilistic contexts of the agents that interact. In the following, we point out two special cases of shapes that the joint distributions for the 2-interaction and 3-interaction matching operators can take on.

A notable special case of a shape of joint distribution created by the n-interaction context matching operator is the *independence shape.* The *independence shape for the* 2*-interaction context matching operator* corresponds to the joint distribution $\mathrm{Pr}_{V\times W}$ with the product space $\Omega(\mathrm{Pr}_V) \times \Omega(\mathrm{Pr}_W)$ and the probability distribution $\mu_{V\times W}$ combining the two probabilistic context databases $\mathrm{Pr}_V = (\Omega(\mathrm{Pr}_V), \mu_V)$ and $\mathrm{Pr}_W = (\Omega(\mathrm{Pr}_W), \mu_W)$. In this special case, since independence holds, it follows that for all $v \in \Omega(\mathrm{Pr}_V)$ and all $w \in \Omega(\mathrm{Pr}_W)$, it holds that $\mu_{V\times W}(v,w) = \mu_V(v) \cdot \mu_W(w)$.

Similarly, the *independence shape for the* 3*-interaction context matching operator* corresponds to the joint distribution $\mathrm{Pr}_{U\times V\times W}$ with the product space

$\Omega(\mathrm{Pr}_U) \times \Omega(\mathrm{Pr}_V) \times \Omega(\mathrm{Pr}_W)$ and the probability distribution $\mu_{U\times V\times W}$ combining the three probabilistic context databases $\mathrm{Pr}_U = (\Omega(\mathrm{Pr}_U), \mu_U)$, $\mathrm{Pr}_V = (\Omega(\mathrm{Pr}_V), \mu_V)$, and $\mathrm{Pr}_W = (\Omega(\mathrm{Pr}_W), \mu_W)$. Again, since in this case independence holds, it follows that for all $u \in \Omega(\mathrm{Pr}_U)$, $v \in \Omega(\mathrm{Pr}_V)$, and all $w \in \Omega(\mathrm{Pr}_W)$, it holds that $\mu_{U\times V\times W}(u, v, w) = \mu_U(u) \cdot \mu_V(v) \cdot \mu_W(w)$.

Another notable shape is the *trivial* shape. For the 2-interaction context matching operator, there are two symmetric instances of this shape, namely the *$V-$trivial* and *$W-$trivial* shapes. For the 3-interaction context matching operator, there are three instances of this shape, *$U-$trivial*, *$V-$trivial* and *$W-$trivial* shapes. More specifically,

1. The joint probability space $\mathrm{Pr}_{V\times W} = (\Omega(\mathrm{Pr}_V) \times \Omega(\mathrm{Pr}_W), \mu_{V\times W})$ created by a 2-interaction context matching operator of the probabilistic context database $\mathrm{Pr}_V = (\Omega(\mathrm{Pr}_V), \mu_V)$ and the probabilistic context database $\mathrm{Pr}_W = (\Omega(\mathrm{Pr}_W), \mu_W)$ has a
   - *$V-$trivial* shape if for every $v \in \Omega(\mathrm{Pr}_V)$, there is exactly one $w \in \Omega(\mathrm{Pr}_W)$ such that $\mu_{V\times W}(v, w) > 0$; equivalently, $\mu_{V\times W}(v, w) = \mu_V(v)$ whenever $\mu_{V\times W}(v, w) > 0$;
   - *$W-$trivial* shape if for every $w \in \Omega(\mathrm{Pr}_W)$, there is exactly one $v \in \Omega(\mathrm{Pr}_V)$ such that $\mu_{V\times W}(v, w) > 0$; equivalently, $\mu_{V\times W}(v, w) = \mu_W(w)$ whenever $\mu_{V\times W}(v, w) > 0$.
2. The joint probability space $\mathrm{Pr}_{U\times V\times W} = (\Omega(\mathrm{Pr}_U) \times \Omega(\mathrm{Pr}_V) \times \Omega(\mathrm{Pr}_W), \mu_{U\times V\times W})$ created by a 3-interaction context matching operator of the probabilistic context databases $\mathrm{Pr}_U = (\Omega(\mathrm{Pr}_U), \mu_U)$, $\mathrm{Pr}_V = (\Omega(\mathrm{Pr}_V), \mu_V)$, and $\mathrm{Pr}_W = (\Omega(\mathrm{Pr}_W), \mu_W)$ has a
   - *$U-$trivial* shape if for every $u \in \Omega(\mathrm{Pr}_U)$, there is exactly one $v \in \Omega(\mathrm{Pr}_V)$ and exactly one $w \in \Omega(\mathrm{Pr}_W)$ such that $\mu_{U\times V\times W}(u, v, w) > 0$; equivalently, $\mu_{U\times V\times W}(u, v, w) = \mu_U(u)$ whenever $\mu_{U\times V\times W}(u, v, w) > 0$;
   - *$V-$trivial* shape if for every $v \in \Omega(\mathrm{Pr}_V)$, there is exactly one $u \in \Omega(\mathrm{Pr}_U)$ and exactly one $w \in \Omega(\mathrm{Pr}_W)$ such that $\mu_{U\times V\times W}(u, v, w) > 0$; equivalently, $\mu_{U\times V\times W}(u, v, w) = \mu_V(v)$ whenever $\mu_{U\times V\times W}(u, v, w) > 0$;
   - *$W-$trivial* shape if for every $w \in \Omega(\mathrm{Pr}_W)$, there is exactly one $u \in \Omega(\mathrm{Pr}_U)$ and exactly one $v \in \Omega(\mathrm{Pr}_V)$ such that $\mu_{U\times V\times W}(u, v, w) > 0$; equivalently, $\mu_{U\times V\times W}(u, v, w) = \mu_W(w)$ whenever $\mu_{U\times V\times W}(u, v, w) > 0$.

These shapes can be generalized in the obvious way to more interacting agents up to the n-interaction context matching operator, where all agents of the MAS interact and share access to their context with each other.

Below in Example 3, a trivial shape for the 2-interaction matching operator can be seen.

Example 3. Let's again consider the combination of the probabilistic context databases Pr_D and Pr_F of Example 2. It is obvious that the joint distribution that is shown on the right side of Fig. 2 is D-trivial, but not F-trivial.

5 Integrating Probabilistic Contexts

We now define query answering over combined probabilistic context databases with a joint n-variate distribution with n probabilistic context databases providing the given marginals as created by the n-interaction matching operators defined in the previous section. For this purpose, we first provide a short discussion on how to integrate the databases meaningfully with mappings in a general data integration scenario in Sect. 5.1. Then, in Sect. 5.2, we consider how to answer queries over the joint probabilistic context database which is created by an n-interaction matching operator.

5.1 Integration

To define query answering over combined probabilistic context databases, we need to not only combine the distributions, but also the vocabulary and the instances of the databases in a meaningful way. W.l.o.g., we assume that we have two agents, each with its context, (a_1, c_1) and (a_2, c_2), and a_1 shares its context with a_2 upon request. Since each context is a database with a schema, the agent receiving the request shares its schema and access to its database $(\mathbf{R}_1, Pr_1)$ with the other agent. Agent a_2 now has access to two sets of databases, D_1 from a_1 and Pr_1 and its own set of databases D_2 from Pr_2, and it also has access to the whole vocabulary of both schemas of both databases $\mathbf{R}_1 \cup \mathbf{R}_2$.

To combine two databases on a semantic level, we need to match the schemas and create schema mappings. We also need to disambiguate several occurences of an instance under different names. Let Δ_{c_1} be the set of instances (or constants) occuring in D_1 and let Δ_{c_2} be the set of instances (or constants) that occur in D_2. To reconcile instances, i.e. discover same instances occuring under different names, we need to match the sets of instances in Δ_{c_1} and Δ_{c_2} and rename several occurences of same instances under different names in order to maintain the unique name assumption. There is a large body of literature and research on schema and instance matching and data integration, see, e.g., [17]. Since a MAS is an integrated system, we assume a typical integration framework with a global schema $\mathbf{R}_{\mathcal{M}}$ used by all the agents of the set $\mathcal{A}$ of the MAS $\mathcal{M}$ with mappings between the global and the local schemas, so that queries over the global schema are translated to queries over the local schemas. In an integrated system with agents creating and growing their databases, a matching service (e.g. performed by an agent solely responsible for creating matchings and mappings and cleaning the databases) can run in the background and update the global schema and the mappings on the fly even before an interaction is about to happen.

We assume that the schema and instance matching are done with a black box procedure consisting of two subprocedures *Integration*(R_1, R_2) and *Integration*$(\Delta_{c_1}, \Delta_{c_2})$ having as output a set of correspondences Γ between relation symbols (and possibly attributes) and between instances. Once the integration and renaming of same instances are done, we can assume that we can rely on the unique name assumption and that we have a common vocabulary that we can use via a global schema and via adequately created mappings. Observe that the mappings can also be very simple correspondences that may also be probabilistic since they are created automatically; for more information on probabilistic data integration, see also [18]. Since we focus on the integration of the distributions over the probabilistic databases, the specific procedure for data integration and for creating the mappings or correspondences is not of much concern for us here. We simply assume a common vocabulary and simple correspondences between the elements of the schema and between the instances, respectively, to be given after the data integration procedure. We also remove the probabilities of the correspondences by using a threshold above that we define the correspondence to be true and below that we define the correspondence to be false.

Since we consider a general data integration framework with a global schema and only want to answer queries and perform distributed data access, we do not need to exchange or merge data, and, hence, we also do not need to worry about integrity constraints and resolving inconsistencies caused by integrity constraints of the databases like in the data exchange framework [19,20].

Summarizing, we assume the integration to happen as a black box procedure and the correspondences that we yield through the integration to be deterministic and simple constant-to-constant and relation-to-relation correspondences. We also assume that the databases are preprocessed by replacing the occurrences of one part of the correspondences with the other in the databases so that we can consider them matched and integrated in the classical understanding of data integration before we answer queries.

5.2 Query Answering

We now define query answering over a combined probabilistic context database created from n probabilistic context databases with the n-interaction context matching operator. We assume a common vocabulary which corresponds to a joint integrated relation $\mathbf{R} = \bigcup_{i \in 1,\ldots,n} \mathbf{R}_i$ as discussed in the former section. We also assume the databases to be preprocessed as discussed in the former section. Below we define query answering over combined databases created with the 2-interaction context matching operator.

Definition 5 ***(CQs over 2 Joint Probabilistic Databases).*** Let $\Pr_{V \times W} = (\Omega(\Pr_V) \times \Omega(\Pr_W), \mu_{V \times W})$ be a joint probabilistic context database created by a 2-interaction context matching operator from the probabilistic context databases $\Pr_V = (\Omega(\Pr_V), \mu_V)$ and $\Pr_W = (\Omega(\Pr_W), \mu_W)$. The *answer to a CQ Q* over the joint probabilistic context database $\Pr_{V \times W}$, denoted $Q_{\Pr_{V \times W}}$, is a set of pairs $\{(t_i, p_i)\}$ such that t_i is a tuple, and p_i is its *marginal probability* with respect to the query $Q_{\Pr_{V \times W}}$ over the combination of the two probabilistic context databases, which is computed as $p_i = p_{i,1} + p_{i,2} - p_{i,3}$ with the $p_{i,j}$ defined by:

$$p_{i,1} = \sum_{v_i \in V : v_i \models q(t_i)} \sum_{w \in W} \mu_{V \times W}(v_i, w)),$$
$$p_{i,2} = \sum_{w_j \in W : w_j \models q(t_i)} \sum_{v \in V} \mu_{V \times W}(v, w_j)),$$
$$p_{i,3} = \sum_{v_i \in V, w_j \in W : v_i \models q(t_i) \wedge w_j \models q(t_i)} \mu_{V \times W}(v_i, w_j)).$$

The above definition works according to the inclusion-exclusion principle and adds up all the joint probabilities where the tuple is entailed in one of the two combined databases, and then the probability that it is contained in both databases is subtracted.

Note that Definition 5 can be extended to combinations of three and more contextual databases.

6 Summary and Outlook

We have proposed a framework to model autonomous interacting agents in a multi-agent system (MAS). In the setting that we consider, the MAS is collecting data from the web and creates a set of connected and integrated distributed databases (the contexts) that can be queried collectively at all times during the collection and afterwards. In the collection phase, each agent autonomously collects and summarizes probabilistic data (as its context) on its own and stores it in its own probabilistic database that it maintains. Since the agents are supposed to complement each other, they need to interact and share access to their databases with each other in order to be able to adapt their data collection strategies and data analysis strategies according to the data the other agents in the MAS have already acquired. Agents can only collaborate and complement each other in the MAS, when they "know" what other agents in their MAS "know".

For the purpose of interacting agents, we have proposed to create n-variate joint distributions of database combinations corresponding to copula constructions, i.e., corresponding to joint probability distributions with given marginals. We consider a simple integration scenario and have defined query answering over the combined probabilistic contextual databases.

This paper is a first step to a more detailed study of data integration over probabilistic databases in the context of an information gathering MAS. We intend to develop the current approach further into several directions. One direction is to devise more elaborate communication and coordination strategies for the agents. Another direction (in connection with the other directions) is to add ontologies and study ontology-based data access in our setting, inspired by our previous work in probabilistic ontological data exchange, see, e.g., [21,22]. Considering more complex mappings and also probabilistic ones is also important. Finally, a prototype implementation is interesting as well.

References

1. Menger, K.: Random variables from the point of view of a general theory of variables. In: Proceedings of the 3rd Berkeley Symposium on Mathematical Statistics and Probability, Volume 2: Contributions to Probability Theory, pp. 215–229 (1956)
2. Suciu, D., Olteanu, D., Ré, C., Koch, C.: Probabilistic Databases. Morgan & Claypool, San Rafael (2011)
3. Fréchet, M.: Sur les tableaux de corrélation dont les marges sont donnés. Annales de l'Université de Lyon **4**, 53–57 (1951)

4. Morgensten, D.: Einfache Beispiele zweidimensionaler Verteilungen. Mitteilungsblatt fuer mathematische Statistik **8**, 234–235 (1959)
5. Sklar, A.: Fonctions de répartition à n dimensions et leur marges. Publication Statistical Institute de l'Université de Paris **8**, 229–231 (1959)
6. Genest, C., Neslehova, J.: A primer on copulas for count data. ASTIN Bul. J. IAA **37**(2), 475–515 (2007)
7. van den Broeck, G., Suciu, D.: Query processing on probabilistic data: a survey. Found. Trends Databases **7**(3–4), 197–341 (2017)
8. Rode, S., Turner, R.M.: Representing and communicating context in multiagent systems. In: Christiansen, H., Stojanovic, I., Papadopoulos, G.A. (eds.) CONTEXT 2015. LNCS (LNAI), vol. 9405, pp. 257–270. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-25591-0_19
9. Garant, D., da Silva, B.C., Lesser, V., Zhang, C.: Context-based concurrent experience sharing in multiagent systems. In: Proceedings of AAMAS 2017, pp. 1544–1546 (2017)
10. Dorri, A., Kanhere, S.S., Jurdak, R.: Multi-agent systems: a survey. IEEE Access **6**, 28573–28593 (2018)
11. Chen, F., Ren, W.: On the control of multi-agent systems: a survey. Found. Trends Syst. Control **6**(4), 339–499 (2019)
12. Furche, T., et al.: DIADEM: thousands of websites to a single database. PVLDB **7**(14), 1845–1856 (2014)
13. Govindaraju, V., Zhang, C., Ré, C.: Understanding tables in context using standard NLP toolkits. In: Proceedings of ACL 2013, pp. 658–664 (2013)
14. Zhang, C., Shin, J., Ré, C., Cafarella, M., Nui, F.: Extracting databases from dark data with DeepDive. In: Proceedings of SIGMOD 2016, pp. 847–859 (2016)
15. Zhang, C., Ré, C., Cafarella, M.J., Shin, J., Wang, F., Wu, S.: DeepDive: declarative knowledge base construction. Commun. ACM **60**(5), 93–102 (2017)
16. Nelson, R.B.: An Introduction to Copulas. Springer Series in Statistics, 2nd edn. Springer, New York (2006). https://doi.org/10.1007/0-387-28678-0
17. Bellahsene, Z., Bonifati, A., Rahm, E. (eds.): Schema Matching and Mapping. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-16518-4
18. Das Sarma, A., Dong, X.L., Halevy, A.Y.: Uncertainty in data integration and dataspace support platforms. In: Bellahsene, Z., Bonifati, A., Rahm, E. (eds.) Schema Matching and Mapping. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-16518-4_4
19. Fagin, R., Kolaitis, P.G., Miller, R.J., Popa, L.: Data exchange: semantics and query answering. Theor. Comput. Sci. **336**(1), 89–124 (2005)
20. Fagin, R., Kimelfeld, B., Kolaitis, P.G.: Probabilistic data exchange. J. ACM **58**(4), 1–55 (2011)
21. Lukasiewicz, T., Martinez, M.V., Predoiu, L., Simari, G.I.: Existential rules and Bayesian networks for probabilistic ontological data exchange. In: Bassiliades, N., Gottlob, G., Sadri, F., Paschke, A., Roman, D. (eds.) RuleML 2015. LNCS, vol. 9202, pp. 294–310. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-21542-6_19
22. Lukasiewicz, T., Martinez, M.V., Predoiu, L., Simari, G.I.: Basic probabilistic ontological data exchange with existential rules. In: Proceedings of AAAI 2016, pp. 1023–1029 (2016)

Why Polysemy Supports Radical Contextualism

François Recanati(✉)

Collège de France, PSL, 75005 Paris, France
frecanati@gmail.com

Abstract. After presenting two forms of Contextualism, I will argue that the phenomenon of polysemy supports the stronger one – so-called Radical Contextualism. My argument will be based on a comparison between indexicality and polysemy.

Keywords: Polysemy · Contextualism · Indexicality · Modulation

1 Meaning and Content

I take *content* to be fundamentally a property of mental states or acts (e.g. belief or judgment) and derivatively of speech acts. Consider someone who believes that elephants have wings, and expresses that belief by saying that elephants have wings. The proposition that elephants have wings is the content of her belief, as well as the content of the assertion she makes when she expresses that belief linguistically.

I take *meaning* to be a property of linguistic expressions (considered as types). The sentence 'Elephants have wings' has a certain meaning, and the words in that sentence do as well.

I take the debate between Literalism and Contextualism to bear on the relation between meaning and content. *Literalism* holds that they are the same thing. It accepts what I call the basic equation:

The basic equation:
meaning = content

Contextualism is the opposite view. It rejects the basic equation.

Because it accepts the basic equation, Literalism takes the meaning of the sentence-type 'Elephants have wings' to be the proposition that elephants have wings. The meaning of sub-sentential constituents is taken to be their contribution to the meaning/content of the sentences in which they occur, i.e. objects, properties, relations etc. or modes of presentation thereof.

2 Indexicals

Indexicals constitute an obvious counter-example to the basic equation. Their linguistic meaning is not the same thing as their content. Their linguistic meaning is invariant,

G. Bella and P. Bouquet (Eds.): CONTEXT 2019, LNAI 11939, pp. 216–222, 2019.
https://doi.org/10.1007/978-3-030-34974-5_18

while their content is contextually variable (whether we take that content to be an object or a Fregean sense). In Kaplan's influential framework, the meaning of an indexical is a 'character' that determines the content carried by the indexical in context (Kaplan 1989).

Indexicals are not sufficient to arbitrate the dispute between Literalism and Contextualism, however. They are not sufficient because Literalists acknowledge that indexicals constitute an exception to the basic equation. *In the case of indexicals, meaning ≠ content.* Literalists accept that. Still, they maintain the basic equation as the *default*, while Contextualists reject the basic equation, even construed as the default.

3 From Literalism to Methodological Contextualism

Contextualism comes in several varieties (Recanati 2005, 2012). I distinguish between methodological and substantial forms of Contextualism, and between two substantial forms (Fig. 1).

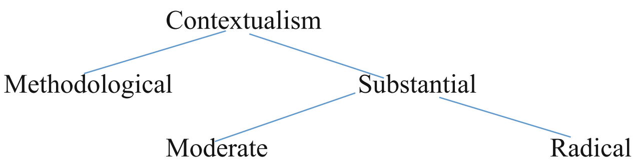

Fig. 1. Varieties of Contextualism.

The weakest form of Contextualism is Methodological Contextualism. It contrasts with Literalism in the following manner. Literalists take the indexical exception to be well-circumscribed: there is a list of expressions known to be indexical (the so-called 'basic set'),[1] and for the expressions that are not in that list the basic equation holds. So the default is: meaning = content (unless we are dealing with an expression in the list). According to Methodological Contextualism, however, we don't know in advance which expressions are indexical and which aren't. Ahead of inquiry, we should assume (by default) that meaning ≠ content, for indexicality is always a possibility. So Methodological Contextualism *reverses* what Literalism takes to be the default.[2]

4 Substantial Forms of Contextualism

In its substantial forms, Contextualism considers that indexicals are not an 'exception': context-sensitivity generalizes to *all* expressions (whether indexical or not). All expressions are such that the content they contribute depends upon the context, in contrast to the (invariant) linguistic meaning of the expression.

There are two forms of Contextualism that count as substantial by my characterization. One is moderate, the other radical. Each appeals to a particular phenomenon. Moderate Contextualism appeals to the phenomenon of *modulation*, while Radical Contextualism appeals to the phenomenon of *polysemy*.

[1] The expression 'basic set' comes from Cappelen and Lepore 2005: 1–2.
[2] On Methodological Contextualism, see Recanati 1994 and 2004: 160.

5 Moderate Contextualism

The meaning of an indexical is gappy and calls for a contextual process of saturation (e.g. an assignment of values to free variables in logical form). That process is mandatory: without saturation, no content can be assigned to an indexical expression. According to Moderate Contextualism, however, there is *another* contextual process that takes place on the way from meaning to content: modulation.

Modulation covers processes of sense extension (loosening/broadening) and sense narrowing (enrichment) as well as semantic transfer (metonymy) and possibly other phenomena (see Recanati 2004, chapter 2 for an overview). It is hard to deny that a sentence like 'The ham sandwich stinks' carries distinct truth-conditional contents depending on whether the description is taken literally as referring to the sandwich or metonymically as referring to the person who ordered it. Similarly, 'John is crazy' carries distinct truth-conditions when 'crazy' is taken literally and when it is a hyperbole. So context-sensitivity generalizes: Just as the content of an indexical depends upon the context of use, the content *actually* carried by an ordinary, non indexical expression also depends upon the context: it depends on *whether, and how,* the literal meaning of the expression is 'modulated' in context.

6 Radical Contextualism

What makes Moderate Contextualism moderate is the fact that, in contrast to saturation, modulation is *optional:* it may or may not take place. Whether or not it takes place depends upon the context, so the possibility of 'zero-modulation' (Recanati 2010: 44–45) is compatible with the generalization of context-sensitivity characteristic of substantial forms of Contextualism.

Cases of zero-modulation correspond to *literal* language use. In such cases, according to Moderate Contextualism, the basic equation still holds: meaning = content. According to Radical Contextualism, however, meaning is *never* identical to content. Lexical meaning is *constitutively unable* to figure as a constituent of content; it does not have the proper format for that ('wrong format view'—see Recanati 2004, chapter 9). This is where polysemy comes into the picture.

7 Polysemy as Ambiguity

As soon as an expression comes into public use, it becomes polysemous – the more frequent its use, the more polysemous it is. The senses of a polysemous expression result from pragmatic modulation (one sense is a modulation of another) but these modulations have become conventionalized and the senses of a polysemous expression are stored in the memory of language users (Benveniste 1974: 227; Recanati 2004: 135).

Since the senses of a polysemous expression are conventionalized (in contrast to *novel* instances of modulation), it is tempting to construe polysemous expressions as straightforwardly *ambiguous*, i.e. as expressions endowed with a multiplicity of meanings (Fig. 2).

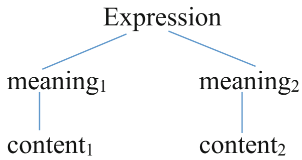

Fig. 2. The ambiguity model.

Ambiguous expressions contribute different contents in different contexts, but *this does not threaten the literalist equation of meaning and content* (since ambiguous expressions possess *distinct* meanings). This contrasts with the case of indexicals which possess a single, invariant meaning that determines different contents in different contexts.

8 Objection: Polysemy vs Homonymy

Two homonymous expressions (e.g. 'bank' and 'bank') are different expressions, with the same phonological realization but distinct meanings. A polysemous expression is supposed to be something else. A polysemous expression admittedly carries distinct senses, but *these senses are felt as related*: they form a *family* of senses. So instead of two different expressions with the same shape but distinct meanings (homonymy), what we seem to have is a single expression, i.e. a *semantic* as well as a phonological unit (polysemy): The expression has a single meaning which (depending on one's theory) either accounts for, or supervenes on, the diversity of its conventional uses.

If we don't allow polysemous expressions such an inherent meaning, distinct from the various senses they contribute in context, we are bound to *deny* that there is a difference between polysemy and homonymy. In other words: either polysemy does not exist (as a phenomenon distinct from homonymy), or, if it exists, it cannot be accounted for along the lines of the ambiguity model.

I call the alternative model we need the 'context-sensitivity model' (Fig. 3) because it posits a single meaning to which there correspond different contents in different contexts (as in the case of indexicals).

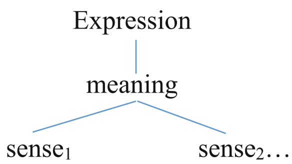

Fig. 3. The context-sensitivity model.

9 What Is the Unitary Meaning of a Polysemous Expression?

For Ruhl (1989), polysemous words possess a highly abstract (and underspecified) meaning which they carry in *all* their occurrences and which is responsible for the various senses they contextually express. Because the underspecified meaning lies below the level of consciousness, what intuitions reveal (and dictionaries record) are the expressed senses. These senses depend on context (both linguistic and extralinguistic), while 'a word's semantics should concern what it contributes in all contexts' (Ruhl 1989: 87). The task of the theorist is to *discover* lexical meaning by extracting from the data some abstract, unitary schema which *all* the uses fit.

I see two problems with this approach. First, it's not clear how it handles metonymies (which Ruhl hardly mentions). Second, even though polysemous expressions are *conventionally* associated with determinate senses which they regularly convey, these senses are not an aspect of the *linguistic meaning* of the expression, in Ruhl's framework. (The linguistic meaning is more abstract than these senses.) This is similar to the idea, floated in the seventies, that there are 'conventions of use' that are not 'meaning conventions': e.g. the convention that 'Can you pass the salt?' is a request that should be complied with rather than a question that should be answered.[3] But this construal of 'meaning' is overly narrow. As Langacker emphasizes, our goal as meaning theorists should be 'to properly characterize a speaker's knowledge of linguistic convention' (Langacker 1991: 268). Now,

> A lexical item of frequent occurrence displays a substantial, often impressive variety of interrelated senses and conventionally sanctioned usages… *Even when all its attested values are plausibly analysed as instantiations of a single abstract schema, or as extensions from a single prototype, there is no way to predict from the schema or prototype alone precisely which array of instantiations or extensions — out of all the conceivable ones – happen to be conventionally exploited within the speech community*. (Langacker 1987: 370; emphasis mine)

For these reasons, following Langacker,[4] I take the meaning of a polysemous expression to be neither a 'prototype' nor an 'all-subsuming superschema', but *the network of senses the expression is conventionally associated with* (including the prototype and/or the superschema, should there be any, as well as the modulation relations between the

[3] See Searle 1975 and especially Morgan 1978.

[4] "A strict reductionist approach would seek maximum economy by positing a single structure to represent the meaning of a lexical category. However, if our goal is to properly characterize a speaker's knowledge of linguistic convention, any such account is unworkable. From neither the category prototype alone, nor from an all-subsuming superschema (should there be one), is it possible to predict the exact array of extended or specialised values conventionally associated with a lexeme (out af all those values that are cognitively plausible). A speaker must learn specifically, for instance, that *run* is predicated of people, animals, engines, water, hosiery, noses, and candidates for political office; the conventions of English might well be different. Equally deficient is the atomistic approach of treating the individual senses as distinct and unrelated lexical items. The claim of massive homonymy implied by such an analysis is simply unwarranted—it is not by accident, but rather by virtue of intuitively evident relationships, that the meanings are symbolized by the same form. A network representation provides all the necessary information: an inventory of senses describing the expression's conventional range of usage; the relationships these senses bear to one another; schemas expressing the generalizations supported by a given range of values; and specifications of distance and cognitive salience." (Langacker 1991: 268).

senses). As can be seen by comparing Fig. 4 below with Figs. 2 and 3, the network model blends features from the ambiguity model and the context-sensitivity model (Recanati 2017).

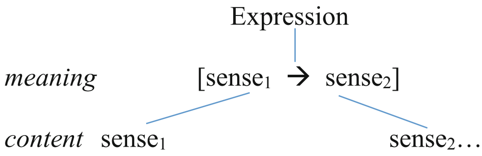

Fig. 4. The network model.

10 Conversion into Sense

In language use, senses multiply and diversify through modulation operations, which are optional in the sense of context-driven. Think of the first time the word ‘swallow’ was used to refer to what an ATM sometimes does with credit cards. The sense of ‘swallow’ was then creatively extended so as to exploit the similarity between the ATM situation and ordinary swallowing-situations. What was extended (the input to modulation) was the standard sense of ‘swallow’ as it applies to living organisms with a digestive system. The output of modulation was the (broadened) sense in which an ATM can be said to swallow a credit card. As a result of conventionalization, the extended sense has become part of the network of senses which makes up the lexical meaning of ‘swallow’, but the modulation relation between the extended sense and the prototypical sense is still alive in the consciousness of the language users. So we must distinguish between three things:

(i) the lexical meaning of ‘swallow’, which has the wrong format for being a constituent of content (it is or comprises a network of senses);
(ii) the standard/literal/prototypical sense of ‘swallow’ (with respect to living organisms), which was the input to modulation in the ATM example;
(iii) the extended sense relevant to ATMs, which was, and is still perceived as, the output of modulation.

On that view there is a principled difference between the linguistic meaning of a polysemous expression and the sense the expression contributes when used in context (even if that sense is the standard or prototypical sense). *Context-sensitivity thus generalizes in a way which supports Radical Contextualism.* An expression cannot directly contribute its lexical meaning, which has the ‘wrong format’ for being a constituent of content. The lexical meaning must be contextually converted into an appropriate sense through various context-sensitive operations (typically a mixture of sense selection and modulation).

11 Conclusion: Three Types of Contextual Process

We must distinguish the relation between the lexical meaning of ‘swallow’ and the extended sense the word takes in ATM-situations, namely a special case of conversion into sense, from the relation between that extended sense and the standard, prototypical sense of ‘swallow’, namely modulation. That distinction tends to be neglected because it is often (wrongly) assumed that the lexical meaning of ‘swallow’ *is* its standard/prototypical sense. But the distinction between modulation and conversion into sense is important because *conversion into sense is mandatory, while modulation is optional.* This makes conversion into sense similar to saturation (and polysemy similar to indexicality). Still, conversion into sense concerns all expressions (to the extent that, to a greater or lesser degree, all expressions are polysemous). This makes it similar to modulation (which may affect any expression), while saturation only concerns indexical expressions.

In terms of these two features—mandatoriness and universality—we can characterize the three contextual processes that map meaning to content (Table 1).

Table 1. Saturation, modulation, and conversion into sense.

	Mandatory	Universal
Saturation	+	−
Modulation	−	+
Conversion into sense	+	+

References

Benveniste, E.: Problèmes de linguistique générale, II. Gallimard, Paris (1974)
Cappelen, H., Lepore, E.: Insensitive Semantics. Blackwell, Oxford (2005)
Kaplan, D.: Demonstratives. In: Almog, J., Perry, J., Wettstein, H. (eds.) Themes from Kaplan, pp. 481–563. Oxford University Press, Oxford (1989)
Langacker, R.: Foundations of Cognitive Grammar. Stanford University Press, Palo Alto (1987)
Langacker, R.: Concept, Image and Symbol. Mouton De Gruyter, Berlin (1991)
Morgan, J.: Two types of convention in indirect speech acts. Syntax Semant. **9**, 261–280 (1978)
Recanati, F.: Contextualism and anti-contextualism in the philosophy of language. In: Tsohatzidis, S. (ed.) Foundations of Speech Act Theory, pp. 156–166. Routledge, London (1994)
Recanati, F.: Literal Meaning. Cambridge University Press, Cambridge (2004)
Recanati, F.: Literalism and contextualism: some varieties. In: Preyer, G., Peter, G. (eds.) Contextualism in Philosophy: Knowledge, Meaning, and Truth, pp. 171–196. Clarendon Press, Oxford (2005)
Recanati, F.: Truth-Conditional Pragmatics. Clarendon Press, Oxford (2010)
Recanati, F.: Contextualism: some varieties. In: Allen, K., Jacszolt, K. (eds.) The Cambridge Handbook of Pragmatics, pp. 135–149. Cambridge University Press, Cambridge (2012)
Recanati, F.: Contextualism and polysemy. Dialectica **71**, 379–397 (2017)
Ruhl, C.: On Monosemy: A Study in Linguistic Semantics. State University of New York Press, Albany (1989)
Searle, J.: Indirect speech acts. Syntax Semant. **3**, 59–82 (1975)

Generics in Context: The *Robustness* and the *Explanatory* Implicatures

Martina Rosola(✉)

FiNO Consortium - Università degli studi di Genova,
Via Balbi, 5, 16126 Genova, GE, Italy
martina.rosola@gmail.com

Abstract. Generics are sentences that express generalizations about a category or about its members. They display a characteristic context-sensitivity: the same generic can express a statistical regularity, a principled connection, or a norm. Sally Haslanger (2014) argues that this phenomenon depends on the implicit content that generics carry in different contexts.

I elaborate on Haslanger's proposal, arguing that the implicit content of generics is complex and constituted by two different propositions. A first proposition, that I here call the *robustness proposition*, characterizes as robust the link between the category the generic is about and the predicated property. This proposition is relatively invariant and is, as I claimed elsewhere, a *generalized* conversational implicature. In this paper, I will argue that a second implicature, the '*explanatory implicature*', arises which crucially depends on what explanation is called for in a certain context. Given its context-dependence, I conclude that this proposition is a *particularized* conversational implicature. While generics convey by default that the category and the property are strictly related, the specification of this relation hinges on the characteristics of the context in which the generic occurs.

1 Introduction

Generics are sentences that express generalizations about a category or about its members. Examples of generics are the following:

(1) a. Ducks lay eggs.
b. The mosquito is widespread.
c. A tiger is striped.

Generics are distinct from quantified sentences. For example, "some ducks lay eggs" is a quantified sentence, not a generic. In quantified sentences, a quantifier ('some', 'many', 'most', or 'all') explicitly states which proportion of Ks, namely the members of the category the sentence is about, have F, namely the predicated property. The subject Determiner Phrase (DP) of generics, however, lacks any explicit quantifier and it is debated whether a covert one should be posited.

G. Bella and P. Bouquet (Eds.): CONTEXT 2019, LNAI 11939, pp. 223–237, 2019.
https://doi.org/10.1007/978-3-030-34974-5_19

Generics can be distinguished into three categories: Bare Plurals, Definite Singulars, and Indefinite Singulars. The DP of Bare Plurals is constituted by a plural noun only, as in (1-a), that of Definite Singulars and Indefinite Singulars is constituted by a singular noun preceded respectively by a definite article, as in (1-b), or by an indefinite article, as in (1-c).

However, not every sentence with a Determiner Phrase of these forms is a generic. For example, (2) is about some salient lions and it makes no generalizations over lions. Therefore, it is not a generic.

(2) Lions are on the front lawn.

Generics display a characteristic context-sensitivity: the same generic can express a statistical regularity, a principled connection, or a norm. Consider, for example, (3):

(3) Philosophers are rational.

The above sentence can, according to the context, mean that a large proportion of philosophers are rational (statistical regularity), that being rational is a characteristic property of philosophers (principled connection), or that philosophers *should be rational* (social norm). Consider the following fictional context:

C1 The subjects' reactions were analyzed according to the profession. Contrary to most subjects, philosophers did not behave emotionally. The experimenters concluded that *philosophers are rational*.

It seems that, in this context, (3) simply points at a statistical fact: the experimenters did not observe a significant number of emotional reactions among philosophers. Suppose now that (3) is uttered in the following exchange:

C2 A: How can Jane be so impassive in this situation?
B: Philosophers are rational.

In this case, (3) has a different meaning: B is saying that being rational is somehow characteristic of philosophers. Finally, consider C3:

C3 You should calm down: *philosophers are rational*.

Here (3) expresses a social norm: it says that being rational is what a philosopher *should* do.

Haslanger [8] offers an account of this phenomenon. I will present her theory below.

1.1 Haslanger's Account

According to Haslanger, what allows (3) in context C2 to express a principled connection is the implicit content that the generic carries. She argues that generics, in certain contexts, convey that "the connection between the Ks and F holds primarily by virtue of some important fact about the Ks *as such*, or by virtue of *what it is to be a K*" ([8]: 370, emphasis in the original). That is, (3) in C2 conveys (4):

(4) The connection between philosophers and *being rational* holds primarily by virtue of some important fact about philosophers *as such*.

This implicit content interacts with two assumptions that, according to Haslanger ([8]: 379), humans generally hold:

Essentialist Assumption. Robust (meaningful?) regularities are not accidental. They are due to the nature of things.

Normative Assumption. Things should express their natures and under normal conditions they will. Abnormal circumstances are not good and should be avoided or changed.

The implicit content conveyed by sentence (3) in C2, joined with the *Essentialist Assumption* further licenses an essentialist statement about philosophers. That is, a statement claiming that being rational is grounded in the generic essence of philosophers. A *generic essence* is the essence of a category: the generic essence of a category K is what individual members of the category have essentially that makes them Ks. Essences explain facts, behaviors, and characteristics of individuals. Thus, Haslanger argues, generics import an explanation: membership to category K explains the possession of property F.

In some contexts, the *Normative Assumption* further induces "that it is right and good for [Ks] to be [F], and [Ks] that are not [F] are defective" ([8]: 380).[1] This is what happens in context C3: given the essentialist statement about philosophers and the *Normative Assumption*, (3) expresses a social norm. Since being rational is grounded in the generic essence of philosophers and, according to the *Normative Assumption*, things should express their natures, then philosophers *should be rational*.

What about C1? Generic (3) does not convey the implicit content carried in the other contexts. Consequently, the sentence expresses a mere statistical claim and no further specifications of the connection between the category of philosophers and the property of being rational are conveyed. Haslanger provides an explanation for why statistical generics (i.e. generics that express mere statistical regularities) lack the implicit content. She first hypothesizes this is due to a contradiction that would arise between the implicit content and preexisting background assumptions. However, she rules out this hypothesis: since implicit content can change the common ground, a contrast with background assumptions is not sufficient to prevent the content from arising. She concludes that "[g]iven that claims of generic essence are efforts to offer an explanation in terms of significant features of the kind in question, when an explanation is not being called for, there is no implicature" ([8]: 382).

[1] In the original text Haslanger uses 'F' for the category and 'G' for the property. For the sake of uniformity, I substituted the letters in the quotation to let 'K' for the category and 'F' for the property.

To sum up, according to Haslanger, if a generic is meant to provide an explanation for a certain fact, it conveys that "the connection between the Ks and F holds primarily by virtue of some important fact about the Ks *as such*". This implicit content combines further with the *Essentialist Assumption*, inducing an essentialist statement. In some contexts, the essentialist statement further combines with the *Normative Assumption*. As a result, the generic expresses a norm.

1.2 Refining Haslanger's Account

As Haslanger acknowledges, her analysis is not fully spelled out. In particular, she does not take a stand on the nature of the implicit content she hypothesizes. Even if she uses the term 'implicature' in some passages, in two footnotes she specifies that the implicit content of generic could be of a different kind:

> I'm actually not sure whether it is better to consider it an implication or a presupposition. I'm willing to adjust my account to accommodate evidence for either. My goal in this paper is programmatic and I am aware that much more work needs to be done on the details.
> ([8]: 370, footnote 9)

> It is a difficult and contested matter how to distinguish what enters the common ground through implicature and what enters through presupposition. For my purposes, little hinges on this [...] I will use the model of implicature to account for the examples we're looking at, but it may be that they are better handled differently.
> ([8]: 377, footnote 32)

In a previous paper, I argued that the implicit content of generics hypothesized by Haslanger is indeed an implicature. Specifically, I argued that it is a generalized conversational implicature. I based this claim on the observation that the implicature is quite constant across contexts. However, this seems to be only part of the story. Indeed, Haslanger argues that the demand for an explanation plays a role in generating the implicature. Thus, it seems that the content of the implicature will vary according to the explanation at stake in a certain context:

> [W]hen generics are asserted to provide an explanation of a phenomenon, they (defeasibly) implicate that there is an explanatorily robust relationship between the kind and the property indicated. Because there are different sorts of explanations that might be called for, the generic may implicate a specific kind of relation that is relevant to the particular form of explanation.
> ([8]: 382)

In this paper, I maintain that the implicit content of generics is complex and it is constituted by two different propositions. The first proposition characterizes the link between Ks and F as robust. This proposition is relatively invariant and is, I argue, a *generalized* conversational implicature. However, a second implicature arises which crucially depends on what explanation is called for. Given its dependence on the context, I argue that this proposition is a *particularized* conversational implicature.

In the next section, I will present my argument that generics convey a generalized conversational implicature that the connection between Ks and F is robust. Then, I will put forward my case for the particularized one. The content of the latter will depend on the context and, in particular, by what explanation is called for in a certain conversational exchange.

2 The *Robustness* Implicature

Haslanger [8] argues that generics convey that the connection between the Ks and F is robust. I will hereafter refer to this implicit content as the "*robustness proposition*", for the sake of brevity. As observed above, she claims that the *robustness proposition* can be either an implicature or a presupposition. Intuitively, a presupposition is what is taken for granted rather than being asserted. An implicature is, roughly, a defeasible consequence of a sentence.

Presuppositions can be semantic or pragmatic. A *semantic presupposition* of a sentence must be true in order for the sentence to be truth evaluable; a *pragmatic presupposition* is a proposition that the speaker assumes to be true. The latter is a broader notion: a speaker pragmatically presupposes all the semantic presuppositions of the sentences they utter, but not every pragmatic presupposition is semantic as well.

Implicatures can be conventional or conversational. *Conventional implicatures* are determined by the conventional meaning of the words used; *conversational implicatures* follow from assumptions underlying conversational exchanges. According to Paul Grice, who introduced the notion of implicatures, speakers conform to the Co-operative Principle, which prescribes to make one's conversational contribution "such as is required, at the stage at which it occurs, by the accepted purpose or direction of the talk exchange in which you are engaged" ([7]: 26). Grice further articulates this principle in four maxims: Quantity, Quality, Relation, and Manner. Conversational implicatures arise from the violation or the flout of a maxim.

Conversational implicatures can be distinguished further, according to the conditions in which they emerge. Those arising in all contexts, unless some unusual contextual assumptions defeat them, are generalized. Particularized conversational implicatures, instead, can arise only provided specific contextual assumptions.

In what follows, I will use some standard tests to determine what category does the *robustness proposition* belong to. I test for projective behavior to check whether it is a *semantic presupposition*; the 'Hey, wait a minute' test will determine if it is a *pragmatic presupposition*. The results reveal that *robustness proposition* is not a presupposition. Then, I test for implicatures. In particular, I check whether the *robustness proposition* can be canceled. Since it is cancelable, it is not a *conventional implicature*. Cancelability is a central feature of *conversational implicatures* and, as other tests show, the *robustness proposition* possesses all the main properties of conversational implicatures. I conclude that the *robustness proposition* is a conversational implicature.

Furthermore, I will take into account the conditions where the *robustness proposition* arises to determine whether it is a *particularized* or a *generalized* conversational implicature. Since it arises by default, I conclude that the *robustness proposition* is a generalized conversational implicature of generics. I argue that such implicature is generated by a general principle, Stephen Levinson's *I-principle*, that applies to unmarked forms.

2.1 Tests for Presupposition

One distinctive feature of semantic presuppositions is their ability to project. That is, if a sentence S presupposes P then so do larger structures embedding S, like negated sentences, questions, and conditionals. Let's check whether the *robustness proposition* conveyed by generic (3) projects:

(3) Philosophers are rational.

(4) [Presupposition?] The connection between philosophers and *being rational* holds primarily by virtue of some important fact about philosophers *as such*.

(5) **Test #1: projection**

a. Philosophers are not rational.
#The connection between philosophers and *being rational* holds primarily by virtue of some important fact about philosophers *as such*.

b. Are philosophers rational?
#The connection between philosophers and *being rational* holds primarily by virtue of some important fact about philosophers *as such*.

c. If philosophers are rational, then Hypatia was.
?The connection between philosophers and *being rational* holds primarily by virtue of some important fact about philosophers *as such*.

As the test shows, (4) does not project. Therefore, it is not a semantic presupposition. The *Hey, wait a minute* test [6,11] "is capable to detect pragmatic presupposition. For if a speaker takes a proposition ϕ for granted in uttering S, then one can object 'Hey wait a minute!'" ([5]: 1410). To show that, Katharina Felka proposes the following example, where speaker A *pragmatically* presupposes that their audience understands German:

(6) A: In der Küche gibt es Kaffee.
(Translation: there is coffee in the kitchen)
B: Hey, wait a minute. I don't understand German.
C: #Hey, wait a minute. I had no idea that there was coffee in the kitchen.

B's reply is appropriate since it questions A's pragmatic presupposition. On the contrary, C's reply is not appropriate since it questions what is asserted rather than what is presupposed. The inappropriateness of the *hey, wait a minute* reply to (3) reveals that the *robustness proposition* is not a pragmatic presupposition of the generic:

(7) **Test #2: hey, wait a minute**

A: Philosophers are rational.

B: #Hey, wait a minute. I had no idea that the connection between philosophers and *being rational* holds primarily by virtue of some important fact about philosophers *as such.*

The *robustness proposition*, as the tests show, is not a presupposition. In what follows, I show that the *robustness proposition* is cancelable, calculable, nondetachable, and indeterminable. I thus conclude that it is a conversational implicature.

2.2 Tests for Implicature

Contrary to logical implications and presuppositions, conversational implicatures "may be explicitly canceled, by the addition of a clause that states or implies that the speaker has opted out" ([7]: 39). For example, while (8-a) conversationally implicates (8-b), (9-a) does not since the addiction of "actually all of them did" cancels the implicature:

(8) a. Some students passed the exam.
b. [Conversational implicature:] *Not all* students passed the exam.

(9) a. Some students passed the exam, *actually all of them did.*
b. #*Not all* the students passed the exam.

Conventional implicatures, instead, are not cancelable:

(10) a. Jane is rich but nice.
b. [Conventional implicature:] Being rich is opposed to being nice.
c. # Jane is rich but nice, *and there is no contrast between being rich and being nice.*

The term 'but' induces the conventional implicature that being rich is opposed to being nice, and the added clause in (10-c) does not prevent the implicature from arising. Rather, it gives rise to a contradiction. Since the *robustness proposition* can be canceled, it is not a *conventional* implicature:

(11) **Test #3: cancelability**

a. Philosophers are rational *only in appearance.*
b. #The connection between philosophers and *being rational* holds primarily by virtue of some important fact about philosophers *as such.*

If the rationality of philosophers is only apparent, then it is not due to some important fact about philosophers *as such*: the clause in italics cancels the *robustness proposition*. This constitutes evidence that it is a *conversational* implicature. Furthermore, the *robustness proposition* possesses two other features of conversational implicatures: nondetachability and indeterminability.

A putative implicature "is nondetachable insofar as it is not possible to find another way of saying the same thing (or approximately the same thing) which simply lacks the implicature" ([7]: 43). The following rephrasing of (3) all carry the *robustness proposition*, showing that it is nondetachable:

(12) **Test #4: nondetachability**

a. Philosophers resist emotions.
b. Philosophers are cold.
c. Philosophers are impassive.
d. Philosophers follow reason.
e. ...

Moreover, as the test below shows, the *robustness proposition* is indeterminable. That is, the implicature conveyed by (3) may not be exactly determinable and it can be expressed in different ways, as for example:

(13) **Test #5: indeterminability**

a. The connection between philosophers and *being rational* is robust.
b. There is a strong link between philosophers and *being rational*.
c. There is something special about philosophers that makes them rational.
d. Philosophers *as such* are rational.
e. ...

Thus, the tests revealed that the *robustness proposition* is cancelable, nondetachable, and indeterminable. These, in addition to calculability, are the characteristic features of conversational implicatures. I will postpone the test for calculability since I need to introduce Levinson's I-principle first. Since the *robustness proposition* possesses these characteristic features, I conclude that it is a conversational implicature.

What kind of conversational implicature is it, generalized or particularized? Generalized conversational implicatures arise in all contexts, unless some unusual contextual assumptions defeat them; particularized conversational implicatures can arise only provided some specific contextual assumptions. Generic (3) carries the *robustness proposition* even in absence of a specific context, when uttered out of the blue. Moreover, the implicature is constant across different contexts. Consider contexts C2–C3 from above:

C2 A: How can Jane be so impassive in this situation?
B: Philosophers are rational.

C3 You should calm down: *philosophers are rational*.

Sentence (3) implicates in both contexts that the connection between philosophers and *being rational* holds primarily by virtue of some important fact about philosophers *as such*. The implicature is present in both contexts and its content does not depend on specific features of the context. Hence, I argue, the *robustness proposition* is a *generalized* conversational implicature.

My proposal is that what induces the implicature of (3), "philosophers are rational", is the generic construction. Specifically, I argue that this depends on the unmarkedness of generics: they are a very simple way of attributing a property to a category. A speaker uttering a generic provides no information concerning what links Ks and F. Thus, they allow the addressee to assume that "the connection between the Ks and F holds primarily by virtue of some important fact about the Ks *as such*". This depends on the assumption that what is simply described is stereotypically exemplified. Levinson [10] proposes that this assumption underlies our conversational exchanges. In particular, Levinson argues that an underlying principle, that he dubs 'I-principle', applies to unmarked forms enriching their content. I will present this principle and how it applies to generics in the next subsection.

2.3 The I-Implicature of Generics

Levinson [10] (pages 31–33) argues that generalized conversational implicatures depend on three heuristics:

Heuristic 1. What isn't said, isn't;
Heuristic 2. What is simply described is stereotypically exemplified;
Heuristic 3. What's said in an abnormal way, isn't normal;
or Marked message indicates marked situation.

The heuristics above serve to increase the informativeness of a sentence: they rule out more states of affairs than the sentence alone would. These heuristics, as Levinson points out, are intimately related to some of Grice's maxims. The first two heuristics are related to the Maxim of Quantity. Specifically, Heuristic 1 is related to the first submaxim of quantity, Q1: "make your contribution as informative as required (for the purposes of the exchange)" ([7]: 27), and Heuristic 2 to the second one, Q2: "do not make your contribution more informative than is required" ([7]). The third heuristic is related to the Maxim of Manner, "be perspicuous" ([7]), and in particular to the first and fourth submaxims, M1: "avoid obscurity of expression" and M4: "avoid prolixity".

Given the relation that the first heuristic has with the first Maxim of Quantity, Levinson dubs it *Q-heuristic*. He names the second one *I-heuristic*, to recall the "Informativeness Principle" of Atlas and Levinson [1], that is a version of this heuristic. Finally, *M-heuristic* is the one associated with the Maxim of Manner. I contend that the I-heuristic applies to generics, inducing the *robustness proposition*. Therefore, I will focus on this heuristic alone.

Implicatures derived from the I-heuristic, that Levinson labels 'I-implicatures', depend on a tendency towards economy that leads the speakers to avoid stating what is obvious. Consequently, the hearers are licensed to conclude from a simple description that the situation is stereotypical. Levinson spells out this principle in the following way ([10]: 114–115):

(14) ***I-principle***

Speaker's maxim: **the maxim of Minimization.** "Say as little as necessary"; that is, produce minimal linguistic information sufficient to achieve your communicational ends.

Recipient's corollary: **the Enrichment Rule.** Amplify the informational content of the speaker's utterance, by finding the most *specific* interpretation, unless the speaker has broken the maxim of Minimization by using a marked or prolix expression.

Specifically:

a. Assume the richest temporal, causal and referential connections between described situations or events, consistent with what is taken for granted.
b. Assume that stereotypical relations obtain between referents or events, unless this is inconsistent with (a).
c. Avoid interpretations that multiply entities referred to (assume referential parsimony); specifically, prefer coreferential readings of reduced NPs (pronouns or zeros).
d. Assume the existence or actuality of what a sentence is about if that is consistent with what is taken for granted.

Generics are unmarked: they are morphologically simple and brief, very common, and neutral in register. Consequently, the I-principle applies to generics, inducing an I-implicature. The I-principle invites the recipient to enrich the content of an utterance and, in particular, to "assume the richest temporal, causal and referential connection between situations or events". Generics, however, do not connect different situations or events. I argue that the assumption in this case concerns the relationship between the category and the property, I-implicating that it is robust. The underlying Gricean reasoning in this case being: if the speaker did not specify what relationship links the category and the property, then they meant the strongest connection, namely that "the connection between the Ks and F holds primarily by virtue of some important fact about the Ks *as such*". We can now see how the *robustness proposition* of (3) can be worked out:

(15) **Test #6: calculability**

a. The speaker uttered (3), "philosophers are rational".
b. Generics are unmarked and the I-principle applies to them.
c. By the I-principle, the recipient should assume the richest connection between the category and the property.

d. Thus, (3) implicates the *robustness proposition*, namely that the connection between philosophers and *being rational* holds primarily by virtue of some important fact about philosophers *as such*.

To recap, Haslanger argued that generics convey what I called the *robustness proposition*, namely that the connection between the Ks and F holds primarily by virtue of some important fact about the Ks *as such*. I argued, based on linguistic tests, that this implicit content of generics is a *generalized conversational implicature*. Levinson's I-principle applies to generics, due to their unmarkedness, inducing the implicature.

If this is correct, then every generics should carry the *robustness proposition*. However, statistical generics do not. According to Haslanger, statistical generics lack the implicature because they are not uttered to provide an explanation and the *robustness implicature* has an explanatory role. My proposal, though, is that the *robustness proposition* arises by default unless defeated and it is quite independent of whether an explanation is called for in a certain context. I claim that the reason why statistical generics lack the implicature may be due to a broader difference between statistical and non-statistical generics, as recognized since the first studies on generics [2–4,9]. Specifically, while non-statistical generics can occur with any generic determiner phrase (Bare Plural, Definite Singular, and Indefinite Singular), statistical generics cannot take an Indefinite Singular form. This fact is exemplified by the contrast between the following examples, taken from [9]:

(16) a. Madrigals are polyphonic.
b. The madrigal is polyphonic.
c. A madrigal is polyphonic.

(17) a. Madrigals are popular.
b. The madrigal is popular.
c. # A madrigal is popular.

The sentences in (16) express a principled connection and are all acceptable, but the indefinite singular (17-c), that expresses a statistical connection, does not receive a generic interpretation. Given the differences in distribution between statistical and non-statistical generics, I argue that we should analyze them separately and that my proposal applies to non-statistical generics only.

In the next section, I will argue that non-statistical generics convey an additional implicature. This implicature details the specific kind of relation connecting the category K and the predicated property F. As anticipated above, the content of this implicature changes according to the context. Specifically, it will depend on the form of explanation relevant in the context. For this reason, I label it the *explanatory implicature*.

3 The *Explanatory* Implicature

As seen above, Haslanger argues that "claims of generic essence are efforts to offer an explanation in terms of significant features of the kind in question" ([8]: 382) and that generics convey a claim of generic essence. Moreover, she claims that "generics may implicate *a specific kind of relation* that is relevant to the particular form of explanation" that is called for in a certain context ([8]: 382, emphasis added). Haslanger, following Aristotle, takes explanations to be of four kinds: formal, efficient, material, and teleological. Different explanations will be relevant in different contexts. Consider the examples below:[2]

C4 A: What defines a philosopher?
B: *Philosophers are rational.*

C5 a. A: How can Jane be so impassive in this situation?
B: *Philosophers are rational.*
b. A: What destroyed Notre-Dame?
B: *Beams burn.*

C6 A: Why is it prohibited to smoke inside?
B: *Beams burn.*

C7 a. A: Why is she withholding her emotions?
B: *Philosophers are rational.*
b. A: Why are they throwing the beams in the fireplace?
B: *Beams burn.*

Context C4 calls for a *formal* explanation; contexts C5 for an explanation in terms of an *agent* or *efficient cause*; a *material* explanation is relevant in context C6; and a *teleological* explanation is salient in contexts C7. The same generic occurs in three different contexts: "philosophers are rational" occurs in contexts C4, C5, and C7, thus working as, respectively, *formal*, *efficient cause*, and *teleological* explanation. "Beams burn", instead, occurs in contexts C5, C6, and C7, where it's offered as *efficient cause*, *material*, and *teleological* explanation, respectively.

My proposal is that the very same generic has a different implicature in each context and that it's this implicature that allows it to provide different kinds of explanations according to what is called for. This implicature, I argue, is additional to the *robustness proposition* that the generic carries by default and that is constant across contexts. Consider for example "philosophers are rational" as it occurs in context C4 above. It implicates, I claim, both the *robustness proposition* ('RP' for short) and the *explanatory implicature* ('EI' for short):

C4 A: What defines a philosopher?
B: *Philosophers are rational.*

[2] An immense thanks to Beatrice Michetti for helping out figuring out what I needed in these examples.

[RP:] The connection between philosophers and *being rational* holds primarily by virtue of some important fact about philosophers *as such.*
[EI:] *Being rational* is what makes one a philosopher.

Similarly, "philosophers are rational" carries both implicatures in contexts C5 and C7:

C5 a. A: How can Jane be so impassive in this situation?
B: *Philosophers are rational.*
[RP:] The connection between philosophers and *being rational* holds primarily by virtue of some important fact about philosophers *as such.*
[EI:] *Being rational* influenced the philosopher's reactions.

C7 a. A: Why is she withholding her emotions?
B: *Philosophers are rational.*
[RP:] The connection between philosophers and *being rational* holds primarily by virtue of some important fact about philosophers *as such.*
[EI:] *Being rational* is the purpose of philosophers.

As we can observe, the *explanatory implicature* changes according to the context, contrary to the *robustness proposition.* For this reason, I argue that the *explanatory implicature* is a *particularized* conversational implicature. In particular, I argue that this implicature arises from the flout of the maxim of Relation. Let's see how this works in context C4. I propose that the implicature can be worked out with a reasoning along these lines:

(18) **Test #6: calculability**

a. The speaker uttered (3), "philosophers are rational", as an answer to my request to define a philosopher.
b. By the maxim of Relation, (3) should be relevant to my request.
c. Therefore, the speaker intends (3) to express a definition of philosophers.
d. For this to be the case, the connection between the category and the property should be definitory.
e. Thus, the speaker has implicated that *being rational* is what makes one a philosopher.

Different relations between the category and the predicated property are relevant in each context C4–C7. Therefore, the specification of the connection implicated (and, thus, the content of the *explanatory implicature*) is different.

The implicature in question is cancelable too, as showed by the test below:

(19) **Test #3: cancelability**

C4 A: What defines a philosopher?
B: Philosophers are rational, *of course, but this is not a definitory feature of philosophers.*

The clause in italic cancels the implicature: B's utterance does not implicate that *being rational* is what makes one a philosopher. The tests I applied show that the *explanatory implicature* possesses two central features of conversational implicatures: cancelability and calculability.[3] I conclude, therefore, that it is a genuine implicature. Notice moreover that the addition in italic does not cancel the *robustness implicature*: B's reply still implicates that the connection between philosophers and *being rational* is robust. That it's possible to cancel one implicature while keeping the other proves that the *explanatory implicature* is distinct from the *robustness proposition*. Thus, the implicit content of generic is composed of *two* implicatures and neither reduces to the other.

4 Conclusions

In this paper, I elaborated on Haslanger's proposal that generics carry an implicit content concerning the connection between the kind and the predicated property in a generic. I argued that the implicit content hypothesized by Haslanger is complex and it is composed of two implicatures. A first implicature, that I labeled the *robustness proposition*, characterizes the connection between the kind and the property as robust. An additional one, that I dubbed the *explanatory implicature*, specifies the kind of relation in question. While the former is relatively invariable, the latter is highly dependent on the context. Therefore, I argued, they are a *generalized* and a *particularized* conversational implicatures, respectively.

In particular, I proposed that the *robustness proposition* is generated by a general assumption, what Levinson calls 'I-principle'. This principle invites the addressee to enrich utterances by assuming the richest connections. Since generics are unmarked, the I-principle applies to them, generating the *robustness proposition*. The *explanatory implicature*, instead, arises from the flout of the maxim of Relation. Since the relevant connections vary according to the context, so does the content of the implicature. As I showed with linguistic tests, both implicatures possess the main features of genuine implicatures. Finally, I showed that the two implicatures are genuinely distinct and should not be conflated.

References

1. Atlas, J.D., Levinson, S.C.: It-clefts, informativeness, and logical form: radical pragmatics (revised standard version). In: Cole, P. (ed.) Radical Pragmatics, pp. 1–62. Academic Press (1981)
2. Burton-Roberts, N.: Generic sentences and analyticity. Stud. Lang. **1**, 155–196 (1977)
3. Carlson, G.N., Pelletier, F.J. (eds.): The Generic Book. The University of Chicago Press, Chicago (1995)
4. Dahl, Ö.: On generics. In: Keenan, E.L. (ed.) Formal Semantics of Natural Language, pp. 99–111 (1975)

[3] I am not able to apply nondetachability and indeterminability for reasons of space.

5. Felka, K.: On the presuppositions of number sentences. Synthese **192**(5), 1393–1412 (2015)
6. von Fintel, K.: Would you believe it? The King of France is back! Presuppositions and truth-value intuitions. In: Reimer, M., Bezuidenhout, A. (eds.) Descriptions and Beyond, pp. 315–341. Oxford University Press (2004)
7. Grice, P.: Logic and conversation (1975). Reprinted in Grice, P.: Studies in the Way of Words, pp. 22–40. Harvard University Press (1989)
8. Haslanger, S.: The normal, the natural and the good: generics and ideology. Polit. Soc. **3**, 365–392 (2014)
9. Lawler, J.: Studies in English Generics, vol. 1, no. 1. University of Michigan Papers in Linguistics (1973)
10. Levinson, S.C.: Presumptive Meanings. The Theory of Generalized Conversational Implicature. The MIT Press, Cambridge (2000)
11. Shanon, B.: On the two kinds of presuppositions in natural language. Found. Lang. **14**, 247–249 (1976)

Modelling Context Awareness for a Situated Semantic Agent

Piek Vossen(✉), Lenka Bajčetić, Selene Baez, Suzana Bašić, and Bram Kraaijeveld

Vrije Universiteit, Amsterdam, The Netherlands
piek.vossen@vu.nl

Abstract. This paper presents a model of contextual awareness implemented for a social communicative robot *Leolani*. Our model starts from the assumption that robots and humans need to establish a common ground about the world they share. This is not trivial as robots make many errors and start with little knowledge. As such, the context in which communication takes place can both help and complicate the interaction: if the context is interpreted correctly it helps in disambiguating the signals, but if it is interpreted wrongly it may distort interpretation. We defined the surrounding world as a spatial context, the communication as a discourse context and the interaction as a social context, which are all three interconnected and have an impact on each other. We model the result of the interpretations as symbolic knowledge (RDF) in a triple store to reason over the result, detect conflicts, uncertainty and gaps. We explain how our model tries to combine the contexts and the signal interpretation and we mention future directions of research to improve this complex process.

Keywords: Robotics · Situated context · Social context · Discourse context

1 Introduction

Without context, we are lost in semantic space. Ambiguity and variation of natural language is so big that meaning is unsolvable without context. Contexts can be defined as knowledge-rich data points that can help in interpreting signals, while signals are structures that convey information. Contexts have predictive power, as they can predict the signal before it is present or when it is masked. Strong evidence for this predictive power comes from current word embedding models trained from Big Data such as Word2Vec [15] or GloVe [16] which predict the direct linguistic context. Embedding models can also be reversed to predict the linguistic context from the signal, showing that the relation is mutual since contexts need to be interpreted as well: contexts define signals and signals define contexts. The difference is more a matter of relevance and focus.

In real-world situations the context can be complex or confusing when interpreted wrongly. This is apparent when modeling (mutual) understanding in

G. Bella and P. Bouquet (Eds.): CONTEXT 2019, LNAI 11939, pp. 238–252, 2019.
https://doi.org/10.1007/978-3-030-34974-5_20

human-robot communication. Robots have limited capacity to perceive and interpret the context in its full complexity. This severely complicates communication. Our robot model therefore aims to take uncertainty and alternative interpretations as a starting point, while using communication to adapt, correct, confirm and reach consensus about interpretation. Possibly wrong interpretations of signals are stored in a 'brain' (an RDF[1] triple store) as symbolic knowledge representations using a Theory-of-Mind model [12] that keeps track of the sources of interpretation and its status. This architecture allows our robot to reason over contextualized signal interpretations and to proactively resolve errors, conflicts, uncertainties and gaps using natural language dialogue.

In this paper, we report on our vision to model contexts of human language communication in real-world situations by building a robot model that communicates about the world and about us. In our previous work [18,19] we introduced the social robot *Leolani* or *L* as a multi-modal semantic agent that uses communication to learn. In this paper, we explain her contextual awareness along three different data layers: the discourse, the surrounding space and the social relationship. We describe how the interpretation of signals and contexts influence each other. In Sect. 2, we explain the foundations for our robot model, while in Sect. 3 we describe the overall robot model. In Sects. 4.1, 4.2 and 4.3, we describe the discourse, spatial and social contexts, respectively, and the way in which our model deals with ambiguity and conflicts in interpreting signals and acquired knowledge. We conclude and look forward in Sect. 6.

2 Background

Our research focuses on *identity*, *reference* and *perspective*. Identity conveys the ability to determine the world we share and the objects, situations and properties within it. Robots perceive the world differently from us and have difficulty identifying situations, entities and properties of the physical world. Identity of the world can therefore be very different across humans and robots. Language commonly makes reference to the physical world and the entities within it. We can make reference to the same things in different ways and different things can be referred to in the same way [6]. We can observe large ambiguity and variation in making reference to the world. While ambiguity can be resolved by context, variation can only be explained by both the context and the perspective of the source that makes reference (a personal context). Perspective can be defined as the personal and social position of the discourse participant with respect to the topic of communication. This position can be defined spatially, where the discourse participants stand with respect to perceived objects, and also socially: what knowledge is shared, what emotions and intentions you have, what the relationship is between the discourse participants.

Identity, reference and perspective are clearly related and ambiguity and variation can only be resolved and explained when dealing with all three aspects in combination. This relation is contextual in nature. What we distinguish and

[1] Resource Description Framework.

consider relevant in the real world is partly determined by our perspective, e.g. intentions and past experiences. The distinguished and relevant world determines references and resolves ambiguity. Such contexts are also defined and established during conversations, which creates a referential context that can be exploited.

In the next section, we will investigate the context created in discourse, the spatial context as the result of awareness, and the way the social context can be established and exploited during conversation between robots and people. We follow a pragmatic modeling approach to study the relation between contexts and the above phenomena, where we also investigate the capacity of the robot to collaborate in reaching an optimal solution. Our robot implementation thus demonstrates context in its full complexity and shows directions of future research to explore. In the next Sect. 3, we first explain the basics of our robot model L .

3 The Robot Model

L is a curious robot equipped with cognitive abilities and communication skills to support social behaviour. When switched on, L scans the objects and people in her environment and relates them to a new instantiated context. Next, she tries to determine her location either by reasoning over previous contexts or asking an available source. Upon encountering people, L tries to discern whether the person is already known and should be greeted as such, or the person is encountered for the first time, in which case a get-to-know sequence is initialised. Subsequently, the robot waits for the person to initiate conversation by asking a question or making a statement. Questions trigger simple (SPARQL) queries, while statements are processed to represent new knowledge along with its provenance. When new information is added, this generates *thoughts*, which are reflections of the current state of the "brain" (the storage of her knowledge) and how this is affected by the newly added information. Through these thoughts the robot raises pro-active questions or comments to the person to improve the state-of-the-brain. These initiatives to improve the state-of-the-brain are defined as inner *drives*, some of which try to harmonize knowledge in relation to the context.

The overall robot model architecture is shown in Fig. 1. We defined four layers: (1) sensor processing layer, (2) communication layer that responds to sensor input or inner drives, (3) language processing layer which deals with questions and statements, and (4) knowledge layer that queries or stores the result of communication, or accesses the Web. We utilize several ready-made modules in the sensor processing layer: WebRTC [2] for speech detection, the Inception neural network [17] for object recognition, OpenFace [1] for face recognition, and Google Cloud Speech-to-Text API [7] for speech recognition. The outputs of these processing modules are used as inputs to the other layers. Hence, we do not address potential conflicts and ambiguities in the signal layer itself, but try to resolve them in the higher-level layers.

Signals are processed either as perceptions of the surroundings or as communication. Visual perceptions are interpreted by object recognition and face

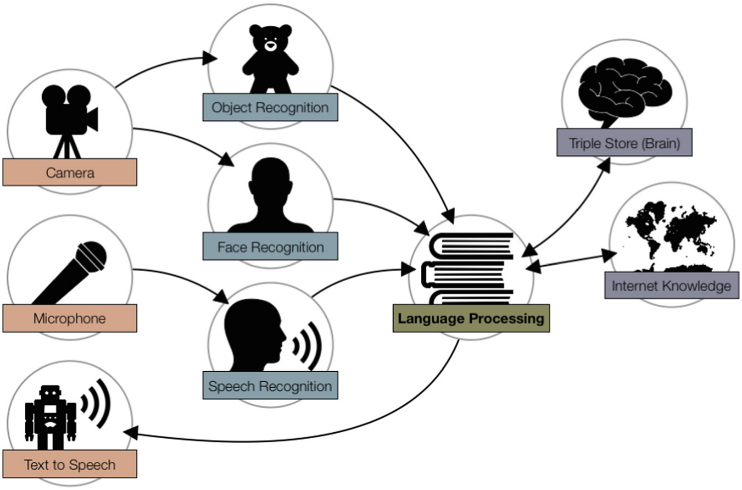

Fig. 1. Overall architecture of the robot model

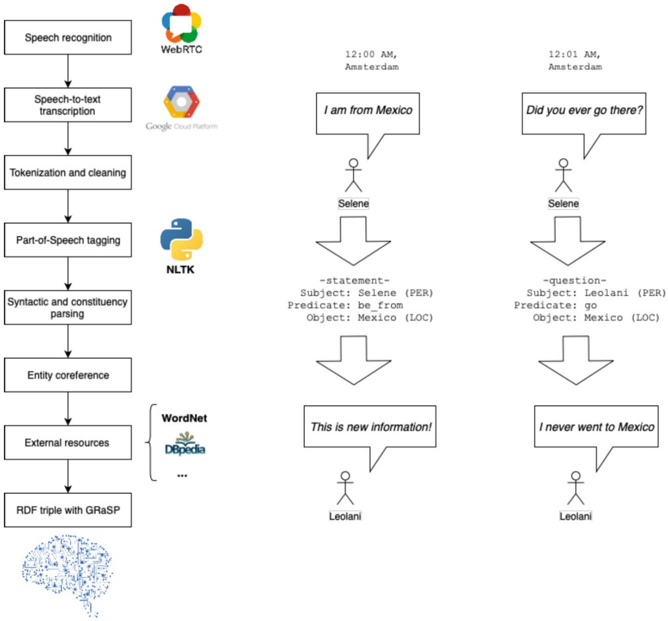

Fig. 2. Natural language processing pipeline

recognition. These are used to define the surrounding context and to identify the people in it. Audio perceptions are processed as language. The result of interaction is stored in an RDF triple store ("the brain"), which stores all interpretations of experiences. The brain forms the basis for the drives of the robot to communicate.

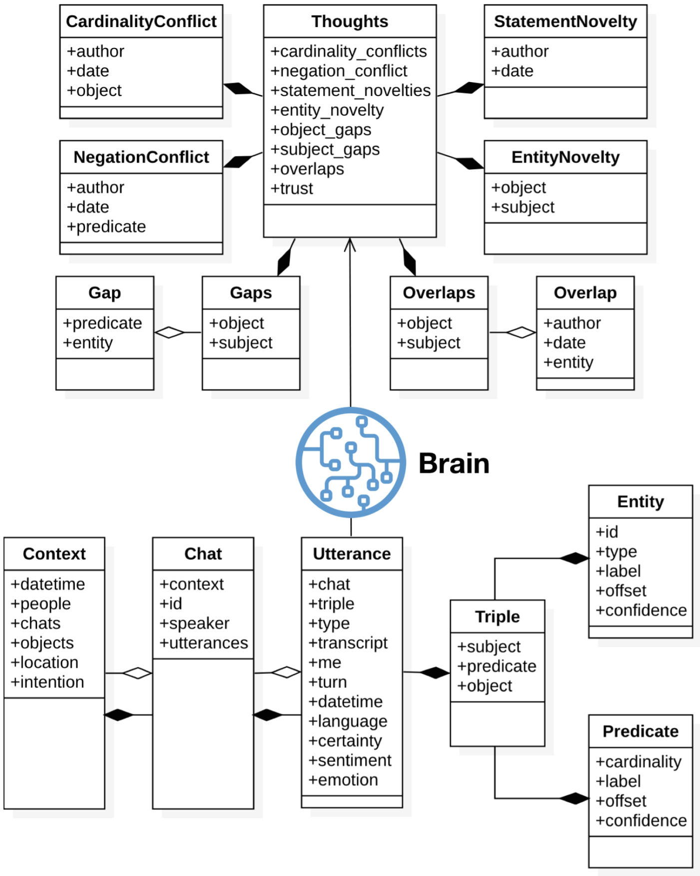

Fig. 3. Data model class diagram

Figure 2, shows the NLP pipeline that is used to process the audio signal. As shown in Fig. 2, the NLP Pipeline consists of several external components, while some are manually implemented specifically for this task. For the sake of transparency, reasoning and control, we resorted to rule-based parsing instead of a neural-net approach. This refers specifically to the syntactic and constituency parsing.

We use the Grounded Representation and Source Perspective (GRaSP) model [5] as a basis for representing content, communication and sources. We have adapted GRaSP to deal with perception and communication by robots. Statements communicated to the robot are mapped to RDF representations, which are stored together with their source. The model also stores the perspective of the source on a property expressed in the statement. The possible perspective values are denial/confirmation, sentiment/emotion, and certainty. One of the purposes of modeling the perspectives is to respond to conflicting or uncertain interpretations of signals by the model: negative feedback, correct or provide additional information.

In Table 1, we show a simplified example of an RDF representation in the brain which is the result of processing an utterance in a chat for which *Tom* is the speaker, within a specific context in Armando's office on the 24th of January 2019 during which she also perceived a chair and a person. *Tom* claimed that *Karla lived in Paris* and expressed a perspective: he confirms the claim and he is certain and surprised. In the meantime, while *L* was listening to *Tom*, she also saw a chair and recognized a person, *Gabriela* in the room where the chat took place, *Armando's office*. The event and the perceptions are all part of the same context, anchored in time and place. The RDF representation gives further details on the source, as well as perspective, entities and relations expressed in the utterance.

Table 1. RDF triples representing a context taking place in a specific time and place, an utterance in a chat, the speaker, the claim made and the perspective of the speaker on the claim

Named graph: lTalk:Interactions		
lContext:context1	a	eps:Context;
	sem:hasBeginTimeStamp	lContext:2019-01-24;
	sem:hasPlace	lContext:armandosOffice;
	sem:hasEvent	lTalk:chat4;
	eps:hasDetection	lWorld:gabriela, lWorld:chair1
lTalk:chat4	a	grasp:Chat;
	sem:hasSubevent	lTalk:chat4_utterance1
lTalk:chat4_utterance1	a	grasp:Utterance;
	sem:hasActor	lFriends:tom
lContext:armandosOffice	a	sem:Place
lFriends:tom	a	sem:Actor, grasp:Source
Named graph: lTalk:Perspectives		
lTalk:chat4_utterance1_char0-25	a	gaf:Mention;
	grasp:denotes	lWorld:karla_livedIn_paris;
	prov:wasDerivedFrom	lTalk:chat4_utterance1;
	prov:wasAttributedTo	lFriends:tom
lTalk:chat4_utterance1_char0-25_ATTR1	a	grasp:Attribution;
	rdf:value	grasp:CONFIRM, grasp:CERTAIN,
		grasp:SURPRISE;
	grasp:isAttributionFor	lTalk:chat4_utterance1_char0-25
Named graph: lWorld:Instances		
lWorld:karla	a	n2mu:Person, gaf:Instance
lWorld:paris	a	n2mu:Location, gaf:Instance
lWorld:gabriela	a	n2mu:Person, gaf:Instance
lWorld:chair1	a	n2mu:object, gaf:Instance
Named graph: lWorld:Claims		
lWorld:karla_livedIn_paris	a	grasp:Statement, sem:Event
Named graph: lWorld:karla_livedIn_paris		
lWorld:karla	lWorld:livedIn	lWorld:paris

Communication modeling starts with representing the `Context`, which provides information about the situation within which conversations take place. Within a `Context`, there are `Chats`, which model one-to-one conversation. Within a `Chat`, `Utterances` are spoken, both by the human and the robot. These `Utterances` are parsed, to obtain a subject-predicate-object RDF `Triple`. The parsed `Utterance` is sent to the brain (represented as in Table 1), which, in response, produces `Thoughts`. These `Thoughts` are the result of the inclusion of the new RDF triple and its reasoning in relation to all stored knowledge.

Figure 3 shows the types of thoughts that we defined so far: *gaps*, *conflicts*, *overlap* and *novelty*. `Gaps` are defined by the ontologies included, and as such relate to the structure of the modelled world. `Conflicts`, `Overlaps` and `Novelty` are defined by the stored triples and relate to the population of the modelled world. A detailed description of the thoughts is given in Table 2. Each of these thoughts represents a state of the brain that requires a communicative action from the robot to improve this state or to inform friends.

4 Context Implementation

In order to provide our semantic agent with sufficient contextual information to properly interpret the world, we split the contextual model into three complementary aspects. Firstly, *discourse* context is modeled as the awareness of what has previously been said. Discourse context is typically distinguished from the *perceptual* or *situational* context. Situational context includes the relevant aspects of the environment, which is non-textual in nature. It consists of date, time, geo-location, and objects and persons present in the space. This allows for a basic spatial awareness of the robot's surroundings.

However, context should not be thought of as a static backdrop to language. Rather, it is dynamic in nature as context not only influences the language used on particular occasions, but is also itself changed by language. Hence, we can perceive an interactive relationship between language and context.

Table 2. Types of thoughts

Cardinality conflict	Claims that cannot coexist because a strict one to one predicate is enforced, but two different objects have been linked
Negation Conflict	A new claim is directly negated by a previous claim
Statement Novelty	Awareness that knowledge was acquired before, along with the provenance, or if it represents genuinely new information
Entity Novelty	Awareness that a new entity is mentioned
Subject/Object Gap	Potential knowledge about a subject/object is absent and provides an opportunity to learn something new
Overlap	Awareness that new claim contain shared, but not equal, information already present in the brain
Trust	Score based on how much people talked, how much the robot learned from them, and how many conflicts they generate

We define this dynamic relationship as the *social* context. The social aspect of this dynamic relationship can be partially explained with Gricean maxims [8]. Essentially, talking to someone means we have a better understanding of their world knowledge as we are creating our mutual social context. Within it, we can describe concepts in a way specific to our individual and mutual world knowledge, differently than we would with someone previously unknown. This personal language is evolving and allowing for shorter references and thus more efficient communication.

The form of referring expression used by a speaker signals their belief with respect to the status of the referent within the hearer's set of beliefs. For example, a pronominal reference signals that the intended referent has a high degree of salience within the hearer's current mental model of the discourse context. [9] Thus, references made in dialogue can be interpreted against the current linguistic and situational context, as well as the growing world knowledge which L shares with the speaker.

In the next subsections, we illustrate the role of context for interpretation and communication within our robot model for the discourse, spatial and social context. The technical reality of this interaction is both challenging and confronting with respect to the questions and problems to be solved for robot-people interaction, setting a research agenda for the future.

4.1 Discourse Context

In discourse models, linguistic context is perceived as a part of general, situational context. This means that within one *situation* there can be many dialogues with different people. Due to the lack of mobility while being in active mode and difficulties of conducting dialogue with more than one person, this world view is well suited for our agent.

Discourse context is stored in a hierarchy of objects which connect the information about the speaker with the current spatial and linguistic context.

Spatial and situational context awareness is modeled with a wide-scope *Context* object. Every context has a unique ID which points to a location such as *Piek's office*. The basic assumption is that people and objects that are present may change but the location and situation may stay the same. Accordingly over time, L will be able to reason over patterns of objects and people present at different contexts, and learn what to expect. A crucial aspect is the identity of the human discourse participant, which is established by face recognition and getting to know new people.

When a conversation starts with an identified person, a new *Chat* object is created which is connected to the speaker. This way, all first and second person personal pronouns can be co-referenced. Within a chat, there can be many *Utterances*, both statements and questions. Keeping track of the types of the things mentioned in recent discourse allows for easier entity and pronoun coreference. For example consider the following utterance: "My sister and I like London, but we've never been there". After syntactic parsing, pronouns are dereferenced using a lexicon and a rule-based system. Coreferring *there* to *London* can be done simply by knowing that *London* is a location and that *there* is a non-ambiguous pronoun, referring to a location. However, in order to connect the other two co-referents, a slightly higher awareness is necessary. Of course, *I* refers to first person singular, and *we* to first person plural, but *my sister and I* refers to two people, which can then be co-referred with *we*, of which *I* refers to the identified speaker and *sister* is the individual identified relative to the speaker.

All aspects of this model suffer from practical limitations, and the model of linguistic context is no different. In a pipeline system such as this one, there are a lot of possible reasons why the input could not be correctly parsed or interpreted. The input sentence could be ungrammatical or incorrectly transcribed, the Part-of-Speech tagger or the syntactic parser could fail, semantic types can be incorrectly classified, etc. In these cases, a good property of linguistic context is that it is dynamic and the agent can change it too. In other words, by implementing a multi-initiated discourse model, our agent asks for clarification and more information when needed. Confusion, conflicts and uncertainty of coreference relations will thus trigger a drive to resolve these and trigger the robot to ask questions to the human.

4.2 Spatial Context

One of the major problems for our robot is distinguishing between separate instances of objects of the same type. Whereas people are identified individually through face recognition, object recognition only yields types. In the first version of our model, only a single instance of each object type is represented in the brain and all knowledge is linked to this instance, i.e. all perceived chairs result in the same object instance of the type chair: one-type-one-instance. We thus create a single URI for a unique instance based on the type, as an instance of this type for all encounters and mentions. The alternative extreme is to treat each perception and mention of an object type as a new instance, but that over-generates instances, i.e. one-perception/mention-one-instance. In that case, we create a single URI for each perception, as different instances of the type. The proper granularity of identities is somewhere in between those two extremes but needs to be carefully crafted in accordance with the human ways of defining instances in context.

Failing to distinguish objects (and also people) results in unwanted errors and conflicts, as all claims made about any chair are stored as claims for the same chair. Failing to identify objects results in dispersed information over false identities and more ambiguity, making it impossible to decide which chair is being referenced or e.g. which laptop is my laptop. How then to define the permanence of objects and their identity, so that we approximate the true number of distinct objects per situation?

Our current solution exploits the knowledge about locations and contexts to reason over object instances. As explained in Sect. 3, situations encountered by L are represented as instances of a *context*. An instance of a context is anchored in time and connected to a location. All objects and people that she meets during a context instance are linked to this context together with the identified location.

Identifying the location and identifying the objects mutually depend on each other and this forms the basis for making reference to identities in a context.

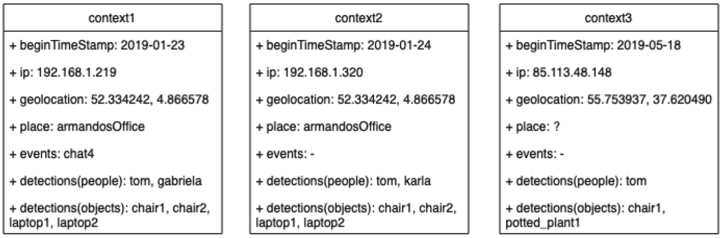

Fig. 4. Example for context construction, and location and object identity

In practice, when switched on, the robot becomes aware of a new context and creates a new instance in her brain. This is shown in Fig. 4, for *context1*, *context2* and *context3* which are created on different days during which she is switched on. Next, she scans the objects and people in her environment and relates them to this new context. People are identified through face recognition and objects are represented as potential new object instances of a certain type based on image recognition. After this first scan, the robot tries to identify her location for which she gathers some initial information (IP, geolocation). She matches all the information of the current context with all previously modeled contexts.

In Fig. 4, the information collected for context2 is compared to context1, whereas context3 will be compared to context2 and context1. Note that only so-called *endurants*, as defined in DOLCE [13], make sense to compare. Endurants, such as object and physical places, persist through time and place and therefore across contexts, whereas perdurants, such as events, conversations, and situations only exist within a time and place boundary and therefore only exist at most for the duration of each instance of a context. Given the basic information on the location derived from the system, the robot thus only uses physical objects and dimensions to compare contexts for determining the potential location.

If there is sufficient overlap with a previous context, L hypothesizes that she is now in the same location. In case of uncertainty, she can ask for confirmation. If she is certain that there is no match, she assumes she is in a new location and will ask for its name. If a new location is detected and confirmed, the robot assumes all objects in this location are new instances. If a known location is recognized, she will map the physical objects of the new context to the objects of the matched location of the most recent context. If there are fewer objects in the new context, these objects are assumed to be absent but still exist in the brain. If there are more objects in the new context, new instances are created to match the cardinality. Object identity is thus determined in relation to location identity, where the robot tries to maximize the permanence of objects for each location across different contexts. Obviously, we need to extend this model to deal with objects that can move to new locations.[2]

In Fig. 4 for example, context2 matches context1 for *Tom* and two chairs and two laptops. On the basis of the match, *L* concludes she is now in *amandosOffice* and the chairs and laptops are assumed to be the same, as there is no cardinality mismatch. What is different is the presence of *Gabriela* in context1 and the presence of *Karla* in context2. In contrast in the case of context3, only *Tom* and one *chair* are matched while the *potted_plant* is new. Therefore, the place value remains unresolved which will trigger her to ask for the location. If that is different from previous locations, both the *chair* and the *potted_plant* will be added as new instances to the brain, each with properties indicating when and where they were perceived and mentioned if referenced during a chat in the same context.

In communication, the robot treats objects in new locations as new instances unless told otherwise. For example, if somebody claims ownership of a chair within a context and location, e.g. *this is my chair*, the property *own* is assigned to that instance. In another location, a similar object can be perceived but it is considered to be a different instance. However, if the same person again claims ownership of this similar object, the robot realizes that multiple similar objects related to different locations are owned by the same person. As a weak conflict, this may trigger questions about identity: *is this the same chair?* On the other hand, if the chair in this new location is claimed to be owned by another person, it does not result in a conflict as it was already represented as a different chair in the brain and both chairs can have different owners.

Our current implementation can only detect a limited range of objects of coarse types and we have only started to detect basic object properties such as colour, size and relative position. The robot awareness of contexts, locations and the objects is therefore extremely shallow and limited compared to human representations. However, our model is open to more fine-grained representations and improvements in detecting differences. If image recognition improves, future versions can detect even more object properties than colour or specific positions in a location using 3-D triangulation. Likewise, we expect that e.g. ownership of very similar objects can be hypothesized from closeness to the owner: my chair, phone and laptop are close to me, yours are close to you.

4.3 Social Context

Social context is on the one hand defined by the drives for social interaction and on the other hand by the shared personal experience built up from previous encounters. The modeling of knowledge through GRaSP enables the robot to consider all the knowledge, experiences and communication that is the result of encounters with a single specific person. This shared social knowledge represents a personalized context, which forms the basis for more efficient communication.

[2] In the future, we plan to use properties of objects (both perceived and communicated) to help to further separate different instances, e.g. *green chair* or *my chair is close by me*.

1. less ambiguity: both lexical and referential ambiguity is limited as the shared vocabulary and shared world take priority over the possible world in interpretation
2. less variation: words used previously to make reference are preferred over alternative variants
3. more relevance: situations, objects, people, concepts previously discussed are more relevant than others and new information in relation to the known is also more relevant

Following the Gricean Maxims, the shared knowledge and experience defines information that does not need to be exchanged and it also provides that background against which all new signals are to be interpreted. The first time I talk about *my sister* the robot can only identify her relative to me but except for that she has no idea about the identity. As a result of my statement a URI will be created to represent her in the brain as an instance of a person (gender female) and the kinship relation is created to link her to me and my parents. As L does not know her name, she will ask for it, which is *Selene*. From now, we share this knowledge and I can refer to her as *my sister* or as *Selene* as shared knowledge.

On the other hand, *Selene* is also the name of another friend of L . By introducing my sister, an ambiguity is created. Of course, there are many Selenes in the world, but in my case the ambiguity only exists if I also know the other Selene. References to *Selene* by me are not ambiguous as long as L derives from her brain that we only talked about one Selene. On the other hand, she may be prompted to ask if I happen to know the other Selene as well. In that case, two Selenes become part of our shared knowledge and from that point on mentions need to be disambiguated, i.e. *Selene, your sister* and *Selene, my friend*.

The drives that the robot has to pro-actively interact can be tuned to such personalized contexts. Drives such as Novelty, Conflicts, Uncertainty, Trust are mainly considered in relation to you. In case of *Selene*, we already saw that Novelty of information that results from our conversation triggered L to inform me about related knowledge she has. Similarly, she may ask if I know people she knows that have any other background that is similar. There can be conflicts and uncertainty coming from the communication with anybody but she will only address me about conflicts and uncertainty that related to what we talked about. Trust is a judgment based on the information I communicated in the past and therefore is personal by definition but also relates to others because it reflects the number of conflicts I am involved in. Finally, even perception is based on our social relationship. She may constantly perceive objects but she will give priority to objects we talked about before or the ones that are close to me or that I own.

5 Related Work

Mavridis [14] gives an overview of natural language processing technologies in human-robot interaction and challenges to be tackled, including 'theory of mind', open-domain communication, varied speech acts, symbol grounding and multiple-turn dialogues. Most human-robot communication models still only handle basic communication using one or two speech acts, limited symbol grounding and single turns. In fact, none of these systems exploits context.

Embodied dialogue agents nevertheless offer many new challenges and possibilities to exploit the multimodal nature of situated dialogue. When the embodied agent has a conversation with a partner, they interact within a situational context. This encompasses the physical world around them, their positions within it, as well as the moment in time and other pragmatic notions that arise from the situation. This affects the referring expressions used within dialogue, and fluent use of these expressions is affected by the mutual knowledge that the conversational partners share [4]. The choice of referring expressions used is affected by their salience, whether in discourse or within the situational context. In most computational models, the choice of possible referents can be found within the Discourse Context, which is accumulated through conversation [3]. Discourse context is commonly differentiated from the mutual knowledge set, which contains information available to both conversational partners that was not referred to in the discourse. A referring expression in an utterance introduces a representation into the semantics of that utterance and this representation must be bound to an entity in the mutual knowledge set (in the case of evoking or exophoric references) or in the discourse context (in the case of anaphoric references) for the utterance to be resolved [10].

To interpret a referring expression, algorithms typically analyze the recently mentioned entities. However, this is not enough for embodied agents which are becoming more and more prominent in a variety of domains. One kind of embodied agents are robots with an integrated spoken interface. Implementations of such agents commonly focus on command-and-control interfaces, rather than placing the user and conversational agent into a shared space which can be talked about in an open dialogue [3].

A common approach used to systematically represent situations for the purpose of modeling situational context are ontologies. The design of ontologies for this purpose needs to comply with semantic requirements regarding the capabilities of representing contexts and situations in a general way. For instance, the Situational Context Ontology combines contextual information (spatial and temporal) with related situations of an individual [11].

Our model is designed for open communication with the explicit result of acquiring knowledge and building a social relationship. Ontological knowledge is used to control the interaction and to interpret the context, e.g. people, friends, locations, space, objects detected by the image recognition, some basic object properties. This basic ontology allows us to model the context of the interaction and the references to these contexts in the communication. Although the ontological model is relatively basic, it allows us to model and experiment the interaction in real-world situations. We defined a context as an episodic element that explicitly gathers everything *Leolani* learns in connection with specific situations. In addition, we defined *thoughts* as reflections on the interpretations of the contexts and any previous episodic encounter, i.e. *awareness* of gaps, relevance, conflicts, uncertainties, trust. Reflections result in drives to interact with the participants or observe the situation, which results in an update of the context interpretation.

6 Conclusion

In this paper, we described a robot model for social communication within contexts that was implemented and can be used for further experiments. We did not take a theoretical perspective but tried to consider all practical aspects from a pragmatic perspective dealing with the full complexity of a real-world physical context. By considering the problems within realistic situations, we present a vision for future research in essential but also down-to-earth aspects of interpreting contexts.

We explained three notions of context: spatial, discourse and social, that interact with each other and with the interpretation of signals. We demonstrated that also the context needs to be interpreted as a collection of signals and that contexts and signals define each other. We showed how our models try to exploit this relationship and what the limitations are. The level of mutual understanding of the context and the signals within is still very limited and our robot still has the capacity of less than a child. Partially, these limitations can be resolved by better image recognition (objects, properties and relations), detection of scenarios, more knowledge acquired over longer periods of time, richer language models, and more. The pioneering work described in this paper, sets an agenda to further experiment with the different aspects of context and interpretation in real-world physical environments and to evaluate different model implementations.

Acknowledgement. This project was funded through the NWO-Spinoza funds awarded to Piek Vossen and by the VU University of Amsterdam. We specifically thanks Selene Kolman and Bob van Graft for their support.

References

1. Amos, B., Ludwiczuk, B., Satyanarayanan, M.: Openface: a general-purpose face recognition library with mobile applications. Technical report, CMU-CS-16-118, CMU School of Computer Science (2016)
2. Project authors, T.W.: Webrtc. In: Online publication (2011). https://webrtc.org/
3. Byron, D.: Understanding referring expressions in situated language: some challenges for real-world agents. In: Proceedings of the First International Workshop on Language Understanding and Agents for the Real World, pp. 80–87. Citeseer (2003)
4. Clark, H.H., Marshall, C.R.: Definite reference and mutual knowledge. In: Psycholinguistics: Critical Concepts in Psychology, vol. 414 (2002)
5. Fokkens, A., Vossen, P., Rospocher, M., Hoekstra, R., van Hage, W.: Grasp: grounded representation and source perspective. In: Proceedings of KnowRSH, RANLP-2017 workshop, Varna, Bulgaria (2017)
6. Frege, G.: Über sinn und bedeutung. Reclam, Philipp (2019)

7. Google: Cloud speech-to-text - speech recognition. In: Online publication (2018). https://cloud.google.com/speech-to-text/
8. Grice, H.P.: Logic and conversation. 1975, pp. 41–58 (1975)
9. Kelleher, J.D., Dobnik, S.: Referring to the recently seen: reference and perceptual memory in situated dialog (2019)
10. Kelleher, J.D., Dobnik, S.: Referring to the recently seen: reference and perceptual memory in situated dialog. arXiv preprint arXiv:1903.09866 (2019)
11. Kolbe, N., Zaslavsky, A., Kubler, S., Robert, J., Le Traon, Y.: Enriching a situation awareness framework for IoT with knowledge base and reasoning components. In: Brézillon, P., Turner, R., Penco, C. (eds.) CONTEXT 2017. LNCS (LNAI), vol. 10257, pp. 41–54. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-57837-8_4
12. Leslie, A.: Pretense and representation: The origins of "theory of mind". Psychol. Rev. **94**(4), 412 (1987)
13. Masolo, C., Borgo, S., Gangemi, A., Guarino, N., Oltramari, A.: Wonderweb deliverable d17. Comput. Sci. Prepr. Arch. **2002**(11), 74–110 (2002)
14. Mavridis, N.: A review of verbal and non-verbal human-robot interactive communication. Robot. Auton. Syst. **63**, 22–35 (2015)
15. Mikolov, T., Sutskever, I., Chen, K., Corrado, G.S., Dean, J.: Distributed representations of words and phrases and their compositionality. In: Advances in Neural Information Processing Systems, pp. 3111–3119 (2013)
16. Pennington, J., Socher, R., Manning, C.: Glove: global vectors for word representation. In: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), pp. 1532–1543 (2014)
17. Szegedy, C., et al.: Going deeper with convolutions. In: Computer Vision and Pattern Recognition (CVPR) (2015). http://arxiv.org/abs/1409.4842
18. Vossen, P., Baez, S., Bajčetić, L., Bašić, S., Kraaijeveld, B.: A communicative robot to learn about us and the world. In: Proceedings of Dialogue-2019, Moscow (2019)
19. Vossen, P., Baez, S., Bajčetić, L., Kraaijeveld, B.: Leolani: a reference machine with a theory of mind for social communication. In: Sojka, P., Horák, A., Kopeček, I., Pala, K. (eds.) TSD 2018. LNCS (LNAI), vol. 11107, pp. 15–25. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-00794-2_2

Author Index