Library of Ethics and Applied Philosophy 32

Markus Christen
Carel van Schaik
Johannes Fischer
Markus Huppenbauer
Carmen Tanner *Editors*

Empirically Informed Ethics: Morality between Facts and Norms

Springer

Empirically Informed Ethics: Morality between Facts and Norms

LIBRARY OF ETHICS AND APPLIED PHILOSOPHY

VOLUME 32

For further volumes:
http://www.springer.com/series/6230

Markus Christen • Carel van Schaik
Johannes Fischer • Markus Huppenbauer
Carmen Tanner
Editors

Empirically Informed Ethics: Morality between Facts and Norms

Springer

Editors
Markus Christen
Institute of Biomedical Ethics
University of Zurich
Zurich, Switzerland

Johannes Fischer
Institute of Social Ethics
University of Zurich
Zurich, Switzerland

Carmen Tanner
Department of Banking and Finance
Center for Responsibility in Finance
University of Zurich
Zurich, Switzerland

Carel van Schaik
Anthropological Institute and Museum
University of Zurich
Zurich, Switzerland

Markus Huppenbauer
University Priority Program Ethics
University of Zurich
Zurich, Switzerland

ISSN 1387-6678
ISBN 978-3-319-01368-8 ISBN 978-3-319-01369-5 (eBook)
DOI 10.1007/978-3-319-01369-5
Springer Cham Heidelberg New York Dordrecht London

Library of Congress Control Number: 2013949768

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Contents

Introduction – Bridging the Is-Ought-Dichotomy

Humans are moral beings, involved in a complex web of social interactions, acting upon biological dispositions and entangled with culture and history. Although the orientation toward the morally good is generally seen as the hallmark of humanity, moral conflicts and ethical dilemmas seem to be inevitable, often painful aspects of our moral lives. The various traditions of ethical thinking, understood as the systematic reflection upon morality, have always tried to disentangle, clarify and maybe even solve the "moral mess" people often experience. The role of facts in this endeavor—not only information about the problems with which we deal, but also about our capacities to deal with them in a moral way—has again and again been disputed within ethics. How sharp is the distinction between the world of facts and the world of norms?

Recently, interest in using empirical approaches to understand (human) morality has surged across various scientific disciplines: Psychologists investigate how emotions and intuitions influence our ethical theorizing; behavioral economists analyze the impact of moral affect on rational decision making; neuroscientists portray the "moral brain"; anthropologists reconstruct the deep history of moral traits; primatologists look for the "building blocks" of human morality in our primate relatives. Ethicists react in various ways toward these developments. Some deny a substantial relevance of empirical facts for normative argumentation, others call for "empirical ethics" and a few even start doing simple experiments in order to understand genuine moral intuitions. Furthermore, from a science-studies and cultural-history point of view, the recent flood of publications on morality invites interpretations on boundary struggles between disciplines (who has which role in disputes on normative issues?) and on the current social climate within our society.

This volume intends to provide an overview of the most recent developments in empirical investigations of morality and tries to assess their impact and importance for ethical thinking. It involves contributions of scholars from philosophy, theology and empirical sciences with firm standings in their own disciplines but also with inclinations to step across borders—in particular the one between the world of facts and the world of norms. Human morality is complex, and probably even messy—and any distinction between facts and norms becomes blurred when looking closely

at the various components that potentially enable and influence our moral actions and ethical orientations. In that way, morality may indeed be located *between* facts and norms. For that reason, an empirically informed ethics less concerned with analytical purity but thoroughly immersed in moral complexity may be an important step toward making the contributions of ethics more valuable and relevant. With this in mind, we hope that this volume introduces the reader into a zone of scientific inquiry, where fruitful new topics emerge at the boundary between the kingdoms of facts and norms.

This book emerged from an international workshop held in Zurich in March 2010. At this gathering, distinguished scholars and young researchers both from moral philosophy and empirical sciences discussed the various implications of empirical research for ethical theorizing. The editors thank the Swiss National Science Foundation and the University Research Priority Program Ethics of the University of Zurich for supporting this workshop and the book that resulted from these discussions. In particular, we thank Kevin Ladd from the Indiana University of South Bend for critically commenting and proofreading the manuscript, an anonymous reviewer for providing his very helpful input, and Christopher Wilby from Springer Science + Business Media for his support in publishing this book.

The Editors

Part I
What Is Empirically Informed Ethics?

Chapter 1
Outlining the Field – A Research Program for Empirically Informed Ethics

Markus Christen and Mark Alfano

1.1 Introduction

"What is the right thing to do?" This question[1] echoes through the centuries and millennia of human history. It alludes to the sometimes disturbing moral dilemmas humans face, and it has produced elaborate ethical theories of the virtues people should foster, the norms societies should promote, and the states of affairs at which people should aim. It is therefore unsurprising that human behavior in moral contexts has become a topic of empirical research, although it was to some extent deliberately excluded as a legitimate research topic at the advent of modern science.[2] The last two decades have witnessed a substantial increase in empirical research on morality—in particular using psychological and neuroscientific methods.[3] This research also influences moral philosophy; in fact, empirical research on morality

[1] Allowedly, the human concern with morality is not represented by a single question, and the focus on moral decision-making and moral action, for which this question stands, is most typical of a recent understanding of ethics as a "toolbox" for helping to solve problems and setting aside questions like "Who should I become?" that refer to virtues and moral ideals; see Pincoffs (1986) and Williams (1985) for critiques of this tendency to narrow the focus of ethics.

[2] A well-known piece of evidence for this point is the draft of the credo of the Royal Society written by Robert Hooke in 1663, in which he articulated the role of the Society as "to improve the knowledge of natural things, and all useful Arts, Manufactures, Mechanic practices, Engines and Inventions by Experiments, not meddling with Divinity, Metaphysics, Morals, Politics, Grammar, Rhetoric or Logic." Although this sentence did not enter the final charter of the Royal Society explicitly, its fragments can be traced to various parts of the charter (quoted after Weingart 2002: 96).

[3] For bibliometric evidence for this claim, see Christen (2010).

M. Christen (✉)
Institute of Biomedical Ethics, University of Zurich,
Pestalozzistrasse 24, 8032 Zurich, Switzerland
e-mail: christen@ethik.uzh.ch

M. Alfano
Center of Human Values, Princeton University, Princeton, NJ, USA

M. Christen et al. (eds.), *Empirically Informed Ethics: Morality between Facts and Norms*, Library of Ethics and Applied Philosophy 32, DOI 10.1007/978-3-319-01369-5_1,

has been the biggest beneficiary of citation transfers into the humanities, compared with other research topics of social neuroscience (Matusall et al. 2011).

Moral philosophers' responses to this trove of empirical data on the evolutionary origin, the biological foundation, the psychological malleability, and cultural diversity of human morality have been ambivalent. One strand of argument—Kauppinen (Chap. 16 in this volume) calls this strand *Armchair Traditionalism*—denies the relevance of empirical data to normative justification, with the obvious exception that it frames the specific problem under investigation (e.g. Nida-Rümelin 2006). Another strand of argumentation—labeled *Ethical Empiricism* by Kauppinen (Chap. 16, this volume)—acknowledges empirical insights for theory building within ethics (Edel 1961), but with conflicting conclusions. For example, research on the psychological foundation of moral intuitions can either be taken as a support for founding normative theories (Nichols 2004) or be used to undermine the normative importance of intuitions (Singer 2005). With respect to the application of ethical theorizing to practical problems, some scholars promote "empirical ethics" that should, in particular, improve the context-sensitivity of ethics (Musschenga 2005). And finally, some philosophically trained researchers have started using empirical methods themselves in order to inform their normative thinking (for an overview see Appiah 2008; Knobe and Nichols 2008; Loeb and Alfano forthcoming).

Of course, the role and relevance of empirical data for ethics depends on the specifics of the problems one wants to solve. Empirical knowledge will affect metaethical theories differently from, for instance, biomedical ethics or business ethics. This divergence in relevance does not necessarily indicate a fundamental conflict within moral philosophy with respect to the role of empirical data. However, there are diverging opinions about what it actually means for ethics to be *informed* by empirical knowledge—and one could even ask to what extent analytically sharp distinctions are blurred by the inclusion of empirical data in normative thinking (see Sect. 1.2.4).

Thus, the endeavor of promoting an *empirically informed* ethics raises various questions. This chapter structures them with respect to the subject-matter, the kinds of empirical methodologies and data that could be useful for ethics, and the types of problems and fundamental questions of ethics for which an empirical approach could be particularly fruitful. It also outlines what is at stake when empirical insights are taken seriously by normative theorists—a point that may affect a competence philosophy attributes to itself: the clarification of concepts and the demarcation of sharp distinctions between them. Morality could indeed be a field where this goal is more difficult to achieve than in other fields—and the facile drawing of distinctions may even mask interesting questions.

Take as an example the basic terms 'morality' and 'ethics'. In particular in the German tradition, these terms are understood to have distinct referents. The former denotes the various norms, practices, virtues, and so on that a specific society or culture holds over a given period of time; the latter is the systematic investigation and justification of these practices, for which the moral philosopher is particularly qualified (e.g. Düwell et al. 2002; Nida-Rümelin 2006). But a closer look at the

practice of morality immediately shows that justifications and reflections are a genuine part of common morality, too—although they are sometimes misleading, doubtful, affected by disruptive factors, and even mistaken. Everyday moral justification lies on a continuum with sophisticated philosophical theorizing about morality (a point that Düwell et al. 2002: 3, acknowledge), which may be a reason for the (frequent) synonymous use of the terms 'moral' and 'ethical' in Anglophone philosophy. Between the covers of this book, we (and the other authors) will try to maintain a robust distinction between these terms, where 'morality' refers more to common practices and discourse upon moral issues within a specific societal or cultural frame and 'ethics' denotes a more reflective approach that is usually connected with some degree of expertise and knowledge in moral philosophy. The distinction may be somewhat artificial, but it remains useful.

Furthermore, this chapter serves as an introduction to the other contributions in this book, as it arranges them into a general framework of empirically informed ethics, which can be called a "research program". We do not understand this term in its sophisticated version used in philosophy of science (Lakatos 1977). Rather, it denotes the endeavor to outline the field, its topics and problems, its methods, and some of the questions we consider most interesting. Section 1.2 presents the phenomenon that empirically informed ethics tackles, which is, we propose, a thorough explanation of 'moral agency' in all its facets. In this section, we also discuss how the understanding of ethics itself influences the role of empirical knowledge for ethics—an aspect that three contributions of this book also examine to some degree (Fischer, Naves de Brito, Krones). In Sect. 1.3, we draw some important methodological distinctions, in order to help clarify the kinds of empirical research that may be relevant to ethics. It's important to distinguish, for instance, quantitative from qualitative research methods. It's also important to keep in mind that explicit, implicit, and behavioral measures of the same phenomenon may diverge. For instance, the subjects of empirical inquiry might explicitly think of themselves as honest, yet exhibit little honesty when their self-concepts are measured implicitly; and both explicit and implicit self-concept may diverge from their actual behavior in honesty-relevant circumstances. In Sect. 1.4, we provide an overview of the different kinds of data that can inform ethics in various ways. The other 11 contributions of this book will be introduced in this section. Finally, in Sect. 1.5, we present several problems that we consider particularly important for an empirical approach to ethics.

1.2 The Phenomenon Under Investigation

1.2.1 Distinguishing 'Moral' from 'Non-moral'

One basic fact about morality is that people are disposed to react to issues of right and wrong, good and bad, virtuous and vicious. This implies both the existence of

some normative frame in which the normative terms obtain their moral meaning and a connection between this normative frame and the real world, in the sense that it guides[4] thought, feeling, deliberation, and behavior[5] of most people much of the time. The connection is bidirectional: our thoughts, feelings, and behavior also influence the normative frame, often in an indirect, though sometimes also in a direct, way—for instance, by expanding it or by changing the semantics of some terms. Various spatial and temporal scales are involved in this interaction and open up a constellation of difficult, interrelated questions. The goal of this section is to structure them in a way that allows a not-too-Procrustean categorization of the various contributions in this book.

Another genuine aspect of morality, which should be mentioned right at the beginning, is its social nature: morality is situated in a social world[6] of actions, judgments, negotiations, and other kinds of expressions made by interacting social beings. This is also the reason why morality matters so much to most people: people get upset when others don't meet their moral standards. This may concern obvious transgressions like harming innocents, but also more controversial issues, for instance, with respect to politics that some might consider outside the realm of morality (Haidt and Graham 2007).

This straightforward observation leads to a difficult question: whether there are uncontroversial criteria that can be used to classify a specific judgment, action, or other phenomenon as clearly *moral*. Various classifiers emerging from different disciplines have been proposed, and all of them have their opponents: Moral philosophers may require universalizability as a property of (justified) moral judgments (a prominent example is Kant 1785/1983), and are then confronted with the objection of moral relativism (for an overview see Moser and Carson 2000). Moral psychologists may focus on the degree of acceptance of norms in order to distinguish between moral and conventional norms (Turiel 1983), but there are important counterarguments with respect to this distinction (e.g. Nichols 2002). Cognitive neuroscientists may use the (measurable) strength of the emotional reaction towards norm-transgressions as markers of morality (Moll et al. 2008b), but are then confronted with the large variability of individual emotional excitability or "affective

[4] In using the term 'guide' we do not mean that the agent necessarily requires conscious awareness of this frame.

[5] We use the term 'behaviors' to denote any observable expression of interacting social entities that includes communicative expressions of a verbal or nonverbal kind ,as well as generating records of behavior using any kind of media (e.g. exposing moral opinions through newspaper articles, blogs etc.) as long as the behavior has the potential to generate social impact. Actions are much more constrained behaviors (including intentionality, free will etc., depending on the theory of action someone holds; Mele 1997). For many philosophers, only actions are object of ethical considerations because the issue of responsibility attribution is clearer in that case. Clearly, our construal of the scope of ethics is more liberal (see Sect. 1.2.4).

[6] We use the term 'social' in a broad way including the possibility that nonhuman creatures can be understood as social beings (an undisputed claim within biology). Surely, the precise definition of 'social' will be adapted to the species under investigation, leading to the question, what kind of behaviors must be present such that the interaction of non-human creatures can be assessed from the perspective of moral agency (see the contribution of Sarah Brosnan (Chap. 5) in this volume).

styles" (Davidson 2004). Evolutionary biologists may focus on the fitness reduction some behaviors have for individuals in order to call them 'moral' (Trivers 1985), but then are accused of unjustified reductionism because they treat morality and altruism as equivalent (Joyce 2006). Carolyn Parkinson et al. (2011) have gone so far as to suggest that literally *nothing* unifies morality. This is not the place to go further into these longstanding issues—it is sufficient to state that we do not have an uncontroversial set of individually necessary and jointly sufficient criteria applicable to all phenomena that would allow us to classify them as being either moral or non-moral. This does not mean that we don't have exemplars of either kind, but there will be a grey zone that is larger than most people are inclined to think.[7]

This demarcation problem is complicated by a further wrinkle: the distinction between 'moral' and 'immoral'—as the term 'moral' has a positive connotation that is hard to avoid in these discussions. Thus, although the "cold, objective observer" of morality may be interested in any kind of entity that is eligible for classification as either 'good' or 'bad' relative to any system of justification, many would insist that the classification of an entity as 'moral' requires an *acceptable* justification—and is thus coupled to some standards of rationality and normativity (e.g. Schaber 2011). However, although there is a well-known asymmetry with respect to 'good/right' and 'bad/wrong' in morality in the sense that transgression of norms causes much stronger reactions than the fulfillment of moral ideals, there is considerable diversity in both space (i.e., between groups/cultures) and time (i.e., with respect to historical development) with respect to what is called 'immoral' (see the contribution of Prinz (Chap. 6) in this volume).

This short outline of the problem of finding adequate criteria for distinguishing the moral and the non-moral, on the one hand, and the moral and the immoral, on the other hand, should remind us to be tolerant in this respect, as we otherwise may overlook important aspects of morality.[8] For current purposes, and with the expectation that revision is inevitable, we tentatively define a phenomenon as moral (as opposed to non-moral) if and only if it is a mental state (e.g., thought, judgment, belief, motive, emotion, sentiment), mental process (e.g., deliberation, construal), behavior (e.g., acting, omitting, refraining), disposition (e.g., virtue, vice, sensitivity) or state of affairs such that the application of the evaluative predicates (e.g., 'good', 'bad', 'right', 'wrong') to it is warranted.

[7] This ambiguity probably results from the basic fact that normativity is woven into the most basic structures of life: all life-forms (including even plants, fungi, protozoa, and bacteria) have built-in "desired states" or "goal states" with respect to basic needs and threats, sensors to detect them, and actors to seek or avoid them. Although there is surely a consensus that most goal-seeking behaviors of life-forms are non-moral, this certitude decreases when social life forms are under investigation. And although we have good reason to couple (sophisticated) morality with language and the ability for conscious reasoning, this criterion may be of little use when the question is how morality evolved.

[8] See Haidt and Kesebir (2010) for an example (from social psychology) of how the initial classification can shape the types of research questions that are asked.

1.2.2 Moral Agency

Having established a basic frame of the problem, we will now outline the subject matter in more detail. We suggest that the key phenomenon *empirically informed* ethics is interested in is moral agency—the fact that patterns of moral behavior emerge from entities whose behavior is somehow regulated by a normative framework that includes an idea of 'good' and 'bad'. We deliberately use the term 'patterns of moral behavior' rather than 'moral action' in this context because we propose to understand this phrase in a broad sense not restricted to mere punctate actions (see also footnotes 4 and 9 and the following explanations).[9]

Empirically informed ethicists want to know *how moral agency is possible* and *how moral agency works*, which (in most cases) includes how reasons and justifications are operative in that framework. Answering these questions requires further specification depending on the concrete issue under investigation, as well as empirical data of various kinds (see Sect. 1.3). However, it would be a mistake to understand this research project as purely empirical, as if the project could be completed merely by decoding the "moral machinery" of the agent and identifying the elements of the normative reference frame (i.e., the norms, virtues, values, and so on that are involved in a particular instance of moral behavior). Although justification claims are an important aspect of moral agency, they can operate on various levels: on the level of the individual agent (e.g., when evaluating reasons for a specific option or action), on the level of direct agent interaction as a demand towards the agent (e.g., after he/she has done something that is criticized or praised by others), on the level of collective phenomena (e.g. with respect to incentives that operate on an institutional level), or on the level of the scientific inquiry of the phenomenon with respect to the question, whether and to what extent a specified behavior could be called 'moral'. Therefore, these practices of justification are not only part of moral agency, they are also entangled in a complex way with the actual understanding of the problem—an important point that we discuss in more detail in Sect. 1.2.4.

Before sketching moral agency in more detail, we propose to distinguish the terms 'moral agency' and 'moral agent' and understand the former in a broader sense. In this way we can include the possibility that there may be phenomena of moral agency, although we are not sure (or even doubtful) whether an agent[10] is

[9] One may object that a different term should be used instead, e.g. 'moral behaver' instead of 'moral agent' and the coinage 'moral behavior' instead of 'moral agency'. However, 'behavior' is too broad in that respect, as only a subclass of behaviors is relevant to moral agency, as social impact of the behavior (pattern) is required and some kind of mechanism that supplies feedback to the agent such that internal states (e.g., through reflection, if the agent has the required abilities) are changed and may lead to a change in future behavior patterns.

[10] We should also note that the concepts 'agent' and 'action' should be distinguished as well in the sense that not all moral behaviors produced by agents are actions. As indicated earlier (footnote 4), the term 'action' refers to a tightly circumscribed set of behaviors operating on a limited time scale. An 'agent' is an entity that is clearly discernible in social space and where no reasonable doubt exists with respect to the fact that the agent is the originator of the behaviors under investigations, some of which may be actions. Therefore, with respect to their specificity, we have a relation in the sense of: 'moral agency' > 'moral agent' > 'moral action' ('>' denotes: 'is more general than').

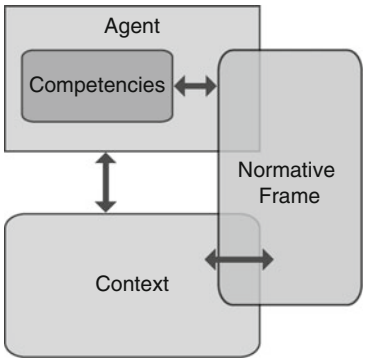

Fig. 1.1 The structural components of moral agency

actually present. This refers in particular to phenomena of collective agency that are also a topic of research in social psychology (Bandura 2001), although within philosophy doubt has been expressed whether a collective of individuals actually can be called an agent (in particular with respect to responsibility attribution; French 1998).[11] Within the research tradition of complexity science, however, we can observe an increasing amount of research on patterns in social space that have moral relevance, although they do not result from intentions or involve top-down control (a recent example is Helbing et al. 2010). These patterns are accessible to empirical research and may even be shaped through politics—i.e. they can become an object both of 'good'/'bad'-attributions as well as intervention, although there is no clearly discernible entity to which this behavior can be attributed. Therefore, some moral behaviors may be interactive or relational in nature and can be understood as an expression of moral agency.

In the following, we propose a basic structure of moral agency that enables us to categorize various research topics with respect to the spatial and temporal time scales involved in moral agency (see Fig. 1.1).

The structure of moral agency as we construe it here is threefold. First, moral agency requires a specified *set of competencies* that must be present in the agent (or a collective of agents). Second, it involves a *normative reference frame* to which the agent has at least partial access. Finally, moral agency is always situated in a *context* (consisting both of other agents and physical boundary conditions that constrain behavior). Competencies, normative frame, and context thus form the structural components of moral agency. A particular investigation of moral agency may refer to just one of these structural components, presumably by examining its content, or to the interaction of two or even all three components.

First, specifying the set of competencies is closely related to defining the moral agent—and the spectrum of proposals for necessary and/or sufficient competencies that qualify for agency is broad. In its simplest form, one may characterize an agent as an entity consisting of sensors, some internal decision procedure to generate actions, inner goal states with which the sensory information is compared, and

[11] There is a longstanding discussion about collective responsibility that we will not outline here (see e.g. Held 1970; Lewis 1948; Narveson 2002).

actors (anything that allows the agent to intervene in the world, such that the behavior of the agent is not completely controlled by factors outside of it). This simple picture—basically emerging from behaviorism and currently the standard definition within agent-based modeling (Bonabeau 2002)—is enriched in social psychology and philosophy with various further capabilities such as authorship, autonomy, intentionality, forethought, learning, self-reactiveness, and self-reflectiveness.[12] These terms refer to rich theoretical concepts, and the picture of moral agency that we end up with depends on how they are spelled out both individually and in their interrelations. At this point, we do not have to perform that task, but we recognize that such a specification is a necessary part of any investigation of moral agency.

Second, moral agency involves normativity, i.e. the idea of a 'goal state' with which an actual (or internally simulated, when assessing options) behavior can be compared, and that includes the implication that this comparison has some effect on the agent and its (future) behavior. Again, the specification of the properties and content of a normative frame is the key point when a particular aspect of moral agency is under investigation. Presumably at least some elements of the frame are accessible to the agent, although accessibility does not necessarily mean that these elements are under the conscious control of the agent when he or she is acting. It just means that these elements are represented in such a way that they can feed into the processes that generate behavior. Defining these elements in such a way that the normative frame can be called a properly *moral* frame is related to the difficulty of distinguishing moral from non-moral phenomena. Indeed, we consider the identification of a proper description of the structure and contents of the normative frame to be one of the central questions to which empirical research can contribute (see Sect. 1.4).

Third, because moral agency is always situated, we can only understand it if we have an adequate notion of the context in which it occurs. Due to the essentially social nature of morality, this context will involve other agents—either as counterparts (affected parties of the behaviors) of the agent or as observers and evaluators of the agent's behavior (e.g., in the sense of third-party-punishment). The latter could also be merely hypothetical, i.e. the agent may have the capacity to internally simulate the evaluative judgments that others would make of some proposed course of action, and regulate his or her behavior in part by reference to these simulated judgments. The context certainly also involves physical boundary conditions that constrain the possibilities for action. The degree of their coerciveness, however, is again an issue of definition. For example, it is often an open question—and one indexed to time-scale—whether a given constraint is better understood as an immutable necessity (e.g., the need for food, water, shelter, and companionship) or as a contingent obstacle (e.g., the demands of a deeply rooted political or social arrangement). In eighteenth-century America, slavery might have seemed an immovable part of the frame, and in the context of a single decision on a short time-scale, it was. Over the long term, however, the institution did turn out to be malleable. We suspect

[12] For an overview see Bandura (2001) and Christen (2009), or the entry on "action" in the Stanford Encyclopedia of Philosophy (http://plato.stanford.edu/entries/action; accessed on October 31st 2011).

that many constraints exhibit this dissociation: in the near term they are best understood as unfortunate but solid constraints, whereas over the long term they are better understood as contingent and mutable.

These three structural components of moral agency correspond to different scientific approaches to morality, each of which has a long tradition. Briefly and with apologies for incompleteness, the psychology of character and traits focuses on the dispositions people should foster in order to be called 'moral', whereas situationism in psychology emphasizes the context in order to understand morality. Finally, various traditions within moral philosophy deal primarily with the normative frame.

In summary, this outline of the structure of moral agency is soberingly complex, as it shows a need to define each single component more precisely whenever a specific problem is under investigation, by taking into account the many (often mutual) dependencies among the relevant phenomena. For example, requiring deliberate access to the elements of the normative frame has consequences for the competencies the agent should have. This observation is neither new nor surprising, but it does remind us that the concrete question under consideration requires a careful elaboration of several interrelated elements. And the outline of these questions further requires a structural clarification with respect to the spatial and temporal scales involved in moral agency, a point we discuss next.

1.2.3 Spatial and Temporal Scales

Moral agency develops on various spatial and temporal scales, which we can use to categorize moral phenomena. We use the term 'spatial scale' here to refer to the number of agents involved and the manner of their interaction. Usually, three different scales are distinguished: the single agent (who, for instance, is reasoning about a specific issue or dilemma), a group of directly interacting agents over a longer timescale (allowing, for instance, the relationships of mutual trust and dependency), and a collection of agents who interact in a more or less anonymous way (for example, by means of social institutions). In many real-world problems, these scales are entangled. However, many problems can be localized on a specific scale (e.g., the "tragedy of the commons" (Hardin 1968) on the scale of society).

With respect to the temporal scale, it's reasonable to distinguish four different levels: The time-scale of immediate acts (on the order of seconds to minutes), the scale of (deliberate) reasoning about a problem (on the order of minutes and hours to days or even weeks), the time scale of the ontogenesis of the agent (on the order of years and decades), and the evolutionary timescale that includes many generations of agents (on the order of decennia, millennia, or more).[13] Again, many real-world problems

[13] The issue of the relevant timescale of evolutionary processes surely depends on the type of phenomenon one analyzes. Furthermore, it seems that there is not a fixed time scale but a strong connection between the speed of evolutionary change and environmental conditions (Kryazhimskiy et al. 2009).

Table 1.1 Examples of behavioral patterns of moral agency, classified along spatial and temporal scales

		Spatial scale		
		Single agent	Group of agents	Collective of agents
Temporal scale	Immediate acts	Intuition-driven behaviors	Instant praise or punishment of actions	Mob behavior
	Deliberate reasoning	Meditation on a moral decision	Collective decision-making in medical ethics	Institutionalized processes of praise, blame, reward, and punishment
	Ontogenetic scale	Development of virtues, character	Development of group reputation	Change in legislation with respect to bioethical issues
	Evolutionary scale	Emergence of moral emotions	Emergence of patterns of cooperation	Cultural change and fragmentation

involve an entanglement of several time-scales, for instance, when many individual immediate acts collectively result in a long-term social outcome (e.g., pollution, climate change, and the tragedy of the commons).[14] However, this classification also allows us to categorize some ethical questions within a specified scale. Based on these distinctions, in total, 12 different categories can be distinguished, to which specific behavior patterns with moral significance can be attributed (Table 1.1).

For each of these behavioral patterns, the details of the agent(s)' competencies, the normative frame, and the context will have to be specified, if they are to become an object of systematic investigation. This will be outlined in more detail below.

1.2.4 What Does 'Being Informed' Mean?

Our discussion so far has focused on the phenomena of moral agency that are especially amenable to empirical research. In the following, we discuss the extent to which these phenomena are relevant to what is often considered the genuine task of ethics: reflecting on normative theories and finding justifications for actions and goals that moral behavior should pursue or promote.[15] To do so, we first outline in more detail the difficulties that arise when attempting to draw clear distinctions

[14] One issue is the possibility of responsibility towards future generations with respect to general behavior patterns of current societies (Birnbacher 1995), which reemerged in the context of debating the moral significance of climate change.

[15] In the following, we will not go into much detail with respect to meta-ethical issues that focus on the logical, semantic, and pragmatic structures of ethical argumentation. However, we recognize that the distinction between ethics and meta-ethics is not easy to draw (e.g. Düwell et al. 2002: 3) and that some of the issues discussed further in this classification may also be considered meta-ethical.

between the different tasks and types of problems that moral philosophy is often concerned with. Next, we discuss how this understanding of ethics affects the appreciation of empirical data within the field.

There are several distinctions moral philosophers consider essential to their task. A closer investigation of them, however, often encounters pitfalls. The problem with one distinction—between 'morality' and 'ethics'—has already been discussed; and it is probably not a crucial one the field has to defend. Likewise, the problem of finding criteria by which to distinguish the moral from the non-moral does not threaten to undermine the whole ethical endeavor; rather, this point denotes a relevant problem with which the field is dealing. However, there are two (interrelated) analytical distinctions that are at once central to the task of ethics and deeply connected to the empirically informed approach to ethics: the is-ought dichotomy[16] (sometimes also called the fact-value dichotomy) and the difference between explaining and justifying behavior. We will not outline the long-standing discussions of these issues, but it can easily be observed that whenever ethics becomes practical—e.g., when training professionals of other disciplines in ethics— these distinctions are mentioned as key instruments of the analytical toolbox of ethics, and accusing a philosophical opponent of committing the "naturalistic fallacy" (see footnote 16) is often taken to be devastating in any practical discussion of ethics (an example is Arn 2009). We are, however, skeptical about both the certitude of these distinctions and their alleged usefulness in practical discourse—and this skepticism is related to a criticism of an understanding of ethics that places empirical knowledge on a distinct plane, detached from the realm of normative justification.

To outline this skepticism, we distinguish three different ways of relating empirical data to ethical theorizing and discuss each of them separately:

1. *Empirical data as a framing of an ethical problem:* All ethical questions—in particular those that concern practical issues, such as stem cell research—have essential conceptual connections to the real world (e.g., one needs to know what stem cells are in order to analyze ethical questions of stem cell research). This is, at first sight, a trivial and uncontroversial inclusion of empirical data in ethical reasoning. However, even this involvement of the empirical may become tricky, as it has what seem at first blush to be obviously value-based conflicts but may actually be factual conflicts (Daniels 1996). Furthermore, the observation that, in practical discourses, it is often difficult to see this difference might indicate that even this seemingly clear-cut involvement of data may blur the fact-value dichotomy. There are two potential explanations for this problem. First, the number and complexity of the facts that must be grasped to understand a problem may be quite great, which complicates the task of identifying hidden normative assumptions about some facts. This problem may be overcome by allowing sufficient time for investigation and deliberation, but this solution may fail when taking the

[16] This distinction goes back at least to David Hume (1739–1740/2003) and should be distinguished from the so-called naturalistic fallacy problem raised by G. E. Moore (1903). See also the contribution of Krones (Chap. 15) in this volume.

second explanation into account: the possibility that many of the crucial predicates and properties (e.g., 'generous' and generosity) inextricably combine descriptive and evaluative components (these are sometimes called "thick" terms and properties, following Williams 1985). If these explanations are correct, the "empirical information" ethics may use with respect to certain problems under investigation is not normatively neutral.

2. *Empirical data as an indicator of the feasibility of ethical thought:* A second potential involvement of empirical data, in particular emerging from moral psychology, acts as a practical constraint on ethical theories. Bernard Williams (1985; see also Flanagan 1991) and others have forcefully argued that an ethical theory that is committed to an impoverished or inaccurate conception of moral psychology has a serious competitive disadvantage. Although this may be a common agreement shared also by antecedent exponents of moral philosophy, the involvement of such facts is more demanding than it seems. First, history of science has taught us that empirical research is an endeavor that is less rational (e.g., with respect to the choice of research topics and theory defense; Kuhn 1962) than initially anticipated. Thus, the empirical data that is expected to constrain normative theorizing is itself the product of a complex and contingent process with respect to what is investigated (and what is not investigated). For example, it is remarkable that current research in social neuroscience has a strong focus on "good behaviors" (empathy, cooperation etc.), whereas a few decades earlier quite different topics were the primary objects of study (Matusall et al. 2011). This makes the constraints of normative theories dependent on culturally shaped trends within the science that delivers the data, and thus ultimately the social and political forces that determine funding priorities. Second, the measurement process involved in establishing such a fact (e.g., how empathy frames perception of moral problems; Singer et al. 2006) involves normativity both by specifying the details of the setting and with respect to the normative frame that serves as reference point (Christen 2010). For example, data emerging from patients with focal lesions in the prefrontal cortex that play a significant role in arguments for the significance of emotions as a "foundation" of moral intuitions and for practical decision making are remarkably imprecise with respect to what kind of emotions are affected. Such findings are also highly prone to misinterpretations driven by prejudices about what the data should demonstrate, as the famous case of Phineas Gage showed (Macmillan 2000). Third, systematic epistemic injustice (Fricker 1998) may lead to biased data-collection and -interpretation, creating a vicious feedback loop in which mistaken normative assumptions lead to erroneous conclusions which in turn are used to support those very assumptions. The data are therefore not independent of the investigators' normative frame, but involved in a complex feedback loop with it. Finally, philosophical interpreters of scientific results are often unaware of raging controversies within the scientific discipline over the validity of those results.[17] Such

[17] Neuroimaging—a central tool in today's social and cognitive neuroscience (Matusall et al. 2011)—is such a complex methodology that recently gave rise to an intense debate, see http://www.edvul.com/voodoocorr.php (accessed on November 3rd 2011) for an overview.

methodological issues require choices with respect to credibility and plausibility of the empirical data that are taken to constrain ethical theories, another way in which normativity comes into play with respect to the fact-value dichotomy.

3. *Empirical data as foundations of normative theories:* Finally, empirical data of a special kind is also involved in a central way in ethical theorizing: when performing thought experiments. Such experiments can be understood as "intuition pumps" (Dennett 1984) and are set up in such a way as to elicit assent to or even certitude in certain philosophical judgments. The inner state of experiencing this assent or certitude is an intuition, which many philosophers are inclined to treat as data against which moral theories are to be tested (Singer 1974). This poses the question of the reliability of this data and its relation to normativity. The first point has been increasingly investigated by experimental philosophers, who find considerable variance in laypeople's philosophical intuitions (Knobe and Nichols 2008; see also the contribution of Shaun Nichols and colleagues (Chap. 9) in this volume), suggesting that cultural diversity is reflected in very basic intuitions about metaphysical and moral issues, too. Surely, one may object that lay intuitions are not data of sufficient quality, but recent investigations focusing on "expert intuitions" of moral philosophers indicate a similar degree of variance.[18] An explanation for this variance may lie in the murky entanglement of the normative and factual aspects of intuitions, which are often taken to serve as both data and "genuine persuaders"—the latter due to an involvement of emotional aspects that probably should be investigated further. It is also important to note that this problem concerns not only individual intuitions, but also the way we combine intuitions, principles, and other elements of a theory into a coherent whole (e.g., by using the method of the reflexive equilibrium; Rawls 1971/1999). We have to expect various similarity relations between such entities (Thagard 1998) that will also rely on intuitions (Christen and Ott 2013). Therefore, the persuasiveness of such intuitions involves a normative component that bridges the fact-value-dichotomy, again.

There are various consequences of this skepticism with respect to clear-cut distinctions between the world of facts and the world of norms. Three of them are outlined in contributions of this volume. Johannes Fischer critically investigates the understanding of ethics as a rational justification of moral judgments. He comes to the conclusion that moral reflection from this orientation cannot do justice to moral phenomena. Furthermore, this view cannot be deduced from the essence of morality, nor can it be substantiated from the fact that we sometimes err morally, nor even can it be deduced from the idea that one of the tasks of ethics is the resolution of moral conflicts.

Adriano Naves de Brito reflects on the foundation of basic values that ethics itself fosters—universalism and egalitarianism. He argues that morality is to be seen primarily as functional and can be understood naturalistically. He provides an example of an explanation of values in terms of preferences, affections, and other

[18] See Schwitzgebel and Cushman (2012) and Krist Vaesen & Martin Peterson, The Reliability of Armchair Intuitions (unpublished manuscript).

agentic dispositions. His recognition of an asymmetry between indignation and shame or guilt, which he considers as fundamental to morality as a system of reciprocal demands, is the key element of the analysis. He concludes that universality and equality are to be defended in any tolerable human concept of morality, simply because they are essential elements of human morality, and not because it is rationally plausible to choose them.

Finally, Tanja Krones reflects on the role of ethics from the point of view of sociology of science and shows the deep entanglement of empirical and normative issues in various practical questions in bioethics. She delineates a context-sensitive, transdisciplinary model of bioethics and (social) science beyond old dualisms and disputes, and presents various results of case studies resulting from empirically informed ethical theorizing.

These considerations and contributions demonstrate that an empirically informed ethics has a fraught relationship with facts and data compared to ethical theorizing that basically uses data as an exploitable resource. It involves both sensitivity to the various ways in which empirical and normative issues are entangled and an understanding of how the relevant data has been generated. The next two sections explore this point in more detail, taking the other contributions of the book emerging from various disciplines as exemplars.

1.3 Methodological Distinctions

As we've seen in the previous sections, the relationship between empirical insights and ethical theory is manifold and complicated. It should come as no surprise, then, that a variety of methodologies have been fruitfully brought to bear on moral issues. In this section, we canvass four of the more important methodological distinctions relevant to empirically informed ethics.

1.3.1 Quantitative/Qualitative

One basic distinction in the social sciences is between quantitative and qualitative research methods. While it is difficult to provide hard and fast definitions of the two methodologies, examples are easy to come by. At a bare minimum, quantitative research aims to establish statistically significant relationships between and among variables; it generates numerical data on these variables, and then tests for correlations in that data. For example, a researcher might ask people to rate their own generosity on a scale ranging from "not at all generous" (coded as −1) through "somewhat generous" (coded as 0) to "extremely generous" (coded as 1), and then provide them with the opportunity to donate money to a charitable organization. The researcher could then test the extent to which self-reported generosity correlates with charitable giving. She might find that these variables are uncorrelated,

meaning that even if you know that someone thinks of himself as generous or stingy, you cannot predict with confidence whether he will donate to a charity. She might instead find that the variables are positively correlated, meaning that people who self-report generosity (stinginess) can be predicted to donate more (less) than the average person. Or she might find that the variables are negatively correlated, meaning that people who self-report generosity (stinginess) can be predicted to donate less (more) than the average person. Any of these findings would be relevant to ethical theories that countenance both virtue and introspection (or some other form of self-knowledge).

We are not, of course, recommending that such a simple and transparent method would yield many insights. The example is merely meant to illustrate how (more complicated and better-designed) quantitative research might be taken to be relevant to ethical theory. Such research includes much of personality psychology (which attempts to develop scales of normatively charged dispositions), social psychology (which investigates situational influences on behavior), behavioral economics (which explores the influence of social, cognitive, and affective factors on economic decision-making), and experimental economics (which uses controlled experiments in laboratory settings to understand preferences, desires, and markets).

Quantitative research methods offer many benefits. Their results can be analyzed statistically, replicated across time and research groups, and modeled in exquisite detail. However, some questions cannot be investigated quantitatively—at least not yet. Furthermore, quantitative research sometimes seems to lack ecological validity. For these reasons among others, it's also important to use qualitative research methods. Qualitative research aims to explore how people experience the world, without imposing the researcher's own agenda and categories on that perspective. Examples of qualitative research include open-ended interviews, sociological observation of group dynamics, some aspects of primatology, and so on. Arguably, Carol Gilligan (1982) would not have been able to develop the ethics of care without going through the painstaking process of interviewing men and women about their moral views and behavior.

Thus, we do not want to make an invidious distinction between quantitative and qualitative research methodologies, but this distinction is important to bear in mind because of the different strengths and weaknesses of the two methods.

1.3.2 Explicit/Implicit/Behavioral

In addition to the distinction between quantitative and qualitative research methodologies, we find it helpful to distinguish explicit, implicit, and behavioral methods. Explicit methods attempt to directly measure whatever variable is at issue. For instance, if you want to investigate industriousness, you might just ask people whether they prefer striving for long-term or short-term goals, whether they think of themselves as industrious, and so on (Duckworth and Seligman 2005, 2006; Duckworth et al. 2010). Explicit research has the advantages of being simple, straightforward, and economical.

However, you might worry that in some cases explicit methods will be subject to systematic bias. For instance, when a personality psychologist asks people whether they have a virtue, participants might be self-deceived, they might want to impress the researcher, or they might tell the researcher what they think she wants to hear. Presumably we can trust people's self-reported extroversion more than we can trust their self-reported honesty or humility. This problem of self-presentational bias is probably more pronounced when moral issues are the object of investigation, as people usually are quite sensitive to what is socially desirable or objectionable. To supplement, complement, or correct explicit research, then, it's often advisable to use implicit or behavioral methods.

One common implicit measure is the so-called implicit association test (IAT), developed by Greenwald et al. (1998). Such a test aims to detect the strength of a subject's automatic associations between various concepts or objects. Subjects are presented with words, images, or symbols one at a time. They classify these items as belonging to one of two disjunctive categories. For example, subjects might have to say whether 'career' or 'Emily' belongs in the *male or work* category or the *female or family* category. The disjunctions are then permuted so that the subjects have to classify the items into either the *male or family* category or the *female or work* category. The answers to these categorization tasks are always easy. What's tested is not accuracy but the speed with which the subjects are able to make the classifications, the assumption being that if you're faster when dealing with *male or work* than with *female or work*, you implicitly associate work with the male gender. IATs have been developed for many categories, but only quite recently has one been successfully developed for morality (Perugini and Leone 2009).

Thus, one could investigate, for instance, honesty with both explicit and implicit methods, and one might find that the pictures that emerge are consonant or dissonant with one another. The distinction between explicit and implicit measures is not exhaustive, though. One might also investigate honesty with behavioral methods. Behavioral economists Mazar et al. (2008) did so, for instance, by providing people opportunities to cheat.

These methods can, and in many cases should, be used in conjunction with one another. We gain a more nuanced and complete understanding of human morality by bringing to bear a variety of perspectives and methods, and by calibrating and correcting some methods with others.

1.3.3 Individual/Social

A third methodological distinction is between individual and social research. As we emphasized above, morality is a social phenomenon. Furthermore, social science research continues to turn up evidence of the mutual interpenetration of the personal and the social. While there would of course be no society if there were no people, and there could be people without society, almost every person who has lived in the past several millennia was enculturated into a social world. When researchers

decontextualize subjects by removing them from that world and putting them into a laboratory environment, they are able to control, to some extent, for social and cultural influences, which in turn enables them to examine the properties of the individual. However, doing so threatens to make their research ecologically invalid.

Moreover, many morally relevant aspects of individuals, such as reciprocity, self-presentation, and benevolent or altruistic preferences, can only be studied in a social setting. For this reason, experimental economists, for instance, have people play economic games such as the ultimatum game, the dictator game, the trust game, and the public good game with other participants. Likewise, anthropologists investigate the interaction of social norms with individual behavior. As with the previous distinctions, the distinction between social and group methods is not invidious: different methods will be more appropriate for answering different questions, and many questions will be best addressed by a smorgasbord of methodologies.

1.3.4 Real/Virtual/Simulated Worlds

One final methodological distinction may become more relevant in the future as a result of the tremendous improvement of computing technology and the pervasive nature of technological information-processing in many aspects of everyday life: the one between real and virtual/simulated worlds. It is a well-established observation that various scientific disciplines currently experience a profound transition through the use of computer simulations (e.g., solid-state physics, chemistry, molecular biology, and climate physics)—and disciplines in humanities and social sciences will surely be transformed by this methodology in the near future as well.[19] Furthermore, sophisticated computer games and virtual worlds are becoming more and more part of the everyday life of many people, providing new "playgrounds" for moral behavior.

Surely, counterfactual thinking in the sense of thought experiments always has been an important instrument in ethical theorizing, but simulation techniques and virtual worlds may become new instruments for understanding moral behavior and perhaps even "testing" ethical theories, thus supplementing empirical research on real world behavior. This approach may prove to be useful in various respects. Computer games—Serious Moral Games (Christen et al. 2012) —may provide frameworks for more realistically assessing the moral behavior of agents compared to simple psychological tests. Simulations also force the researcher to conceptualize a specific problem in detail in order to make, e.g., an agent-based model run properly. They also allow creating and replicating more complex thought experiments. So far, this novel methodological approach is only rarely used in ethics (see Danielson 1992 as an early example of using computer simulation methods in ethics),

[19] An example is the FuturICT project, a mega-science-proposal to build a research community and supercomputer infrastructure to simulate whole societies for understanding social change and predicting crises; see http://www.futurict.eu/; accessed on November 4th 2011.

but it is not hard to predict that research in morality will make more use of virtual worlds and simulations in the near future. Furthermore, they obviate to some extent concerns about the ethical treatment of research subjects. Many interesting ethical questions cannot be empirically investigated because the requisite research would involve harming subjects. Simulated subjects, however, can sometimes be "harmed" with moral and legal impunity.

1.4 Relevant Data for Empirically Informed Ethics

Various empirical research traditions—sometimes referred to as "descriptive ethics" (Düwell et al. 2002: 2)—study morality, including (developmental, moral, and social) psychology, (moral) sociology, and history of morality. As explained in the previous section, though, morality has been increasingly recognized as a worthy object of study by other disciplines in the humanities and social sciences.[20] The same holds also for some disciplines within the natural sciences (in particular, neuroscience, anthropology, and primatology). Even in medicine, "deviations" in moral behavior have (again) increasingly become an issue (Christen and Regard 2012). The following overview will admittedly not be complete, but focuses on the contributions of this volume to sketch the variety of empirical approaches that are in use today.

1.4.1 Phylogeny of Moral Agency

Asking the question of the origin of morality has long been a central topic for scholars interested in morality, and, since the groundbreaking work of Charles Darwin and his prominent followers (Herbert Spencer, Julian Huxley and others), empirical approaches to this question referred to the concept of evolution when looking for answers (Joyce 2006). This search for the "phylogeny" of moral agency requires a specific framing of the problem and goes along with several well-known questions and problems that have been intensively discussed by evolutionary biologists and philosophers (e.g., Boyd and Richerson 1985; Caplan 1979; Kitcher 2011). Among them are the normative significance of a genealogy of morals and the interplay of cultural and biological evolution.

One characteristic of this endeavor is that the search for the phylogeny of morality is heavily framed by one's understanding of moral agency, especially the competencies needed to count as a moral agent. A "traditional" understanding may outline that morality requires sophisticated abilities with respect to language, cognition, and reasoning, such that these abilities have to be fully developed and present

[20] Examples can be found in behavioral economics (Gintis et al. 2005), political science (Haidt and Graham 2007), and pedagogy (Huff and Frey 2005).

within the agents *before* anything like morality could develop. This understanding is reflected by a myth prevalent in many (theistic) religions, according to which otherwise fully-developed people obtained their system of moral rules from an external, divine source. However, this is not a convincing framing of the problem given the emerging knowledge about the deep embedding of moral behaviors in our biological nature. But the price is then a rather loose understanding of morality—in particular with respect to the justificatory aspect of morality, as there is at present no way to analyze when, how, and to what extent people actually started to use justifications in evolutionary history.

Given the structure of moral agency described in Sect. 1.2.2, a crucial issue for any phylogenetic explanation of moral agency is the identification of the competencies agents need in order to produce behaviors that are candidates to be called 'moral behaviors'. Paleontological data is hard to obtain on this issue (an exception may be archeological excavation of burials indicating some degree of care towards the dead), which is why the behavior of primate relatives has become a source for investigating such behaviors and the competencies required for them. The contribution of Sara Brosnan (Chap. 5) to this volume is an example of this approach, as she looks for those behaviors in primates that may be related to social norms, as well as for potential mechanisms for moral behavior, such as empathy. Emerging from observation studies (and increasingly from experimental investigations) of primates living in both natural and "artificial" (zoo, etc.) environments, a remarkable increase in research with respect to "precursors" of morality in species other than humans can be observed. This research is relevant to empirically informed ethics because it allows us to obtain a clearer picture of the "basal" or "paradigmatic" moral behaviors that form our moral lives in the sense that they are shared with our socially sophisticated evolutionary cousins.

With respect to the relevant context for the phylogeny of moral agency, there is little disagreement among the researchers in the field: the specific environmental conditions and the lifestyle of human foragers—i.e., the spatial scale of the (small) group with strong mutual interdependencies and relations—shaped (human) moral agency in a decisive way. In the contribution of Carel van Schaik, Judith M. Burkart, Adrian V. Jaeggi and Claudia Rudolf von Rohr (Chap. 4), an extended hypothesis building on a large body of research in anthropology, ethnology, and related sciences is presented. They propose that moral emotions are the subjective side of the proximate rules (motivations) that regulate human cooperation, which in turn is an evolutionarily novel adaptation to enable the uniquely derived lifestyle of human foragers, which requires generosity and sharing due to extreme mutual interdependence. For an empirically informed ethics, such a theory is relevant not only to understanding the origins of human morality, but it also has normative implications that are hard to ignore. For example: What follows from the fact that the current human lifestyle is far removed from the one our ancestors had over many thousands of years (e.g., with respect to establishing cooperation)?

An emphasis on the role of context for the emergence of moral agency leads to some difficult questions, as we can expect that (on an evolutionary time scale) the context was pretty stable with respect to the decisive elements (e.g., scarcity of

resources). One such question is: how is diversity with respect to the content of the normative frame possible? This diversity is obvious from a historical perspective, unless one restricts morality to those very few behavior patterns (e.g., represented in the "Golden Rule") that seem to be quite robust across cultures and times. This is where culture comes into play, the focus of the contribution of Jesse Prinz (Chap. 6). The examples provided by Prinz are, however, not only a reminder to be cautious in seeing (deterministic) connections between biological human nature and the moral systems that emerge out of them. They also remind us that we currently lack a systematic investigation of the "normative knowledge" humans have accumulated within their history. Rather, this type of knowledge seems to be dispersed in many different disciplines, including history, theology, philosophy, and political science, among others.

1.4.2 Ontogeny of Moral Agency

A second major question for scholars interested in morality is: How do human beings become moral beings in their lifetime?—A question whose practical relevance is especially pertinent to moral education. Again, the time-scale partially defines the problem, whereas all spatial scales may have effects on the ontogenesis of moral agency. The contributions of the book associated with this major question refer to several topics associated with the ontogenesis of moral agency.

One major issue to solve is the question of which competencies a moral agent should develop in the course of the first few years (or perhaps even decades) of life. Carmen Tanner and Markus Christen (Chap. 7) investigate this issue in their contribution by presenting a broad overview of the various competencies that have been proposed in the literature (in particular, moral psychology) as essential for moral agency. In their contribution, these abilities are arranged in a model called moral intelligence, which highlights two particular aspects: First, motivation—captured by the competence of "moral commitment"—gains a central role in influencing all other competencies of the psychological model of moral agency. Second, the model proposes a way in which the normative frame—captured by the notion of a "moral compass"—is integrated into the psychological processes that generate moral behaviors, mediated by the competence of moral commitment. The framework of "moral intelligence" stresses the importance of making such competencies measurable in order to have a basis for evaluating the effects of "moral training." This type of research informs ethics by emphasizing that any normative theorizing requires an understanding of the competences needed for moral actions whenever these theories are applied for practical purposes.

A second important aspect of the ontogeny of moral agency is the context in which the moral agent develops. With respect to humans, it is increasingly appreciated that early life experiences have important repercussions throughout an individual's lifetime. The contribution of Darcia Narvaez and Daniel Lapsley (Chap. 13) outlines this point by presenting both empirical data on early childhood

development and a theoretical framework called "triune ethics," which distinguishes three basic ethical systems with respect to their functional effect (Safety, Engagement, and Imagination Ethics). These systems are differently shaped by early childhood experiences. To move towards moral expertise, extensive focused practice is required under the guidance of a mentor. Such education involves the cultivation of a deliberative mindset along with immersion in environments that foster appropriate intuitions.

A third aspect when assessing the single moral agent concerns the "neuronal infrastructure" that implements the competencies necessary for moral behavior. This very recent research—also called "neuroscience of ethics" (Roskies 2002)—is presented in the contribution of Kristin Prehn and Hauke R. Heekeren (Chap. 8). A question of particular importance is whether this research will be able to distinguish between domain-specific and domain-general capacities needed for moral agency, and how individual differences reflected by the variance of this neuronal infrastructure influence moral judgment competencies. Important methodologies used in this research are neuroimaging (in healthy subjects as well as in subjects with specified brain lesions), transcranial magnetic stimulation (TMS) and other non-invasive tools for casually influencing neuronal processing, and behavioral experiments. The authors claim that neuroscientific research helps to disentangle the processes involved in moral judgment and behavior, and that it enables us to test the numerous assumptions made by psychological theories of moral agency.

A fourth aspect of the ontogenesis of moral agency is the difference in expertise among moral agents. Many studies within cognitive psychology have shown the superiority of experts over novices in nearly every aspect of cognitive functioning—leading to the question of whether one can speak of moral expertise in a similar way. Bert Musschenga (Chap. 11) investigates this issue in the framework of reflexive equilibrium by examining whether the quality of a reflective equilibrium can be strengthened by requiring that the initial judgments come from moral experts. He comes to the conclusion that this expertise is domain-specific: the reflective equilibrium of local ethical theories can indeed be strengthened by giving special weight to the judgments of moral experts. These judgments are superior to those of laypeople if they stay within the locally accepted moral framework. Sometimes, however, moral intuitions transcend accepted moral frameworks. In such cases it is not up to moral experts to determine whether such intuitions are relevant and should be accommodated within an ethical theory.

Finally, the issue of developing moral abilities and expertise becomes a normative question when it refers to persons operating at decision points of institutions and societal systems. This aspect is investigated by the contribution of Markus Huppenbauer and Carmen Tanner (Chap. 14), who address "ethical leadership": the abilities and values that should be fostered in and by people who occupy positions of responsibility in companies and other institutions. In their contribution, they sketch various areas of intersections between ethics and psychology, where each field can learn and benefit from the other when exploring ethical leadership and ethical decision-making. They thus show how empirically informed ethics can become very practical indeed.

1.4.3 Reasons and Moral Agency

A third group of ethical questions within ethics for which empirical data is relevantly important involves the role of reasons and their foundation in intuitions and standards of rationality—a field that requires a sophisticated understanding of morality with respect to basic competencies such as the capacity to reason and deliberate. It concerns in particular the normative frame and the question of how this frame is operative with respect to actual behavior.

There is a longstanding debate about the basic structure of normative theories within ethics: Should they be founded by some basic principles, implying a hierarchy of reasons (foundationalism; see e.g. the contributions in DePaul 2001), or is the core of ethical theorizing more a non-hierarchic network of moral beliefs that "cohere" in some sense (coherentism; e.g. Thagard 2000). The latter approach was of particular importance for ethics when the methodology of "reflexive equilibrium" was introduced by Rawls (1971/1999). Ghislaine J.M.W. van Thiel and Johannes J.M. van Delden (Chap. 10) investigate in their contribution how this theoretical construct can be made more empirical by analyzing the role of intuitions (of third parties) in arriving at reflexive equilibrium. Their so-called Normative Empirical Reflective Equilibrium uses empirical research to obtain information about these intuitions and creates a framework for understanding moral wisdom, grounded in the idea that ethics requires rich, complex and context-sensitive reasoning.

How intuitions actually can become an object of empirical research is shown in the contribution of Shaun Nichols, Mark Timmons, and Theresa Lopez (Chap. 9). They offer a case study at the intersection of moral psychology and normative ethics by investigating the phenomenon of "moral luck." Psychological evidence indicates that people make harsher blame judgments about unlucky agents than equivalently situated lucky agents. Their research suggests that our commitment to allotting greater blame to unlucky agents is an entrenched commitment that runs fairly deep in human psychology and carries some initial normative authority. This initial authority is not beyond critique, but as it happens, people's commitment to outcome-based blame is more sensitive than has been recognized. People are much more likely to embrace outcome-based blame when agents are negligent than when agents are conscientious. This, according to the authors, provides the basis for a more plausible rendering of the control principle—a basic moral intuition according to which a person can only be blamed for what is within their control. Thus, the psychological research not only helps us to assess the nature of our normative commitments, it also helps us to articulate normatively plausible principles. The contribution is an exemplar of how empirical data influences normative thinking.

Finally, Erich H. Witte and Tobias Gollan (Chap. 12) investigate actual justification patterns used in political discussions that include moral issues. They operationalize the main positions of moral philosophy by developing a questionnaire and a content-analytic category system. With these instruments they measure ethical justification as prescriptive attributions in the form of rated subjective importance (questionnaire) or frequencies (content analysis). Both measures enable researchers

to obtain empirical data on how ethical justifications are actually used and to test hypotheses empirically, for instance concerning the dependence of the justification pattern on the kind and quality of action, as well as on culture, role, and mode of group discussion. For example, they find a large difference between the Arabian and the Western culture in justifying war and terrorism. Both direct and indirect utilitarian argumentations seem to be typical for the Western groups; however, emphasizing the bad consequences of the enemy's action for a certain group appears more often in the justifications of the Arabian parties. This data is of particular importance when collective phenomena of moral agency come into the focus of normative theorizing.

1.5 Some Focus Questions for Empirically Informed Ethics

What, then, is empirically informed ethics, and what are its prospects? The contributions of this volume do not paint a complete picture of the empirical investigation of normative issues, but they do serve as exemplars of the many ways in which empirical information can be brought to bear on ethical questions. We have chosen to construe ethics quite broadly as the study of the phenomena of moral agency, which includes three components: agentic competencies, a normative framework, and situational constraints. We can now distinguish between first-order aspects of empirically informed ethics, which draw on data concerning just one of the components of moral agency (e.g., the biological underpinnings of agentic competencies, the ontogeny of the space of reasons, and the psychology of situational influences), and higher-order aspects of empirically informed ethics, which examine either the feedback loops within one of the components across different spatial and/or temporal dimensions (e.g., the effect in later life of praise and blame for infants) or the relations among the components (e.g., the social ontology of virtue and vice). A comprehensive empirically informed ethics would include both first-order and higher-order aspects.

In the following, we briefly summarize how some of these questions can be aggregated to more general questions that empirically informed ethicists should, we suggest, be particularly interested in:

– *The ontology of the moral space:* First, we need a better understanding of what might be called the "ontology of the moral space"—the question of discerning the basic moral entities and their interrelations. We suggest understanding the term 'ontology' more in its information science sense, i.e., by asking how our knowledge of morality is organized as a set of concepts within a domain. We may understand the moral space as an abstract "space of reason" (Sellars 1956) populated by entities that can be understood as interrelated cognitive-affective units, such as beliefs, desires, values, principles, expectations, emotions, and sentiments (Mischel and Shoda 1995). Pairwise connections between units may be weaker or stronger, and the number of connections between a given unit and other units may be lesser or greater. The structure of this network of relations, thus, determines a

topology for the space of reasons, which, when traced using a clustering approach (Christen and Ott 2013), can be subdivided into classes. In the literature, various classifications have been proposed (Autonomy, Community, Divinity: Shweder et al. 1997; Care, Fairness, Loyalty, Authority, Sanctity, Liberty (the sixth class has been added later): Haidt 2007; Self-Direction, Stimulation, Hedonism, Achievement, Power, Security, Conformity, Tradition, Benevolence, Universalism: Schwartz 1992) that only partially overlap. This is neither surprising nor problematic when we understand the partitioning of moral topology as a bottom-up classification problem, but it shows us which problems should be at the center of our focus: evaluating the effect of different kinds of similarities between moral beliefs, their physiological implementation in the agent, and connections between the topology of the space and actual behavior patterns.

– *The function of morality:* Many empirical approaches to morality (in particular with respect to the origins of morality) assume a functional view of moral agency, i.e., see it as something that favored survival (and flourishing) of groups and societies. This is obviously a restricted view with a conformity bias that does not take into account that moral behavior may also include non-conforming behaviors that directly threaten social cohesion (e.g., conscientious objection to military service). Furthermore, a functional view may have irritating normative consequences (see, for example, the sociobiology debate; Caplan 1979). And even an alternative functional understanding of ethics as an instrument that helps to solve moral problems—e.g., using the "just community" approach of Kohlberg (1980)—may be rooted in the preconception that moral problems require solutions. But it is reasonable to assume that at least some moral problems are expressions of cultural diversity within societies and can only be solved by eliminating this diversity. In other words: there are various and conflicting interpretations of the function of morality that need to be investigated further.
– *Understanding moral change:* Another relevant phenomenon to be investigated further is moral change, especially the interrelations between the diversity of moral systems present at a specified time point and their further development. We need a better understanding of what drives moral change given the evidence that rational inquiry is probably not the main driving force in this dynamic, whereas "mavericks" in the ethical discourse (e.g., honor; Appiah 2010) may be of more importance. In this context, there may be important feedback loops between moral behavior, on the one hand, and deliberation about holding agents morally responsible, on the other hand, which could be investigated empirically. In cases where moral behavior arises emergently as a result of the interactions of many agents over a long timescale, the effort to break the system down piece by piece in order to locate the nodes of primary responsibility may instead destroy the system, thereby making the phenomenon harder to understand and ameliorate. Attempts to hold a single person or small set of people responsible may end up scapegoating them and ignoring the structural problems embedded in the social, political, or economic system—problems that would arise regardless of the identities and motives of the individual agents involved.

– *Dealing with moral complexity:* A further interesting phenomenon refers to the complexity-simplicity relation with respect to morality. On the one hand, the involvement of ethical thinking in problem evaluations is often experienced as an act of "complexification" (Casti 1995) by outlining the various facets of a problem. On the other hand, a moral appraisal of a situation often has a simplification effect (in particular in non-dilemmatic situations and when the agent has a clear reference scheme, such as protected values; Tanner and Medin 2004); it makes a seemingly complex problem easier to decide. In short we may say: ethics makes things more complex, morals make them simpler. This is an interesting interplay given the ongoing discussion of the role of intuitions in moral reasoning (triggered by the contribution of Haidt 2001) that deserves a further investigation.
– *Ethical theory building and standards of rationality:* Finally, another core problem of empirically informed ethics refers to the act of theory building itself. What are the historical contingencies that shape the appreciation of ethical theories both within the community of thinkers and on a broader societal scale? What influences the basic intuitions of moral philosophers when they seek to justify their theories? These questions have been addressed already, of course, but usually without systematic empirical investigation and primarily in the gladiatorial mode of attempting to refute or undermine an opponent's view. It would be interesting to investigate this phenomenon from a more external, empirically informed position.

This set of basic questions is not complete. But we suggest that it grasps relevant problems where both ethical thinking and empirical data on moral agency expressions will be required to arrive at genuine insights. And such an outline of these basic problems—something like a "Hilbert list" of problems for empirically informed ethics[21]—could serve as a guideline for future research.

Such a "Hilbert list" for empirically informed ethics powered by the general popularity of methodological naturalism in philosophy, however, should not set aside critical philosophical thinking. As Antti Kauppinen analyzes from an "armchair point-of-view" in his Critical Postscript, many even modest Ethical Empiricist arguments are unsound or at least dubious, and the empirical evidence provided often fails do the work it is alleged to do. Thus, empirically informed ethics should not pursue the "dream of scientism" that empirical science constitutes the most authoritative worldview also within ethics. Rather, the challenge will be to find the junctures between "armchair thinking" and "empirical ethics" such that our understanding of humans as moral beings is promoted.

[21] On August 8th 1900, the German mathematician David Hilbert presented a list of 23 unsolved mathematical problems at the International Congress of Mathematics in Paris that turned out to become a programmatic outlook influencing the research within the field for many years.

Chapter 2
What Kind of Ethics? – How Understanding the Field Affects the Role of Empirical Research on Morality for Ethics

Johannes Fischer

2.1 Introduction

The principles underlying morality have occupied ethics since its genesis during the Greek Enlightenment. Aristotle was one of the first to ponder the matter and then, much later, David Hume formulated the alternative that became so crucial to the ethics of modernity, namely whether morality has its basis in reason or sentiment (Hume 1751/2002: 3). This interest in the principles underlying morality is not solely attributable to moral psychology, but is also concerned with the comprehension and very conception of ethics. In the light of this historical background, it is worthy of note that, today, many ethicists are wont to display a certain reticence and skepticism toward new empirical psychological and neurobiological research into moral behavior. The reasons for this are manifold. The first is almost certainly to be found in the widespread view that ethics is concerned with normative questions, as well as the examination of moral language and the logical structure of moral arguments. Empirical research on morality seems incapable of making any contribution here. A second reason has, in all probability, to do with the excessive aspirations and expectations linked by some authors to this type of research, which come close to a "naturalization" of morality and ethics.[1] Third, such research frequently raises the question of whether what is being experimentally examined as "morality" is not in fact based on models that are far too simple, ones that might relate to a pre-theoretical everyday comprehension of morality, but that do not grasp the phenomenon in the differentiating approach adopted by ethicists.

The most important reason, however, is the fact that findings from state-of-the-art empirical research on morality run counter to a perception of ethics characterizing

[1] An example is Casebeer (2003: 843): "The goal of naturalized ethics is to show that norms are natural and that they arise from and are justified by purely natural processes".

J. Fischer (✉)
Institute of Social Ethics, University of Zurich, Zurich, Switzerland
e-mail: fischer@sozethik.uzh.ch

M. Christen et al. (eds.), *Empirically Informed Ethics: Morality between Facts and Norms*, Library of Ethics and Applied Philosophy 32, DOI 10.1007/978-3-319-01369-5_2,

large parts of modern ethical thinking. According to this perception, the task of ethics is to justify moral judgments rationally (i.e., argumentatively).[2] The mere defining of this task already implies a particular comprehension of morality, deeming that moral judgments can be rationally justified. An obvious tension exists between this view and empirical moral research findings stating that moral evaluations are based on the emotional evaluation of actions or situations. If, namely, moral evaluations develop in this way, then they cannot be demonstrated to a third party argumentatively. Rather, the morally right or wrong can only show itself to a third party when that party takes a close look at the action or situation in question and evaluates it emotionally. The basis and criterion for the truth of moral judgments is then in the viewing and the imagining rather than rational justification, meaning that the task of ethics has to be defined differently.

Current empirical moral research therefore raises the challenge of subjecting a widely accepted view of ethics to a critical reappraisal. This is what the following will attempt. I shall proceed by, first, examining the arguments used to validate the comprehension of ethics as a rational justification of morality, in order to expose their untenability against a background of the current empirical findings on morality. I shall then ask what types of reasons are encountered within morality and how they differ from arguments. Finally, I shall advocate the theory that rather than providing a rational justification of moral judgments, ethics has the far more modest task of guiding us toward correct moral thinking that does justice to moral phenomena.

2.2 Three Arguments in Favor of Comprehending Ethics as a Rational Justification of Moral Judgments

Enquiring into the foundations of this view of ethics, three arguments emerge. The first argument deduces from the very essence of morality an obligation to justify moral claims argumentatively. The second argument points out the unreliability of our moral intuitions and their susceptibility to error, making a rational revisal indispensable. The third argument claims that ethics is primarily concerned with moral conflicts, the latter as such requiring rational clarification.

2.2.1 The First Argument: An Obligation to Justify Moral Judgments Rationally Is Rooted in the Very Essence of Morality

The first argument can be summarized as follows: orienting oneself morally means to orient oneself toward moral values, or judgments of the type ‘X is right/wrong, good/bad.’ Judgments are connected to a claim of general or intersubjective validity.

[2] Unless stated otherwise, in the following I shall use the expression ‘rational’ in the sense of argumentative rationality, as distinct from purposive rationality (cf. Höffe 1984).

In pronouncing them, the speaker assumes an obligation to honor this claim if required. This occurs in the form of an argument. Arguments are reasons aimed at proving the truth of an assertion, so that others are compelled to acknowledge them. Ideally, an argument takes the form of a deductive conclusion, the validity of which can hardly be refuted. Moral judgments thus imply an obligation to justify argumentatively and, in this sense, rationally. Since this takes place via rules, principles or general criteria, ethics then has to be conceived of as an ethics of rules. The following quotation is characteristic both of this view of morality and of ethics as an ethics of rules (Nida-Rümelin 2006: 3; translated from German):

> Ethics starts out from moral convictions. Moral convictions deem which things are good, which actions are morally impermissible, which distributions are just, etc. Ethical theory attempts to develop some general criteria for good, correct, etc., which, on the one hand, harmonize with apparently unrelinquishable individual moral convictions and, on the other hand, can provide orientation in cases where our moral views are uncertain or even contradictory.

What is wrong with this first argument are its premises. First, it is not true that all courses of action attributed with possessing a moral quality have their orientation in moral values. If we witness an accident in which people are injured, then the most obvious reason for helping is the fact that injured people require help, and not the notion that it is morally right or imperative to help them. Here we have a situation before us that demands a certain course of action, and it is the situation that forces us to take this course of action. Let us assume we would miss an important appointment we had agreed to keep and need to make our apologies if we would help these injured people. We will name this situation as the reason for our absence and, in so doing, will narrate it to the other person, giving him a narrative account of it in order to convince him of the correct, indeed imperative nature of our course of action. Narratives bring the moral significance of situations to mind. Narrative reasons, as shown in this example, are different from arguments in that here it is not the speaker showing his counterpart that something is true or right and proving it to him, but the something showing itself to the counterpart when the latter depicts the situation in his mind or has it depicted for him. Unlike the necessity resulting from the "non-coercive coercion" (Habermas 1981: 52f) of an argument, here we have a form of necessity that emerges from a given situation and from which we are unable to escape as a result of visualizing it (cf. Winch 1987). I shall return to this point later.

The idea that moral orientation is an orientation toward moral evaluations is what gives rise to the view that the task of ethics is to justify moral evaluations. Ostensibly this forms the basis for the plausibility of an ethics of rules or norms, an attempt to construct general criteria for moral evaluations, as expressed in the above citation. But this incurs a problematic shift in the focus of moral orientation. Let us imagine an ethics committee which is called upon to comment on active euthanasia. If it were to behave as if moral orientation meant an orientation toward moral evaluations, then the task of the committee would have to be perceived as submitting a judgment about whether active euthanasia is morally right or wrong, in order to aid public orientation in this issue. But if moral orientation has its foundations in the visualization of actual cases and imaginable situations, it becomes a different thing

altogether. In this case, the task of a committee would be to promote a public awareness of the situation in which those affected by this issue find themselves, in order thus to enable the addressees of the comment to form their own responsible judgment regarding how to deal with the issue of active euthanasia and any legal regulations that need to be put in place concerning it.

Second, the idea that moral judgments are connected with a claim to general or intersubjective validity is incorrect. With judgments, whether they are moral or empirical, we stake a claim to *truth*, referring to the relationship between statement and fact. In contrast, we stake intersubjective claims to validity with *assertions* or *theses*. With such speech acts we assume a *discursive* obligation to find proof for that asserted if required, and this in turn takes place in the form of arguments. Narrative reasons are too weak for such *proof* since their power to convince contingently depends on whether the moral significance of a situation presents itself to the other person in the same way as it does to the narrator. In many cases we refrain from making assertions or hypotheses because we do not believe ourselves to be capable of honoring the concomitant discursive claim in a manner convincing to all. But this does not mean that we refrain from forming our own judgment in a controversial moral issue. This point is particularly significant in the situation of moral and ideological pluralism and especially within the global context.

This distinction between judgments and assertions, which is so important for the comprehension of morality, is seldom made in the relevant ethics text books. Dieter Birnbacher, for example, writes (2003: 24; translated from German): "He who judges morally, usually comprehends himself… as somebody who makes an assertion and who expects that the addressees of his judgment can understand and follow what is being asserted." Accordingly, Birnbacher counts the *claim to general validity or intersubjective bindingness* as one of the characteristics of moral judgments (Birnbacher 2003: 13, 24).[3] With reference to moral judgments, Michael Quante (2003: 27) speaks of "asserting statements." And yet, saying: "Abortion is morally wrong" and saying: "I assert that abortion is morally wrong" are quite obviously two different things. The second statement, the assertion, stakes a claim to intersubjective validity, which in turn entails an obligation to produce proof of the truth of this statement if required, whereas the first statement, the judgment, entails no such obligation. In line with what has been said so far, this does not rule out the possibility that we would nevertheless put forward *reasons* for this judgment if asked; but these reasons would not take the shape of, or stake the same claim as arguments.

In equating, on the one hand, moral orientation with an orientation toward moral evaluations, and, on the other hand, morally evaluating arguments with assertions, ethics can well be comprehended as an argumentative, and thus rational justification of morality. Neither Birnbacher nor Quante examines whether statements with a moral content can even become the object of assertions. Does it make sense to say: "I *assert* that in certain cases active euthanasia is morally correct or defensible"? Can any argumentative proof be found for this? I shall return to this point shortly.

[3] From this Birnbacher deduces an argument against particular ethics (e.g., Christian ethics). Namely, these types fail to satisfy a claim to general validity (cf. Birnbacher 1991: 113).

Following what has been said, the view that ethics has its task in the rational, argumentative justification of morality cannot be deduced from or justified by the manner in which we morally orient ourselves in our courses of action. As shown by the example of the accident, communication about our moral actions does involve *reasons*, and crucially so, but the latter are not put forward as arguments. So where does this view of ethics come from?

As has already been hinted at, modern ethical thinking is characterized by ethics being assigned the purpose of justifying morality rationally, in order to constrain human coexistence in a reasoning from which no sensible person can escape. From the perspective of discourse ethics, this development may be reconstructed as follows. Societal order and cooperation demand more than simply that the members of society bind their actions to universally consensual reasons. The consensus could be merely factual, due to a coincidental concordance of opinions and convictions. Societal coexistence would then depend on contingent conditions that could change at any time. Far more, what is required is a binding to reasons that everybody *necessarily* has to agree to, independently of the coincidental circumstances of their own origins and conditioning. Reasons of this type are arguments. As I have said, arguments are reasons connected with a claim to be able to produce proof underpinning the truth of an assertion, in such a way as to *compel* consent. If all members of society were to bind their judgments and courses of action to such reasons, then the non-coercive coercion of the better argument would render superfluous every other type of coercion intended to guarantee societal order and cooperation. Modern ethical theories claim to provide such reasons. They are the result of subjecting ethics to the purpose of rationally justifying morality. Ancient ethics was not familiar with such a purpose, and therefore not familiar with such theories (cf. Tugendhat 1984: 33–56).

This means, of course, that moral reflection is subjected to a purpose that is *alien* to morality itself, i.e. a purpose that is not implied by morality itself, and that has considerable consequences. Modern ethical thinking is characterized by an attitude of *disengagement*. Disengagement entails the objectification of a domain, dispossessing that domain of the normative power with which it usually affects us (cf. Taylor 1989: 160). In the example of the accident involving injured parties, it is the moral significance of the *experienced* situation that causes us to act. As has already been said, its linguistic articulation is encased in a narrative. But if the rationality ideal dominating modern ethics is to prevail, then it must cease the importance of how a situation is experienced or how it may be visualized as the result of a narrative, because that would mean a dependence on contingent conditions; only arguments count. Situations are then turned into cases, where rules are applied, and here their *descriptive characterization*, as distinct from their *narrative visualization*, has the upper hand. Grasped purely descriptively, they forfeit their normative effect on us. The hiatus between (descriptive) facts and values that is characteristic of modern ethical thinking and that underlies phenomena such as the naturalistic fallacy, has its origins in this objectification of morality. This hiatus does not exist in our lifeworldly experience, as shown by the example of the accident: *as one* with the narrative visualization of the situation in question, we are simultaneously aware of

the rightness of the relevant course of action. If, in contrast, we view this situation in terms of its descriptive characterization, then we will additionally require a moral rule or norm to stipulate the right or imperative course of action in situations of this kind.

One illustration of modern ethical thinking is an essay originally written in the 1970s by Peter Singer on the question of whether there is an individual moral obligation to help fight global poverty (Singer 2007). Even though Singer makes it obvious that what motivated him to write his essay and what actually moves him, is the suffering of people in the former region of Bengal, nowhere in his essay does this suffering appear as a reason for helping these people. Singer does not open our eyes to what it is like for a person to live in extreme poverty in a bid to make us more sensitive toward the moral significance of this circumstance. Rather, his essay is governed by the idea that only rational justification of a corresponding obligation to help will constitute a sufficient moral reason to help these people. This idea also dominated the debate later triggered by Singer's essay.[4] In the example of the accident, this would be equivalent to the mere fact that the injured parties urgently require help not representing a sufficient moral reason to help them. Only the argumentative justification of a corresponding obligation to help would constitute such a reason.

2.2.2 The Second Argument: Only the Rational Justification of Moral Judgments Makes Moral Knowledge Authentic

Let us now take a look at the second argument backing the view that the task of ethics is to justify morality rationally. It claims that orientation toward an experienced or narratively visualized situation, as in the example of the accident, is *prone to error* and that we can therefore be *mistaken* with regard to our true obligations. In order to find out what our true obligations are, we have to assume a disengaged standpoint and orient ourselves toward arguments alone. A typical example of this view is the following quotation (Rippe 1998: 363f; translated from German):

> More crucial is the differing attitude held by professional ethicists [as opposed to ethical laypersons] towards ethical questions. On the basis of their professional training, they must be prepared to question everything. Faced with a course of action which to a layperson would be quite obviously condemnable, they ask: 'What is really wrong with it?'. What is wrong with torture, murder, slavery, discrimination? The mere fact that they are even asking this question, believing it necessary to examine the relevant arguments, does, of course, hint at amorality. For can there be any better indication of the deficits of experts than that they question what is seemingly obvious to every single person brought up to be a moral subject? This alone marks them out as suspicious. Philosophical ethicists really do have to live with this deficit. Not taking the intuitions of their time, the doxa, as given, they must subject them to rational examination…

[4] Cf. the contributions in Bleisch and Schaber (2007).

Following on from the Enlightenment, modern thinking has been profoundly marked by this view of philosophy and philosophical ethics. As the quotation demonstrates, to a certain extent it is concerned with the professional ethos of the philosophical ethicist. In contrast to the *doxa,* that which presents itself in the foreground and by which the philosophically uneducated abide, this ethos stands for *aletheia*, the truth. Laypersons might *think* that they know that a murder or a rape is a terrible thing, a moral evil, but ethicists possess *true* (i.e., rationally justified) knowledge. Unlike laypersons, ethicists are prepared to question everything, right down to whether a murder or a rape is really as bad, in moral terms, as laypersons believe.

In the midst of all this enlightening pathos, it is easy to overlook the fact that moral prejudices, as highlighted in the above quotation, are not the only kind; philosophical prejudices also exist, in the shape of convictions that are deemed true and then passed on without ever being subjected to unbiased examination. If you say that moral intuition, perception and experience are all susceptible to error, then no sensible person will contradict you. The crucial question here is: what can be used as a yardstick to measure truth and error? In the above quotation, there is *already an implicit assumption* that rational argumentative examination is to be this yardstick. Accordingly, moral intuition or experience is susceptible to error because it can seduce us into holding opinions that are not the same results arrived at by argumentative examination. And yet this implicit assumption is anything but self-evident, rather requiring justification itself. Advocates of this view are usually unaware of this point because, for them, the assumption is so totally clear. What other yardstick for truth and error can there possibly be, if not the incorruptible logic of the argument? What the advocates of this view need to prove is that moral knowledge does indeed conform to this logic.

Thanks to empirical moral research we are now aware of the significance that emotions have for morality (Fischer 2010). If this is so, how is moral knowledge to be arrived at solely through cognitive operations, in the shape of logical conclusions or arguments? Here we have another distinction between modern ethical thinking and ancient ethical conceptions, such as those of Aristotle. Due to the objectification of morality and the resulting attitude of disengagement in modern thinking, reason and sentiment become potential alternative foundations of morality (Hume 1751/2002: 4), whereas, for Aristotle, in moral decision-making emotion and reason worked hand-in-hand (Eth. Nik. VI, 2 1139b 4–5, in Aristotle 1999). The Platonism in the above quotation, distinguishing the *doxa*, i.e. the manner in which (im)moral phenomena, such as torture, murder, rape, discrimination, etc., exist in lifeworldly perception and experience, from the *aletheia*, which presents itself not to lifeworldly experience, but only to rational, argumentatively justified thought, contradicts our current knowledge of morality. The type of radical questioning propagandized in the above quotation thus only ostensibly merits the label 'philosophical.' It does not really deserve this epithet because the all-crucial premise on which the questioning itself is based remains unquestioned.

It is lifeworldly experience that makes us aware of the fact that moral perception of situations or courses of action is prone to error, and not standards of rationality.

For example, yesterday I may have been of the opinion that I treated my child justly, and yet today, after going through yesterday's situation in my mind again, I realize that I behaved wrongly toward him. Only *one* conclusion can be drawn from this, namely to watch out more carefully and check more precisely when evaluating similar cases and situations in the future. The consequence is *not* fundamentally to mistrust my perception of situations and actions and to orient my future behavior toward my child solely in accord with rationally deduced rules and principles, devoid of emotional involvement.

Harold Arthur Prichard (1912) elaborated on these issues, in his classic essay "Does moral philosophy rest on a mistake?", in which he compared moral philosophy to the Theory of Knowledge. The latter had its origins in a doubt concerning whether what we believe to be knowledge really is, in fact, knowledge. This resulted in a search for criteria, on the basis of which we could then be certain that we do, in fact, know. According to Prichard, the mistake behind this view is that the object of doubt was not actually knowledge at all. Therefore, what was at stake was not a criterion for knowledge. "For when we *say* we doubt whether our previous condition was one of knowledge, what we *mean*, if we mean anything at all, is that we doubt whether our previous *belief* was *true*, a belief which we should express as the *thinking* that A is B. For in order to doubt whether our previous condition was one of knowledge, we have to think of it not as knowledge but as only belief, and our only question can be 'Was this belief true?'" (Prichard 1912: 35). But in order to discover this, we have to re-examine what brought us to this belief, which is equivalent to doing a sum again. And, in order to do this, we do not require general epistemological criteria for knowledge.

Prichard diagnoses the same mistake within moral philosophy. It, too, has its origins in a doubt, namely a doubt about whether what we consider to be obligations really are obligatory. "We then want to have it *proved* to us that we ought to do so, i.e. to be convinced of this by a process which, as an argument, is different in kind from our original and unreflective appreciation of it" (Prichard 1912: 36). This has also led to a search for criteria for moral knowledge with regard to our obligations. And here, too, the mistake is that our doubt does not refer to whether we actually know, requiring criteria for moral knowledge, but rather refers to a *belief* or *conviction*, meaning that we should re-examine what led us to this conviction. Modern ethical theories, as stated, claim to provide criteria for moral knowledge. As Prichard makes it clear, both with a view to consequentialist theories, and with reference to deontological theories, they are unable to honor this claim.

Prichard comes to the conclusion "that we do not come to appreciate an obligation by an argument" (Prichard 1912: 29). Far more, the "sense of obligation to do, or of the rightness of, an action of a particular kind is absolutely underivative or immediate. The rightness of an action consists in its being the origination of something of a certain kind A in a situation of a certain kind, a situation consisting in a certain relation B of the agent to others or to his own nature" (Prichard 1912: 27). Instead of establishing general criteria for 'right' or 'imperative,' it is therefore a case of taking a look at the situations in question, using them as a yardstick to judge whether courses of action are right or imperative. If we are in any doubt about

whether a course of action is right or not, these doubts can only be expelled if we re-evaluate what convinced us that it was right in the first place, this being the situation in question or the relationship between the course of action and the situation. Accordingly, Prichard distinguishes between *moral thinking* and *non-moral thinking*, with the latter pertaining to the moral philosophy he is so critical of, a philosophy believing it can deduce obligations argumentatively.

2.2.3 The Third Argument: Moral Conflicts Require Rational Clarification

This brings me to the third argument for viewing ethics as a rational justification of morality. It claims that ethics is essentially concerned with disputes such as dilemma decisions or distribution conflicts. But these disputes require rational argumentative clarification. For example, in the case of how to distribute scarce organs for organ transplantation, the criteria determining the assignation of organs must be decided using argumentation.

This argument is sound in that many ethical questions involve subquestions that require for their clarification not moral considerations, but deliberations of another kind (e.g. purpose-rational deliberations). If a moral imperative to preserve as much life as possible is pre-given, in terms of life duration and quality of life, then purpose-rational deliberations can determine the distribution criteria for the assignation of organs best able to achieve this goal.

This argument is called into question, however, when *moral* conflicts are the object of consideration. Let us assume that we are faced with a dilemma decision between two morally imperative options. If it is true that morality has its basis in emotions (and, as already stated, in the light of current empirical moral research this cannot seriously be refuted), then what makes the options *moral* options is the fact that we evaluate them emotionally. From this fact both options derive the moral weight they carry for us, and for our dilemma it is decisive which of the two has the greater weight in the given situation. In contrast, assuming a disengaged standpoint equates to blanking out the emotional evaluation, thereby not viewing the options *as moral ones*, with a reference to their respective *moral weight*. The decision that is then reached between the two is likewise not oriented toward this weight. We are then not judging *morally*. Herein is the error of believing that moral conflicts can be decided on the basis of general rules or criteria.

This point is significant, not least in conjunction with the opinion that an ‘ethics of rules’ way of thinking is the chief path toward innovation or progress in the field of morality. If it is true that morality has its foundations in the emotional evaluation of situations and courses of action, then morality cannot be influenced by this way of thinking. The latter may produce provocative theories, issues that are assigned a moral significance, and they may, as in the case of certain theories by Peter Singer, cause quite a public stir and arouse the impression that something is moving and changing. But these theories do not have the character of *moral*, emotionally

founded and evaluating judgments and insights, and therefore neither do they have the power to change existing moral attitudes and convictions. In debating whether we should follow a theory of this kind, diametrically opposed to our own moral insight, or whether the theory should follow our insight, we will decide in favor of the latter simply because the theory is not a moral insight.

2.3 With What Type of Reasons Is the Field of Morality Concerned?

Examination of the three arguments that appear to advocate viewing ethics as the rational justification of morality has thus led to the conclusion that this view cannot be deduced from the essence of morality, nor can it be substantiated from the fact that we sometimes err morally, nor can it be deduced from the idea that one of the tasks of ethics is the solving of moral conflicts. This view subjects moral reflection to a purpose that is alien to morality, one that raises the question of whether a moral reflection with this orientation can do justice to moral phenomena. A response to this question would require detailed analysis of the phenomenology of morality, and this would have to include the more recent philosophical debate about emotions and their significance for the understanding of morality,[5] as well as the findings from current empirical moral research.

The more recent philosophical debate on emotions has expounded the problem that morally relevant emotions are really *perceptions* with an affective content, i.e. things which possess a cognitive component and in which, therefore, *knowledge* is imparted. For example, an essential component of compassion is a perception of the suffering of a third party. Without this perception there would be no reference to the person *for whom* one is feeling compassion. On the other hand, the component thus felt is involved in this perception by rendering empathetically accessible what is being perceived, namely the suffering of the third party. Thus the affect is not simply a reaction to what is perceived, but an essential part of this perception as a factor that renders the perceived accessible and, to this extent, itself has a cognitive significance.

This sheds light on the significance of *narrativity* for moral reflection, particularly in the way that narratives, as distinct from pure descriptions, present moral issues for *emotional cognition* by clearly showing situations and actions in their experiential quality. Narrativity is thus the linguistic form of expression that corresponds to the emotional foundations of morality. The fact that narrativity generates moral knowledge needs to be stressed in the light of a misunderstanding, which has a long tradition traced back to some receptions of the ancient distinction between rhetoric and dialectic, namely that narratives are merely able to influence affects, and thus can have no validity as reasons and can contribute

[5] This debate is expounded in Ammann (2007).

nothing to moral cognition. If what is denoted by the word 'reason' has (at least in part and very essentially) anything to do with *reasoning* about situations and the significance attributed to them, then a description such as the example of the accident involving injured parties appeals to the *reason* and not merely one-sidedly to the sentiments of the listener. The listener can, of course, only reason appropriately if his emotions play a part. Yet what makes him realize that helping the injured is the right thing to do is not an affective reaction to the imagined situation, but a *grasping* of the situation, involving cognition and affect in equal measure. The *reason* for his judgment that it is right or imperative to help the injured is not located in this perceptual grasping: the latter is merely an *explanation* for this insight. Far more, it is located in the perceived situation itself: the injured parties find themselves in a state of emergency and are dependent upon help. This is the reason why it is right or imperative to help them. Thus the *justification* for this judgment exists in imagining this situation, which in turn occurs via its narration. To this extent it makes sense to speak of *narrative reasons*. Such reasons are basal to all moral cognitions because of their emotional foundations. Their character is completely different from that of arguments. When Prichard (1912) writes that our sense of obligation is *immediate* and deduced from nothing else, this captures precisely what is characteristic about reasons of this type, calling to mind the moral significance of situations and courses of action.

In the literature, a standard objection to narrativity as a method of accessing and forming moral judgments is the potential of narratives to manipulate (Düwell 2008: 52–54). In most cases, this objection focuses on major narratives, such as journalistic treatment of events or cinema films. The narrative justification of judgments, decisions or courses of action, however, is concerned with narratives of a different kind. Let us assume that a person does not turn up for work and the next day justifies his absence with the information: "Yesterday my wife was really ill, and I had to take her to hospital, then sort out a whole lot of things for her." This information is quite obviously a narrative, or a *tale* explaining the situation that caused the person in question to stay away from work and leaving no room for doubt that he had a cogent reason for doing so. This is the way in which we morally justify our actions in everyday communications. To what extent is this manipulative? In the worst case, one could indeed suspect the person in question of 'telling tales,' with his wife not being ill or needing to go to the hospital at all. This is a matter of facts, presenting an accurate *description* of the situation. In this way, narratives and the effect they have on us can be controlled using descriptively addressed reality. This to-ing and fro-ing between narration, on the one hand, and description of the factual aspects of situations and courses of action, on the other, also characterizes deliberation in ethical contexts, for example within the field of clinical ethics during case consultations.

As these deliberations are revealing, a study of the phenomenology of morality can deliver crucial insights into the type of reasons involved in moral reflection, and thus in ethics. These reasons are ones that, because they are directed at emotional cognition, at the same time have a *motivating* force. This distinguishes them from

the arguments that are involved in the thinking behind the ethics of rules.[6] Logical relationships and deductions cannot convey moral cognition. We can see this quite clearly from Peter Singer's argumentation in the abovementioned essay on global poverty. Singer states that, in the case of a drowning child, we would see it as our obligation to save the child, and he uses casuistry in an attempt to translate acknowledgement of this obligation to the situation of people living in poverty. With the example of the child, he appeals to our moral perception based on emotion. Here we find ourselves forced to acknowledge an obligation to help because we imagine the situation as if we were actually *experiencing* it. As a result, it has a normative effect on us. This would not be the case if we were to view the situation purely descriptively, i.e. from a disengaged standpoint. In contrast, in the casuistic comparison the two cases are regarded and compared with reference to their *descriptive* features. Consequently, the emotional and motivating effect that the imagined situation of a drowning child has on us is not translated to the situation of global poverty. Any coercion to acknowledge an obligation to help in the latter case is arrived at purely cognitively, via a comparison of the descriptive features of the two situations. To this extent, in the light of the situation concerning the people in poverty, Singer's argumentation does not convey any *moral*, emotionally evaluating, cognition regarding an obligation to help. Here we have an argumentative coercion to acknowledge such an obligation, but not an internally felt coercion to help, as in the example of the drowning child. Therefore, such argumentations are unable to achieve their desired aim, namely to *motivate* people to corresponding actions.

2.4 Ethics as a Guide to Right Moral Thinking

What conclusions may be drawn from what has been said concerning the purpose and task of ethics? Taking Prichard's (1912) deliberations as one's orientation, the task of ethics consists of being a guide to right moral thinking that does justice to moral phenomena. Instead of subjecting moral reflection to a purpose forced upon it from the outside, we have to question the extent to which *morality itself* demands reflection, as well as *which type of reflection* it demands. The irony of defining the

[6] It may be tempting to object that even arguments can have a motivating force. Anyone with a compulsive reason for an action thus has sufficient grounds to carry out that action. Indeed, he has sufficient grounds *as far as his reasons are concerned*, but this does not mean that he has sufficient grounds for actually carrying out the action. It is possible to have sufficient grounds for an action and still not carry it out, due to a lack of motivation or some inner resistance. *Moral* reasons are characterized by the fact that what is focused upon here as a reason. For example, "my wife was really ill" has a simultaneously motivating effect because the matter in question is experienced and evaluated emotionally (moral internalism). This implies that people whose emotional skills are restricted (e.g., because of brain damage) not only suffer from a lack of motivation, but are also incapable of having *moral reasons* (Fischer 2010). They may be able to think logically and subsume situations under rules, but they do not have at their disposal the skill of moral cognition, or cognition that is emotionally evaluating.

task of ethics as a guide to right moral thinking becomes clear if we compare it to theoretical disciplines such as mathematics or physics. Here we would not say that their task consists of being a guide to right mathematical or physical thinking. What we connect with these disciplines is the idea that a way of thinking is right when it grasps the mathematical or physical facts of the matter appropriately. Thus, the object of these disciplines is these facts themselves and not thinking about these facts (the task of providing a guide falls to teachers and lecturers).

This once again exposes the basic misunderstanding facing ethics in the modern age. The paradigm of the theoretical disciplines is translated to ethics as a practical discipline. Correspondingly, it is assumed that ethics is not concerned with our moral thinking, but far more with morality as something that is a pre-given for our thinking, and that is a yardstick for its rightness. This is closely linked to the phenomenon of objectification or disengagement, leading to experienced or narrated reality being replaced by the dualism of objectively given descriptive facts, on the one hand, and equally objectively given moral facts, on the other. The idea is then that moral expressions like 'right' or 'imperative' evaluate actions with regard to attributes they possess, such as the attribute of bringing about a certain effect in a given situation, rather than with regard to the moral significance which they themselves or the situations in question have.[7] The moral rightness or wrongness of an action then seems to be given with these attributes, and it seems to be of the same objectivity as these attributes. This is a misunderstanding because that which constitutes the sphere of morality is only open to an engaged attitude and a thinking process in which, as stated, emotions play a crucial role. Thus the idea, borrowed from the theoretical disciplines, of a distinguishability between objectively given facts and the subjective cognition of these facts cannot be translated to morality. As a practical discipline, ethics has the task of clarifying and guiding a thinking that

[7] In meta-ethics this leads into the debate about the relationship between moral and natural attributes. This debate reveals that an automatic consequence of the objectification of morality is a *naturalism*, in the sense that a moral value (right/wrong, good/bad, moral status, *thick moral concepts*) is ascribed to natural attributes. Accordingly, values supervene on constellations of natural attributes. What is questionable about this way of thinking is that it creates a sufficient condition from a necessary one. To compare: light waves must have a certain wavelength in order for us to perceive the colour 'red'. But redness is not already given with this wavelength of light independently of our perception. Likewise one can say that a specified behaviour needs to show particular natural attributes in order to be able to be experienced or perceived as cruel. But cruelty is not given with these attributes. It is independent of how we experience or perceive and thus *emotionally evaluate* behaviour. Emotional evaluation is directed toward experienced reality, as distinct from the objectified, descriptively broached reality to which talk of 'natural attributes' refers. This means that moral value, to the extent that it is based on emotional evaluation, is also not given independently of our experience. Therefore, it cannot be given with something that is obviously independent of it, namely with natural attributes. This does not lead us to conclude that moral value is "merely subjective." We can communicate intersubjectively about whether or not an action is cruel. We can do this because the word 'cruel' is the linguistic articulation of a (perceptional) pattern which we have jointly internalised via language. It refers, as we have said, to experienced reality, and we can recognise it in many individual actions. The *sufficient* condition for an action to be cruel is that it refreshes this pattern which, in turn, presupposes certain natural attributes as a *necessary* condition in order for actions to be able to be perceived within the framework of this pattern.

is engaged in a specific way, in which we as moral subjects are always automatically involved. It is not there to oblige us to another, disengaged, or in Prichard's terminology (1912): non-moral thinking, with the aim of rationally justifying morality, but to clarify this thinking.

But what can be used to measure its rightness? Following what has been said, there can be no hope here of an objective yardstick. According to Prichard (1912), moral thinking consists first and foremost, of an understanding of situations and circumstances in their moral significance. It is by this that we measure the rightness of actions. If someone has really understood what extreme poverty means, has he not at the same time also understood that people should not be subjected to such a circumstance, and that therefore there is a moral reason to protect them from this circumstance wherever possible? And if someone should doubt the latter, would we then not doubt whether he has really grasped what poverty means? The path to such reasons involved in moral questions leads via understanding. The ethicist can only try to visualize situations, moral phenomena, conflicts or problems in their complexity and moral significance, in the hope that this will also be helpful and transparent for the moral orientation of others. With regard to the question of moral rightness, however, there is nothing else to be said other than that the right thing is ultimately, after taking into consideration all the relevant aspects, what proves to be right according to the moral insight of those involved. For, as we have said, cognition of the morally right thing is immediate, meaning that the right thing is not beyond or independent of the emotional evaluation and insight of those involved, nor is it given with something else and deducible from this something.

This implies that the frequently expressed opinion deeming that this type of ethical thinking merely leads to a confirmation of existing moral views and prejudices is, in turn, itself a prejudice. Quite obviously, existing opinions can be shattered and corrected in this manner. For example, the prejudice that active euthanasia is always and in all circumstances morally wrong and contemptible is best countered by imagining the situations and circumstances of people with severe diseases who are actually calling for active euthanasia.[8] In the light of real people and real situations, moral condemnations prove to be abstractions that have little to do with the realities of life.

As this example shows, we should not falsely conclude from the fact that cognition of the morally right is immediate and not deduced from arguments that there are no ethical arguments. We can argue with examples, i.e. show another person something in which a moral state of affairs reveals itself, and with such forcefulness that the other person cannot escape it. This is true not only of moral judgments, but also of the clarification of moral concepts. Various philosophers, for example, have proposed viewing human dignity as the right not to be humiliated, i.e. violated in one's self-respect (e.g., Schaber 2003). Testing this definition with examples reveals it to be far too narrow. When Serbian troops drove their truck over Muslim prisoners in the Bosnian war, for example, we undoubtedly viewed this as a violation of the

[8] On how to treat and evaluate a moral problem like active euthanasia in the light of the concept of ethics represented here, cf. Fischer (2009).

prisoners' human rights. Yet can the manner of killing really be considered a violation of the self-respect of the Muslims? That would appear to belittle what actually happened. After all, the Muslims were brutally killed. The same can be said of reducing human dignity to the idea of autonomy, as found in the works of Kant. In our example, can the violation of human dignity really be limited to a disregard of the Muslims' autonomy? We can thus argue ethically, using narrative examples, in order to reappraise and review definitions of concepts. We can do this because moral concepts (e.g., cruelty, humiliation, human dignity, etc.) are the linguistic articulation of a reality experienced or narratively visualized, meaning that their content can only be expatiated upon via a back reference to this reality.

As these last comments make clear, the preceding deliberations would have been misunderstood had they been seen as a plea in favor of a "narrative ethics," more or less constituting the alternative to the criticized view of ethics as a rational justification of morality. Ethics does not take place narratively. The expression "narrative ethics" is therefore potentially misleading. It may be true that morality has a narrative structure, due to its foundations in emotion; and yet, as a critical reflection on morality, ethics thinks about what is revealed in the narratives forcing us to look at moral issues, in order then to capture these revelations in concepts and develop them into a coherent system of moral convictions.

In this endeavor, empirical knowledge on human morality is not detached from ethical thinking, but is actually an important element of it, as it allows us to understand mechanisms of moral insight people gain when experiencing specific situations. This knowledge can build trust in moral intuitions—and it also can uncover situational elements that may influence our intuitions in a way that allow people to reconsider their moral judgments. Actually, we can expect that an empirically informed ethics may even be better suited in guiding us to the right moral thinking compared to an ethics that purely understands itself as the rational justification of morality.

Acknowledgement This text has been translated by Sarah L. Kirkby, B.A. Hons.

Chapter 3
Moral Behavior and Moral Sentiments – On the Natural Basis for Moral Values

Adriano Naves de Brito

A central question of a naturalized ethics is whether moral values can be grounded in the natural traits and behaviors of the human species. In this paper I intend to answer this question positively. An exhaustive defense of this assertion demands, of course, more than I can offer in this paper. I can, nevertheless, offer a portion of what I think is necessary for this notion to be considered plausible and an outline of the way it may work. Thus I will be concerned here with revising, in favor of naturalism, the traditional contractualist concept of morality under which moral values are usually viewed. As examples of the way moral values can be explained in terms of human traits, I will use two of the paramount values of contemporary ethics: universalism and egalitarianism. My hypothesis is that moral values can be understood and their authority explained on the basis of how humans are naturally disposed to behave in groups. If successful, this will vindicate my claim for a program for naturalizing values.

The key concept connecting moral values with human behavior in groups is that of "moral sentiments," which are sentiments associated with moral evaluations. I will consider especially indignation, guilt and shame in order to explain universalism and egalitarianism. I will also treat morality as a system of reciprocal demands within which individuals are already contained, and in which features such as universality and equality are constituent components, since they are inherent to the way individuals affectively react to moral demands. Both the connection between judgment and sentiment and the assumed concept of morality must be explained and revised in relation to the traditional contractualist stance. I will dedicate the first three sections of this paper to these themes The contractualist view will be challenged by means of the following anthropological premise: in terms of evolution, to be part of a group is a major advantage, appreciation of which is deeply anchored in affective human dispositions and is not primarily a matter of reasoned

A. Naves de Brito (✉)
School of Humanities, Philosophy, Unisinos University,
Av. Unisinos, 950. 93022-000. São Leopoldo, Brazil
e-mail: brito@unisinos.br

M. Christen et al. (eds.), *Empirically Informed Ethics: Morality between Facts and Norms*, Library of Ethics and Applied Philosophy 32, DOI 10.1007/978-3-319-01369-5_3,

and justified choice. As a corollary to this, I will defend the position that morality, in this naturalistic sense, is crucial to keeping life in groups functional, that is, fit for cooperation, a notion that is compatible with an evolutionist approach to morality.

The fourth section will deal with the challenge of expounding the natural bases for universalism and egalitarianism. I will then underscore and explore the asymmetry between the scope of indignation and shame or guilt among groups. By connecting the anthropological premise with the affective asymmetry concerning moral sentiments, I will sketch out a naturalized way to understand universalism and egalitarianism. Whilst universalism will be depicted as the readiness to defend values beyond the limits of the groups to which individuals belong, egalitarianism will be characterized as the tendency to keep the importance of the members of these groups in balance.

3.1 On the Concept of Morality

From the point of view of Ethnology, morality has been conceived of as a system of reciprocal demands within a human group. This concept of morality has been used in the social sciences and in philosophy (cf. Rawls 1971; Tugendhat 1993; Scanlon 1998), and I take it as my starting point for this paper. There are two main facets to this definition: consideration of the group as the basic circle of morality and taking as the fulcrum of its operation the demands its members make on each other. Both aspects are, however, only vaguely determined in the definition I have put forward. Neither the size of the group, nor the way in which its size can be established, nor even the specific character of the reciprocal demands made within it have been defined, though all of these are essential to the purposes of this paper.

As regards the first question, i.e. the size of the human group within which a system of morality operates, a response might be included in the answer to the question concerning the naturalization of moral values, which is the main motivation for the current study. As such, that first question cannot be answered immediately, so a response will be provided in two stages in the second and third sections. The reason for the difficulty in giving a direct answer to this question lies in the fact that the determining factors, which delimit a human group governed by moral demands, are interwoven with the description of the operation of affections in terms of egalitarianism and universalism. Therefore, before giving a response to the question of the limits of human groups where morality plays a role it is necessary to better understand the dynamics of the affections within them, something which I will endeavor to explain below. The thesis concerning the question of whether boundaries of moral groups depend on affections is that the sociology of moral groups, unlike the sociology of society, cannot dispense with the dynamics of sentiments among group members.[1]

[1] The approach to morality I am trying to defend here is one that is based on groups and sentiments. An example of a group-based theory of morality is the society-centered moral theory, as defended

As regards the second question concerning the distinctive character of the mutual moral demands made within the scope of the group, affections play a leading role. Indeed, many social systems have reciprocal demands as one of their main elements, and it is therefore necessary to establish a criterion for distinguishing between these and other types of demands. One way of doing this is to consider the intentions with which the members of a system express their demands, and to analyze these intentions by means of the way in which member's affections operate in relation to these demands.

In aesthetic appreciation, for example, the demands of agreement concerning good taste are highly restricted, and this can be verified by considering the feelings involved. There is no point in becoming indignant with people simply because they do not share the same aesthetic views. Whoever does, has failed to understand the personal nature of aesthetic appreciation. Indeed, this appreciation should be considered as merely subjective, in the sense that one cannot demand of others (at least not to the point of becoming indignant) that they share the same view. Indignation is, in fact, symptomatic only of moral evaluations.

Playing games is another situation related to reciprocal social demands in which the strength of moral feelings is mitigated. Since participants always have the option to leave a game, the indignation of opponents is limited to the decision of a player to stay in the system or not. The same is not true of the system of morality as a whole where people are included whether they want to be or not. In the realm of morality[2] no one, be they a member of the group or not, is immune to the mutual evaluations that are made, nor are they immune to their own self-evaluations. For this reason, people may feel ashamed if their self-evaluations are negative, or feel enhanced self-respect if their self-evaluations are positive.

Thus, the main characteristics of the type of reciprocal demands that exist in a system of morality are that they are not merely subjective and they take the form of obligations. The demand for the fulfillment of these obligations is based not only on moral feelings such as indignation, but also on feelings such as guilt and shame (when viewed from the perspective of the individual rather than from that of the other members of the group).

One may ask where members of the group acquire the legitimacy to make the reciprocal demands (and self-demands), which possess the characteristics peculiar to a system of morality. What is the source of their justification to objectively insist that all members behave as demanded? What is the basis of the obligations, which are typical of moral systems?

---

by David Copp in *Morality, Normativity and Society* (Copp 1995). However, since families and other small groups are not considered to be societies (Copp 1995: 124–128), and since small groups are crucial for understanding the dynamics of the affections determining moral communities, "society" is not a helpful concept in the present case. This goes against Copp's theory in a number of relevant aspects.

[2] The difficulties in establishing the size of groups governed by moral demands (as discussed above) are of significance here. As we shall see, in terms of indignation, moral demands tend to be directed to anyone, regardless of group membership. This will be one of the themes of the fourth section of this paper.

The way in which a moral system may become legitimate varies significantly both in theory and in practice. Let us focus on two of the most frequently posited sources of normativity in groups: divine authority and agreement. On the one hand, a system of morality may be viewed as being rooted in divinely inspired principles which give legitimacy to the whole and which are not questioned, as is the case with theological moral systems. Alternatively, a system of morality may be viewed as being based on the collective will of its members, either in terms of the legitimacy of the principles from which the moral norms of the system are derived or, at the highest level, in terms of the very existence of the system. As regards the first concept of legitimization, which has its basis in some authority that transcends the will of the participants in the system, the members of the group recognize their heteronomy *vis-à-vis* the established principles and, of course, the values of the system. In the second concept, the emphasis is on the autonomy of the participants in the group *vis-à-vis* its principles and moral values. This concept is typical of contractualism and is a direct descendant of humanism.

Moral systems whose basis for legitimacy is thought to be transcendental by their very nature run counter to the naturalization of values, and therefore I will not deal with them any further here. Moral systems whose basis for legitimacy is contractualism[3] are of great interest to the study of the naturalization of values. In these systems, however, individual will is caught in an embarrassing circularity. If, in accordance with this concept, the legitimacy of the system of morality is based, in the final analysis, on individual free will, then those involved must want[4] the principles and the values governing their system in order for them to be legitimate. Nevertheless, since the reciprocal demands typical of moral systems are obligatory, and the subjective will of the individual is a secondary consideration, members of a contractualist moral system must desire principles and values which override the will of the individual.

The typical alternative solution to this difficulty has been to resort to some way of justifying these principles that satisfies individual members even though it may not satisfy certain circumstantial wishes. There is thus a tendency to conceive of a qualified will as a basis for legitimization of the system, a "morally due will" (Tugendhat 1993) that differs from the mere circumstantial will of the group members and, in order to possess these qualities, is thereby justified. It is unclear how effectively this can resolve the problem once and for all since the normative (obligational) aspect of such a justified and morally due will (a will that one *should* have) would also demand a justification. As a result, in the immanent spirit of the contractualist concept, it would be necessary to resort to some form of will, thus causing once again an embarrassing circularity.

[3] Contractualism is used here in a broad sense where individuals are autonomous in terms of a particular moral agreement, which they accept because they believe it is sufficiently justified.

[4] What is meant by "want" is an object of dispute among contractualists. In Gauthier's contractarian view (Gauthier 1986), for example, the important issue is what people would rationally desire. In Scanlon's contractualism (Scanlon 1998), however, the issue is what people with the desire to justify their actions to others could reasonably want. (This point was made to me by D. Copp).

An alternative, of course, is to take rationality for granted and to base the "second order will" on it. The "first order will" would then be backed up by a rational will. In fact, in Gauthier, the demands of rationality are viewed as not needing justification. Similarly, Scanlon seems to view the idea that moral motivation is a matter of aiming to justify ourselves to others as a conceptual point, which does not need the same kind of justification as do the substantive demands of morality.[5] Rationality, however, is a hard notion for a naturalist to swallow (and so it should be!). The demand for a justification as grounds for legitimacy in the contractualist system of morality, in an attempt to provide an objective basis on which to place obligation, thus implies the circularity of will or, at least, the non-trivial and non-naturalistic (and certainly non-Humean, as I will stress below) assumption that reason is in charge.

The emphasis on justification does not solve the initial problem but creates another one—that of meshing the justification with the moral sentiments embedded in the moral evaluation. If, for example, a person shows indignation regarding the breaking of moral rules, this indignation must be justifiable in terms of the validity of these same rules in order to retain its legitimacy. Therefore, from a contractualist standpoint, moral sentiments must regulate themselves by means of the justifications of the rules of the moral system in question, and it is presupposed that in order to ensure that this happens, moral affections are to be determined by judgments[6] in a way that will be discussed in the next section.

The naturalization of values depends on whether these values can be based on the preference of individuals. In other words, values need to be based on individual's "first order" will rather than on a disputable second, and somehow autonomous (because of its rationality) level of will. The contractualist option fulfills this condition only in part. Eventually, it qualifies the fundamental will by submitting the affections (which are the expression of individual preferences) to a corrective process based on judgments and on reasons. In my view, this step puts at risk the program of naturalization of values, as it tends to establish a level of justification over and above the actual preferences of individual persons, and is, therefore, anathema to a naturalized point of view. At the end of the day, things would look very much the same as in the theological system, simply with the divine authority replaced by a rational authority. Once certain rational principles have been established, they legitimize the system without considering the individual will.[7] However, this claim is still somewhat premature, and in order to consolidate the argument we

[5] Again, this point was made to me in D. Copp's comments on this paper.

[6] This would be considered as the cognitivist aspect of contractualism.

[7] Michael Smith (2004) is certainly among those who think that the circularity problem may be avoided by claiming an a priori truth about morality, which is true because, of course, it is what is rationally demanded. The most significant champion of this position in my view is still, however, Kant, and as sound as he is, he himself accepts that the will determined by such an a priori truth is not of this world, since it would be impossible for us, as sensitive beings, to experience it (cf. Kant 1781: A547, B575). In this sense and contrary to the opinion of some naturalists (e.g., D. Copp), these truths cannot be taken into consideration in a naturalistic approach. Like the free will they determine, they are not of this world.

need to investigate in more detail the relationship between affections and judgments, especially in relation to the themes of this paper. This is the subject of the next section.

3.2 Affection and Judgments: A Critique of a Tradition[8]

In the contractualist tradition it is customary to place more emphasis on justifications than on feelings when it comes to explaining moral obligation. Rawls (1971), and later Gauthier (1986), Copp (1995) and Scanlon (1998), in addition to Habermas (1981) and Apel (1973–1976) set out the tradition by defending a concept of morality in which giving reasons is essential for the system of obligation to operate properly. Tugendhat (1993), whose practical philosophy is critical to the transcendentalism embedded in Habermas and Apel's position, but who moved closer and closer to the contractarian tradition (especially to Rawls and Gauthier), makes a statement that touches on the essence of the theme I intend to discuss in this section. The central point in Tugendhat's statement[9] is that affections must be based on value judgments, or they would be senseless. To put this in another way, sentiments —especially moral sentiments— must be based on some kind of adjudicative evaluation, or they run the risk of losing their intersubjective significance. Thus to follow moral validity is a matter not of being guided by one's own subjective sentiments, but it is a matter of being guided by intersubjective judgments whose basis is some form of justified reason. Therefore, Tugendhat gives us, even though inadvertently, the cue for a transcendental element to become part of his conception of morality, thus avoiding a deepening of his naturalism.

The problem raised by this view of a naturalized morality is as follows: does universalization in the practical realm depend entirely on justified moral judgments, and if so, can that justification be naturalized? At first sight, a link between the contractualist tradition and the above question is not obvious. However, I must insist on the following premise: it is because contractualism seeks to place the basis for moral obligations not on a divine theological authority but on rationally justified moral judgments, and also because it seeks to give to the justifications supporting moral judgments a universal validity that we can question the basis of this validity. Since this basis cannot be outside this world, it must somehow be built upon our natural dispositions. The answer to this question is negative, i.e., the justification of morality cannot be naturalized, at least not if the correlation

[8] An earlier version of the arguments in this section and in Sect. 3.4 appeared in De Brito (2008b).

[9] "For all affections, what Aristotle has shown is valid with authoritative clarity for the whole tradition (Rhetoric, Book 2; accessible at http://rhetoric.eserver.org/Aristotle; last accessed on January 3rd 2013), namely that affections are positive or negative feelings (pleasure or displeasure) which build their own sense on a judgment, more specifically, on a value judgment" (Tugendhat 1993: 20).

between justification and affection is asymmetrical in favor of the former, which is exactly the case with Tugendhat.

In addition to his contractualist roots, Tugendhat follows an Aristotelian tradition regarding the connection between affections and value judgments.[10] Aristotle[11] established that there is a direct relation between certain manifestations of affections and a belief in the corrective value of their corresponding judgments. Might it not be the case that, over and above this, Aristotle was also postulating that there is a certain asymmetry in this relation, in the sense that feelings should be directed by value judgments, but not vice versa?

There are two assertions here, one that says an adjustment between affections and judgments is required for the legitimacy of the latter, and a second that says the direction of adjustment between affections and judgments in the case of moral judgments is from judgments to affections and not the other way round. In the case of the first assertion, the interdependence between affection and judgment can be well-illustrated by the role of the belief in the acceptance of the judgment as a legitimate currency in a moral interaction, a point well made by Aristotle. We presuppose a direct connection between the affection and the belief in the corresponding value judgment,[12] and that this connection is real and essential for the legitimacy of the judgment. Without this connection the judgment loses its strength and changes its illocutionary force completely. The capacity for dissimulation (which is by no means only a human trait) is an excellent example (although a negative one) of this. Indeed, dissimulation only works if the individual concerned is able to convince her peers that her judgment has been made in good faith, i.e., that there exists a corresponding affective evaluation, which, if it were true, would give the judgment the authenticity it needs to be legitimate. In species that possess adequate development of the brain and lead complex social lives, individuals are capable of deceiving their peers by means of verbal or non-verbal signs professing beliefs they do not in fact hold and that do not correspond to their real affections, on which legitimate judgments are expected to be based.

As for the second point, if Aristotle, or anyone else for that matter, considers there to be a one-way link between value judgments and affections, and that the latter are directed by the former (and for the purposes of this discussion it is possible to restrict this affirmation to the dimension of morality), then it must be admitted that the basis of the moral distinction between what is good and bad is not to be found in connection with affections *prima facie*, but originally in connection with reasoning. As such, this basis is not to be found in the sentiments, but in

[10] In a previous article (Brito 2008a), I dealt with this theme more extensively in connection with Tugendhat's thought.

[11] For Aristotle's assertion *apud* Tugendhat, see footnote 10.

[12] It is probable that the reason for the close link between affections and beliefs resides in the fact that their sources may be nearly the same, i.e., their origins are both in the limbic mechanism. This would favor an essentially Humean reading of human nature (Hume 1739–40), which would bring together practical and theoretical reason.

understanding.[13] Hume (1739–40) clearly showed that emotional approval or disapproval is essential to moral appreciation, and that from this perspective, the discussion of the primacy of reason over affections in morality tends to be frivolous (Hume 1751, section 1).[14] Could it be that Aristotle conceived things in these terms by giving understanding precedence over affections? I personally do not think so. My thesis, however, is not exegetical. It is related, above all, to a question of fact, and I can therefore leave open the interpretation of Aristotle. What concerns me here is to determine what is at stake if in morality there is an asymmetry between affections and value judgments in favor of the latter.

On the one hand, it seems obvious that we judge that something is bad because we consider it to be bad. On the other, it does not necessarily follow that if what is evaluated is considered bad, it is objectively bad. It may be that it is not appreciated; it may be that it is not liked. Viewed from this perspective, value judgments are judgments concerning how the world and its trappings affect us. Therefore, we can certainly affirm that there is a link between affections and judgments. However, this link appears to be both direct and symmetrical.

The philosophical consequence of interpreting the relationships between affections and judgments in this way is that value judgments become fundamentally subjective and are the result of individual preferences. Yet, whilst morality implies mutual constraints caused by the establishment of obligations, the justifications relating to other demands that each individual makes have a fundamentally subjective basis. It is clear that a basis that is not merely subjective for the mutual constraints is missing here. If, however, moral affections could only make sense if they relate to adjudicative appreciation, then it would be the value judgments on which these affections are based that would need substance, and it would thus be necessary to identify objective bases for the validity of the judgments and not for the affections. This is exactly how the contractarian tradition proceeds.

Whereas affections are merely subjective entities, it is supposed that judgments can only be justified objectively. In terms of justification, judgments have a clear advantage over affections. Judgments inhabit the universe of discourse, and it is in this universe that justifications are acceptable, since it is there that reasons can be given in an articulate and coherent way, something that is only possible in a

[13] Developments in the cognitive sciences have caused changes in our way of understanding the cerebral processes involved in making judgments in general and moral judgments in particular. In a recent study in the field of empirical ethics (Nichols and Knobe 2007), concerning the connection between the assessment of moral responsibility and the affections, one hypothesis to be considered is that the affections should not be held responsible only for deviations in judgments, but that they create the conditions for these to take place, and as such are central to the act of judgment. As no firm conclusions have yet been drawn in this case, I will use the vocabulary of the modern philosophical tradition in order to express the difference between a decision taken because of reasons, and one taken because of affective inclinations.

[14] An interesting discussion on the topic of the relationship between emotion and moral judgments was undertaken by Jesse Prinz (2007a) while defending his emotionism. His defense is interesting largely because of the fact that it is based on experimental psychology, which gives epistemic substance to the debate. He seems, however, at least in the first part of the book, to favor emotions in the dispute, whereas I personally believe symmetry should be preferred.

rule-bound system. Thus, all that is required is to find a principle, a criterion, a basis, or whatever is able to sustain the rationality implicit in discourse.

Having affirmed this, I can then conclude that the following is at stake in a substantial part of the contractarian tradition of moral philosophy concerning the relation between affections and value judgments: to satisfy the legitimate desire for a morality with universalist foundations. Indeed, who could deny that this is a legitimate wish? And who would be prepared to give up their claim for the validity of their moral evaluations in such a careless fashion? However, there is a caveat in this affirmation. The tacit supposition of the tradition concerned is that, as regards the question of foundations, justification belongs in the sphere of discourse. The agenda for the universalization of morality necessarily involves the justification of the objective validity of moral judgments. In the descending direction, which moves from justification to morality, once the adjudicative evaluation has been corrected by means of a discursive (and therefore rational) principle, affections can be assessed accordingly. Therefore, if we assume that affections can no longer be the basis for the correction of value judgments, then this tradition must accept a moral foundation that cannot belong to the sentiments. According to this scenario, the cost of universalization in the contractualist tradition is a form of practical rationalism, in which Kant (1785, 1788) is the dominant and paradigmatic figure. For Kant, and for all those philosophers who share his view (which is also a contractualist one), universally value-related distinctions can only be established by reason. All other values are based on sentiments and are therefore fundamentally subjective.

Although it may happen in an indirect and secular way, contractualism tends towards the construction of a transcendental level of moral fundamentals. It is thus possible to confirm the suspicion raised at the end of the previous section that recourse to a qualified will as a basis for the legitimacy of obligation in a moral system is fatal to the objective of naturalization of values and, therefore, to the naturalization of morality.

The question that arises as a result of this discussion is as follows: is there an alternative to the universalization of moral premises which does not simply become a variant form of Kantism or of contemporary contractualism? If there is, it will be an authentically naturalistic alternative, since it would have to spring from a denial of the asymmetry between value judgments and affections, and would have to explain the normative authority of moral values on purely immanent bases. In other words, it would have to recur to the economy of affections, in the light of the network of evaluations which characterize moral systems. It might even be possible to consider an alternative of a contractarian type, but there would be no place in it for a higher order will and, therefore, no place for the autonomy of the individual in the sense that rationalism gives to the concept.

What I hope to do in the fourth section of this paper is to make a plausible sketch, in terms of the affective-judgmental nature of human beings, of the dynamics of moral values by means of universalism and egalitarianism. First of all, however, we need to return to the concept of morality and deal with the problem of the limits of moral communities. The discussion of this issue, which was left open in the

previous section, can now be taken up again, with two dividends: the relative value of autonomy and of the freedom of the individual in the realm of morality, and the alignment of morality with the evolution of the human species.

3.3 Different Scopes of Morality

What are the limits of a group governed by moral commandments? How do human beings behave in relation to this group and its limits? If morality is a system of reciprocal demands within a human group, then these questions are relevant to our understanding of the moral human phenomenon.

It seems somewhat intuitive to think about the limits of a human group on the basis of the agreement of its members. As such, the group would become larger as the number of its members increased. This model is convenient and possibly appropriate for a group based on agreements that may be tacit or not and are adjusted to fit the circumstances, and for a moral group emanating from a transcendental authority of a divine-moral nature. On the one hand, it is obvious that this method of delimiting groups by signing a contract (or by adhering to a faith) exists in practice, but on the other hand it does not unilaterally determine the limits of a moral group. The fact is that, given the compulsory nature of moral demands, something that is clearly demonstrated by the analysis of the intention of utterances used to express moral judgments and that contrasts with that of other value judgments (for example, those related to aesthetics or play), everybody is, in one way or another, included in the group one may make moral demands of, whether they want to be included in this group or not. We can speak here, at this very broad level in order to continue using the term "group" for quite small communities of a system of morality. Thus, *lato sensu*, the system of morality tends to include the whole human species. I will return to this issue in the next section when I discuss the affective bases of universalism.

It is nevertheless clear that the system of morality, in its widest sense, is not homogeneous and that differing and often conflicting values co-exist within it. I will also return to this subject later when I discuss egalitarianism. Within the all-embracing sphere of morality, therefore, there exists a myriad of smaller moral circles, in relation to which there is no uniform way of belonging. These are moral groups that are interwoven with each other within the system of morality, which is definitely just a useful conceptual abstraction. With regard to membership of these inner circles of the system of morality, i.e., the moral groups, there are two related phenomena that need to be distinguished from each other. The first is the phenomenon of voluntary membership of a group –through affiliation to a faith by conversion or by reflective consent to a convention. Reflective consent is a form of membership that is typical of groups and is characterized by the desire to react affectively to oneself and to other people and to break norms, contracts and rules of the system, but which stems from the principle of consent. In such circumstances, the autonomy and freedom of members are significant. These circumstances, however, do not demonstrate how human beings originally became part of moral groups.

This brings us to the second phenomenon, that of the original belonging to groups governed by moral demands. In terms of origins, that is, in terms of the conditions in which human family groups (*lato sensu*) were constituted as hunter-gatherer groups, and the conditions in which the human species has evolved, the state of insecurity in which individuals existed at that time made it impossible for such groups to be formed as a result of choices or previous agreements, tacit or not, via the strengthening of affective ties. In such conditions, the economy of affections is a determining factor and is responsible for defining the limits of a particular moral group.

By referring to the two phenomena above, I wish to emphasize that their limits are very flexible, but at the same time, they are fixed in a varying manner. The limits of groups formed by tacit or explicit agreements depend largely on established conventions. Nevertheless, as far as original human conditions are concerned, the limits of groups formed under these conditions did not depend on conventions, but on the dynamics of the moral affections involved. Affections were the most original means by which human groups were defined and maintained, and this is something that was inherited from the evolution of mammals.

Therefore, when we consider the differences between the ways moral groups are formed and maintained it is possible to speak in terms of scopes of morality, that is, more or less well-defined areas within which individuals can expect their moral demands to have value, from the small family group to the overall system of morality which encompasses everyone. In addition, if we take morality in this latter, wider sense, humanity constitutes a single moral community.[15] This means that there will be moral demands that are restricted in scope, but that there will also be moral demands that are made by the human species as a whole.[16] The narrowest scope corresponds to family circles *lato sensu*. In terms of the evolutionary development of the human species, the family (groups of close relatives) is the most basic unit of the system as a whole. The widest scope, as a result of the connections between all the more basic moral circles, therefore corresponds to the whole human species. In this sense, *the* system of morality is a system that consists of smaller moral circles. These inner circles of morality mesh with each other within the moral mosaic right up to the very highest level, which is the whole of humankind.

To sum up, morality is a system of reciprocal demands that individuals are part of whether they want to be or not. It entails obligations (rules) that have to be followed by all the members of the system, and the fact that they obey them is sustained by moral sentiments. If we look at this from the perspective of the inner circles of morality, i.e. the moral groups, individuals are ready to obey the obligations of the group if they feel themselves to be part of it. This is the case if, for instance, they feel ashamed of themselves for not corresponding to the group's demands. So unless they do what

[15] This is a community which could encompass animals as well, in line with such writers as P. Singer (1975/2002). The notion of a world community is, as I have already noted above, a useful conceptual abstraction with no real emotional reality. We are, indeed, emotionally limited creatures, who could, however, despite our many flaws, construct an impressive patchwork of moral groups.

[16] A paradigmatic example of demands placed on humankind would be Human Rights.

the group demands, they will be ashamed of themselves, or they will feel guilty about their transgression. In other words, as far as individuals are concerned, belonging to a particular circle of morality is defined by their readiness to react to a breach of the norms, whether committed by themselves or by others who have moral sentiments, through, for example, indignation, guilt or shame. This readiness, of course, does not depend on voluntary agreement and does not seem to be ensured by having affective ties to other members of a group, which is a sign that we "belong" to moral groups that we have not tacitly chosen to belong to. We may even emotionally disagree with those groups to which we have the closest ties.

What is missing from contemporary moral naturalism is a comprehensive explanation of values and the potential for conflicts between the values of the restricted scope of morality and the values from the wider scope of morality. Obviously, the problem lies in the difficulty of reconciling the interests of restricted groups with those of other human beings and, in particular, of societies[17] (political units which are less inclusive and abstract than "humanity"). Indeed, when it comes to forming values on the basis of feelings, the range seems to be limited. The link between the values in the small circle of family groups, *lato sensu*, and their validity for the moral system in general and in societies and humanity requires an explanation, which, in line with the naturalist agenda, must be rooted in affections. The explanation of how moral values compete for validity ultimately provides the elements that are needed to describe the moral bases of society. Following on from this, what needs to be clarified is how moral values forged in the different affective circles are intertwined within the society and the system of morality.

In the remaining part of this paper, I will try to explain this link with reference to the dynamics of affections. There is a premise in this explanation that is essential in linking the individual and society in terms of morality. My starting point is that a fundamental inclination of individuals is to avoid being excluded from the community to which they feel they belong, and that this premise is perfectly compatible with a naturalistic view of morality.[18] Moreover, the premise is compatible with our evolutionary history regarding our extreme individual vulnerability.

3.4 Naturalizing Universalism and Egalitarianism

As has already been pointed out above, one of the most serious problems with a system of morality based on sentiments is that the validity of the values shared by its members cannot be universalized, since sentiments are subjective or at least

[17] Society in the sense in which D. Copp (1995, Chapter 7) defines it.

[18] Empirical research on empathy has helped to explain the role of this emotion-related trait of our species in morality (e.g. Batson 1991 and Hoffman 2000), and neuroscience has shown how this is so. For a species which has a group of neurons which is specialized in emulating the sensations of others, as well as motor events, the so-called mirror neurons (cf. Rizzolatti and Fabbri-Destro 2010 and Ferrari et al. 2003), which we use to live in groups, to live in groups cannot be an option, but is more like an evolutionary karma.

parochial. To re-state the problem, is this assertion a negation of any attempt to derive from sentiments a certain universalism in morality?

It is questionable that there are universal values, or to put it in terms of the problem of normativity, that there are values whose validity is recognized by everyone. The problem can, of course, be formulated in a less census-related way so that universal values can be understood in the sense of values that oblige everyone so that they *should* act in accordance with them. The issue here, as already discussed above, is either dogmatism or circularity. Since one can always ask why these values should be obeyed, a non-naturalistic answer (in the strict sense I have been trying to give here to the term "naturalism") would be one of the previously discussed alternatives, i.e., either an authority-based morality, or some version of Kant's rationalism. In fact, a great (perhaps too great) philosophical effort has been made to demonstrate that some values are universals in their occurrence and scope, that they should be obeyed by everyone, and that everyone should be taken into account. Considering that one cannot demonstrate this by means of an empirical investigation (since conflicts of values are the rule rather than the exception), the fulfillment of the task of proving the universality of values entails all too frequently the defense and acceptance of an authority, be it a divine theological one or a rational one (e.g., a general principle).

The census-related formulation is not compromised by either of these strong presuppositions, which, as I have stated above, run against a blunt naturalistic account. It is therefore preferable to follow this formulation, so that a naturalized approach to morality should be concerned with a general demand with respect to values (and to the rules leading to them) and not with the purpose of a universalized system of values. The former is the basis for normativity, and the latter is the objective of a normative theory of value, or of a political process. What is at stake here is that, although there may be no universal value in a census-related sense, it is a fact that individuals and groups are prone to treat their values *as if* they were universal, or as if they should be seen as such. How does this come about? How can we deal with this fact in merely naturalistic terms? In this sense, the purpose here is not normative in itself, but rather descriptive.

At this point, some terminological and metaphilosophical considerations should be pondered. First of all, it is important to draw a distinction between universalism and egalitarianism in relation to a theory of value. While the former is a quality of values, the latter is a value in itself. The former refers to the scope of validity of the values shared by individuals, and the latter is an important value among many different groups and a paramount value in western societies. The problem concerning the naturalization of universalism is, therefore, related to the search for the natural basis of normativity in morality, while the problem concerning egalitarianism is related to the natural basis for considering equality (or equity, at least) as a moral value. The former is a much more general problem, to which the assessment of a value in terms of good or bad is not relevant, whilst the latter is precisely about making such an assessment about a specific value. In order to give a naturalistic account of universalism, I will hereafter be concerned with the affective basis of normativity, and as regards egalitarianism, I will try to identify the affective basis necessary for it to be a common value among human groups.

Normativity implies demanding from others respect of values and rules of behavior, where "others" remains as an open concept whose meaning extends from the circle of the family to the whole of humankind. Notwithstanding this, it also implies that the individuals involved recognize these values as such, i.e. as demands on them and, therefore, as possible grounds for changing their behavior. There can be no room for normativity among groups whose demands are not even understood as demands. It should be noted at this level that whether these demands are legitimate or not is not a concern at this juncture. The readiness to make changes in behavior because of the demands of others, and not because of justification, is what is at stake at this basic level.[19] The moment of assessment depends on the prior recognition that there is something requiring such assessment.

From the perspective of nature alone, the only basis for normativity is the will of the individuals involved. I accept this starting point (which is also a basic contractarian assumption) and thus I also accept that whether a particular norm will take effect or not depends on the will of the individual. However, whilst there can be no inexorable reason for an individual to endorse a norm or a system of norms, she is not free to react or not react affectively to moral demands, and this is decisive both for morality as a system of reciprocal demands and for the dynamics of values within groups. The kind of affective reaction implied here has evolved along with strategies for keeping groups together and for making them functional.[20]

What I mean by "functional" can be understood in the sense in which the word is used in Ethology. In this usage, it has a teleological character, since a particular behavior serves a particular end, although this end is, in fact, purely immanent to the species. As a consequence of this, to ask about the function of a behavior is to ask what benefits that behavior can bring to the species. A functional behavior is, therefore, one that provides the species with some adaptive advantage.[21] To say that a group is functional in a relevant moral sense means that it is capable of cooperative

[19] It is clear, therefore, that morality, in the sense in which I am trying to define it here, should not to be reduced to a code, and not even to a justified code, but is tantamount to an effective system of behavior control. Contractualism is very much oriented towards justification, since it is concerned with the rational legitimacy of moral demands made by propositions. The authors I mentioned in Sect. 3.2 (Scanlon, Gauthier, Rawls, Copp and Tugendhat, Habermas and Apel) are all good examples of this characterization.

[20] A. Gibbard (1992: 61–68) develops a similar idea and talks about coordination of emotions from the point of view of evolution and of game theory. Moreover, he considers that the internalization of norms, which for him is something of a linguistic character, is biologically connected with the coordination of emotions, since these have a motivational character (cf. pp. 68–71). He, however, with regard to humans, establishes a direct connection between language and morality that I do not share in my approach.

[21] Primatologists, for instance, in describing chimpanzees' response to inequity, are interested in the function of this behavior for the species, and they consider the hypotheses that it may increase the payoffs of cooperation: "Evidence is beginning to emerge to support the hypothesis that cooperation and the response to inequity are linked in species besides humans. Such a link may provide evidence for how the response evolved. This question is more than academic. Understanding the evolutionary trajectory of a behavior can help elucidate its evolutionary function." (Brosnan 2011: 3).

behavior, which is crucial for humans.[22] What I am suggesting, therefore, in line with an extensive bibliography concerning cooperative behavior in humans and primates (cf. for instance, Boehm 1999; Brosnan and De Waal 2003; Brosnan et al. 2009b, 2010a; Bekoff 2001), is that some affective reactions have evolved as a result of the adaptation to life in groups and under the pressure of reciprocal demands, and individuals are not in full control of them.

In this sense, the general tendency to demand respect for the values the group shares is not a matter of individual choice, but is an essential element of human behavior that should be included in any satisfactory account of morality seen as a human characteristic. My strategy for explaining the trend towards universalism in moral systems will be to analyze how indignation interacts with guilt or shame when individuals are part of the whole system of morality while also being part of some of the inner circles of that system, namely, the groups to which they most intimately belong.

In relation to this, the asymmetry between moral sentiments is a notable trait and can provide the basis for that explanation. It is also interesting to note that indignation, asymmetrically related to shame and guilt, does not recognize any boundaries between the inner circle of morality (the family circle, for instance) and all the other groups, or even humanity as a whole. This means that one cannot prevent others from becoming morally angry with oneself by drawing lines separating one's own group from others. Whenever two groups interact in any way, indignation can occur on either or on both sides. In this sense, we cannot simply be *indifferent* to the behavior of others if they are capable of showing moral sentiments just as well as we do and if their behavior affects us. Indignation is, therefore, a moral sentiment that is potentially directed to any person, and is hence universally applicable. Consequently, by means of indignation, normativity tends inexorably towards universalization, not because of a principle, but because of our fundamental disposition as beings capable of having moral sentiments.[23] Since the approach here is naturalistic, I cannot *prove* the existence of universal values, which would be typical of a rationalist point of view. However, this is not necessary. The whole point here is that the trend towards universalization of any moral value rests on the limited freedom of preference that governs individual moral reactions, something that is at the very heart of the way indignation works. Indignation is an affection, which is a sign of

[22] For an extensive discussion of the evolution of morality in conjunction with cooperation, see Ridley (1996).

[23] Strawson's reply to pessimism (Strawson 1974/2008) is, as he himself describes it, in many senses based on commonplaces. Pointing out "reactive-attitudes", should show to the pessimistic opponent how our life is determined by reactions we cannot fully control. My remark about the link between indignation and a universalistically oriented normativity is similarly a commonplace. Nevertheless, an important part of the philosophical task is to prevent philosophers from forgetting the world they are living in when they are doing their work. As Strawson puts it: "The object of these commonplaces is to try to keep before our minds something it is easy to forget when we are engaged in philosophy, especially in our cool, contemporary style, viz. what it is actually like to be involved in ordinary inter-personal relationships, ranging from the most intimate to the most casual." (Strawson 1974/2008: 7)

the intention that the members of a particular moral group give to their demands to each other and to others outside their group. This is an intention they are not free to discard, since it is connected with their deep feelings of belonging to the group. The same pattern of analysis may also work with another intuitive moral value, i.e. egalitarianism.

In order to plead the case for egalitarianism, I must turn again to the asymmetry between indignation and shame or guilt. While indignation can traverse any boundaries existing between moral communities, shame and guilt are the sentiments that define membership of a group. By sharing with others the same grounds for feeling ashamed or guilty, an individual can be considered a member of a particular moral group. It is important to reiterate that moral groups are the only subjects of analysis here. Groups based either on circumstantial agreements or on tyrannical force are not prima facie under consideration in this paper, since they do not rely on moral sentiments and therefore are not moral communities in the sense defined here. In these groups, individuals may have either contractual duties or fear as binding ties, but shame or guilt is only encountered therein as a misplaced sentiment.

Nevertheless, no matter how atavic the sentiment for belonging to a group may be, human beings are not as hard wired as many other animal species. The complexity of the dynamic of their affections is proportional to the complexity of their nervous system. The individual's ties to the group must be constantly nurtured. They may not be asked to give their acceptance to be part of the family they are born into, but they are certainly expected to the give their affective consent in order to retain their membership after they become adults. An adult can dissent and leave the group if her ties with it are broken. Since the same condition applies to each and every member of the group, and since the ties between its members are of an affective nature, each individual should be treated as if she was important to the group, and this importance (or, indeed, unimportance) is affective. In this sense, everybody in the group should be treated as equally valuable, and this is what the concept of respect means.

Breaking away from the group demands, without any doubt, is more than an episodic act of will or a rational decision. It is, rather, a process of keeping one's distance from the group until the feeling of belonging disappears. Individuals are most certainly interested in nurturing their ties with the group they are attached to, but the reverse side of this is that the group must encourage the respect of each of its members for each other, and its interest in steadily renewing its alliance with its members. God's alliance with His people, which is the basis of a number of religions, can be understood in the same specific sense. A celebration of this alliance is always a celebration of reciprocal respect.

Seen from this affective perspective, egalitarianism is presented as a condition for the existence of functional groups composed of members with complex nervous systems, such as mammals. As the tendency to universalize values is the naturalized version of universalism, the tendency to balance the affective importance of the group members to each other is a naturalized version of egalitarianism.

I can now put all of this together with the previous discussion of the nature of morality. The naturalization of universalism and egalitarianism has shown that they

are conditions for the functionality of the system of morality. Moreover, it has highlighted a feature of morality that is not taken into account in traditional approaches, i.e., that morality must have a functional character for humans in terms of facilitating cooperation. Traditionally, too much weight has been placed on the justification of moral principles without taking into account the fact that the main source of this justification is the functionality that morality provides for life in groups. When separated from this functional basis, the principles have to be justified by abstract authorities who are irreconcilable with the affective and social characteristics of the species and with its evolutionary history. From the standpoint of nature, morality is functional, and the values forged and maintained in the groups governed by moral constrains are components of the system that operate in favor of its functionality.

The fact that morality has both a universalistic and egalitarian characteristic within the substance of its structure does not imply that individuals have to (or even should) adopt a universalistic or egalitarian attitude toward others. There can be no strong normative outcome in this case. My account implies, however, that morality tends to move steadily towards both universality and equality, and that humans tend to behave according to universalistic and egalitarian values by means of the sentiments on which the groups are based. Whenever a limit is imposed on moral constraints and whenever there is an imbalance between individuals, there will be tectonic pressures and seismic movements to restore on the grounds of morality and the balance of the forces, which are at the heart of the system. Conflicts between and within human groups bear witness to this fact time and time again.

3.5 Some Conclusions

It is worth noting certain outcomes of the above analysis of morality and values in a naturalized perspective. In order to account for our moral intuitions, which give precedence to universality and equality, it is not necessary to resort to any plausible defense of a rational concept of morality since these characteristics belong to the innermost structure of morality as a functional tool for improving life and cooperation in groups. A normative approach to morality is certainly an urgent philosophical task. To make our way through the labyrinth to the place where morals converge, as is inevitable in a highly interactive society such as the one in which we currently live, we need rules that can only be arrived at through a discussion of reasons on which each side has to agree. Notwithstanding this fact, it is also an extremely important philosophical task to understand what the moral basis of an agreement that might be reached would actually look like. The above analysis makes at least three contributions to the fulfillment of that task. The first is that morality is to be seen primarily as functional. By highlighting this aspect of morality and its values, they can finally be understood within a naturalistic approach. The second is an explanation of values in terms of preferences, affections and others human traits. The recognition of the asymmetry between indignation and shame or guilt, which is

fundamental to morality as a system of reciprocal demands, but which does not characterize specific morals based on agreements only, is the key element of this analysis. The third is that universality and equality are to be defended in any tolerable human concept of morality, simply because they are constituent elements of human morality, and not fundamentally because it is rationally plausible to choose them.

As for the problem concerning the conflicts between values from the restricted scope of morality and values from the wider scope of morality, to which I referred at the end of the third part of this paper, it remains an open question, though a naturalized view on morality seems to be indispensable to finding a satisfactory answer.

These outcomes are compatible both with our common intuition concerning moral values, and with what science has already discovered about human development from its most primitive stages onward. I consider these outcomes as relevant criteria (though they are certainly not the only ones) for evaluating a particular philosophical hypothesis. At any rate, it will be interesting to continue submitting to scientific scrutiny (paleoanthropology and paleoethnology, for instance) the anthropological premise whereby I have assumed that human beings do not demand benefits to enter a moral group, but that life in a hunter-gatherer group governed by moral constraints is rather a benefit in itself, since it is an essential condition for human existence in terms of its status as an evolutionary advantage. It is in order to keep that benefit available that morality, as a system of reciprocal demands governed by values and sustained by affective predispositions, is indispensable. After all, the interest in, and the necessity of, belonging to a moral group is the invisible force that keeps our moral compass pointing in the required direction.

Acknowledgement This paper has been prepared with the support of CNPq. I am also greatly indebted to D. Copp for having generously read an earlier version of this text and having made a number of apposite remarks.

Part II
Investigating Origins of Morality

Chapter 4
Morality as a Biological Adaptation – An Evolutionary Model Based on the Lifestyle of Human Foragers

Carel van Schaik, Judith M. Burkart, Adrian V. Jaeggi, and Claudia Rudolf von Rohr

4.1 Introduction

4.1.1 Clarifications

A biologist studying the behavior of a species, when confronted with a seemingly costly behavior that is highly persistent in that species, would certainly entertain as her null hypothesis that this behavior is adaptive, and that the psychological mechanisms underlying it were therefore adaptations put in place by some form of natural selection. The aim of this chapter is to develop the outline of such an adaptive hypothesis (Alexander 1987) for the evolution in humans of our moral psychology, the moral emotions of which it consists, and the moral behaviors it produces. Briefly, we will propose that moral emotions are the subjective side of the proximate rules (motivations) that regulate human cooperation, which in turn is an evolutionarily novel adaptation to enable the uniquely derived lifestyle of human foragers, which requires generosity and sharing due to extreme mutual interdependence.

This biologically oriented approach offers a radical departure from the traditional way many philosophers (e.g., Kant 1785/2002) have thought about morality as being rooted in rational reflection. The adaptive approach has gained much traction during the past decade due to work showing that moral actions are often based on snap decisions guided by moral intuitions that cannot be articulated (Haidt 2008) and that at least some of these moral intuitions may have an innate basis (Tomasello

C. van Schaik (✉) • J.M. Burkart • C. Rudolf von Rohr
Anthropological Institute and Museum, University of Zurich,
Winterthurerstrasse 190, 8057 Zurich, Switzerland
e-mail: vschaik@aim.uzh.ch

A.V. Jaeggi
Department of Anthropology, University of California Santa Barbara,
Santa Barbara, CA, USA

M. Christen et al. (eds.), *Empirically Informed Ethics: Morality between Facts and Norms*, Library of Ethics and Applied Philosophy 32, DOI 10.1007/978-3-319-01369-5_4,

2009; Hamlin et al. 2007). Of course, it is not truly novel, going back to the moral-sense theory of the Scottish Enlightenment philosophers, such as Adam Smith and David Hume (Monroe et al. 2009), but the latter obviously did not think of the sentiments as products of evolution through natural selection.

Before developing this hypothesis, we must clarify our terminology. Behavioral biologists distinguish between fundamentally different ways of answering questions about the causes of any behavioral phenomenon, often referred to as proximate and ultimate causes (Tinbergen 1963). The common way to explain the behavior's occurrence is to show the direct causal mechanisms that bring it about: the so-called proximate control or regulation. The focus is on motivations, hypothetical variables that affect the probability that a particular behavior is produced in the presence of eliciting stimuli. Their existence was postulated by the classical ethologists to account for variation in behavior among individuals or within individuals over time in the absence of changes in the state of the external world (Tinbergen 1951). Hunger, fear, sexual desire, etc. are examples of motivations that can be readily defined and measured in animals. Motivations can also be described using neuro-endocrine and other physiological variables, although as yet there is no exact one-to-one correspondence between these levels of description.

There is a second dimension to the proximate mechanisms, uniquely accessible in humans and therefore avoided by behavioral biologists studying animals. It is the level of experienced motives, the feelings accompanying high motivation. An outside observer may experimentally establish that an individual has a high feeding motivation (measured as a high probability of engaging in feeding behavior under specified conditions), but we humans experience this motivation subjectively as hunger and reducing hunger as psychologically rewarding. And, solipsism aside, we know that other humans have very similar subjective experiences. We will refer to this as the subjective dimension of proximate causation, often described in terms of emotions. Emotions can be defined as subjectively experienced intense mental states; they are accompanied by characteristic physiological changes (Frijda 1986). In the case of morality, the subjective dimension of proximate causation includes such emotions, but also the intentions (moral preferences) underlying the actions often discussed by moral philosophers.

To date, we have no methods to demonstrate the presence or absence of this subjective side in species other than humans. However, unless the presence of emotions is directly linked to uniquely derived human features such as language, parsimony suggests that species closely related to us, such as great apes, have similar subjective emotions to ours (Flack and de Waal 2000). Accordingly, we postulate the existence of moral emotions in chimpanzees later in this paper (after all, they also cooperate, albeit in different ways).

In addition to asking about the direct causation of a behavior pattern, behavioral biologists also examine its ultimate causes. The ultimate cause for the presence of a behavior is its function, and we assume (and can sometimes show) that natural selection installed this particular set of motivational predispositions and response tendencies (and the associated emotions), because it prevailed over other ones. Thus, the behavioral predisposition is adaptive, and the behavior an adaptation, when the predisposition toward the behavior contributes to survival and reproductive

success in the particular context of that particular species, relative to specified alternatives. In the case of morality, this task amounts to identifying the adaptive significance of the moral psychology in the original human social organization, that of nomadic hunter-gatherers (henceforth: foragers).

An important fact about the connection between proximate and ultimate factors is that it is usually entirely statistical: instead of; The individual does not represent the function of the behavior, but is exclusively interested in pursuing the psychological goals that evolved in the context of that function. The perceived reward is that of reaching the psychological goal, not the fulfillment of the evolutionary function. Put simply, the performing animal has no clue of the function of the actions it performs. As a result, we can experimentally create conditions or offer stimuli where the existing proximate mechanism produces patently dysfunctional behaviors, as when gulls prefer to roll super-sized egg models into their nest rather than their own, species-specific eggs (Hinde 1970). This is also often true for humans, given recent radical changes in our social and physical environment (Tooby and Cosmides 2005). This may mean that the range of stimuli eliciting our moral emotions may have become far wider than in the original situation, and also that we may be ignorant of their functional significance (as in our disgust of incest).

4.1.2 Human Morality as Adaptive Behavior

In this chapter, we use the distinction between ultimate and proximate causation of behavior. We further divide proximate causation between motivations and emotions, to understand the nature of human morality as expressed behavior, the part most easily measured. Morality, say most dictionaries, is the quality of being in accord with standards of right or good conduct. This definition neatly encapsulates two key features: (1) morality is about doing "right," and (2) what is right is largely defined by the social norms of the society one lives in (the "standards"). The moral psychology of humans that underlies this behavior is characterized by a set of predispositions to actions and responses that function to maintain both reciprocity at the level of dyads and 'service to the group' (production of public goods). This has the effect of making the interests of the individual subservient to those of the community (Alexander 1987). Our goal is entirely descriptive: we wish to identify our moral psychology, as used in everyday behavior, and not to delineate the correct content of moral behavior.

As always, one can do worse than to go to Darwin for inspiration to explain the adaptive significance of this moral psychology. In his *Descent of Man* (1871/1981), he wrote: "morals and politics would be very interesting if discussed like any branch of natural history." In other words, he believed that morality evolved through natural selection. As to the function of these behaviors and emotions, he had some specific ideas as well: "the praise and blame of our fellow-men" and "love of approbation and the dread of infamy" act as "powerful stimulus to the development of social virtues." For Darwin, then, morality is largely about fitting into society and maintaining a good reputation. What he did not explain was why these emotions seem to be limited to humans.

Huxley (1894) and various influential evolutionary biologists since then (Dawkins 1976; Williams 1989) explicitly rejected Darwin's approach, arguing instead that morality represents a recent cultural invention, a successful rebellion against the tyranny of our selfish genes (see de Waal 2006). More recently, however, other biologists have followed Darwin in claiming that morality reflects an evolved adaptation (Alexander 1987; Rosas 2007).

Our contribution to this adaptive approach is that we propose a detailed account of the nature of this adaptation and the context in which it evolved. Our core proposal is that our moral emotions are the subjective dimension of the evolved proximate regulation of adaptive cooperation among human foragers. Life as a nomadic forager requires generosity and sharing due to extreme mutual interdependence. This lifestyle is radically different from that of the two extant chimpanzee species (*Pan*) or of any other ape, and is thus derived relative to the last common ancestor of humans (*Homo*) and *Pan* that lived some seven million years ago. We therefore aim to explain the presence and nature of our moral emotions, but will not claim that their referents, i.e. the conditions that elicit their deployment, are universal, because the biological substrate can interact with cultural evolution and thus add or even modify referents.

The most likely reason that this obvious adaptive hypothesis has not been proposed before is that most modern observers find it hard to imagine the foraging lifestyle. There are two reasons for this. First, this lifestyle is largely defunct, since virtually all people now live in pastoral, agricultural, or, increasingly, industrial societies, which are organized very differently. Indeed, most people, and virtually all scholars, now live in cities, large anonymous aggregations. As a result, the contents of moral judgments made using these emotions may have diverged after the radical departure and divergence of lifestyles since the invention of agriculture. Nonetheless, the hypothesis expects a core set of emotions, and presumably even social norms that are universal. Second, the foraging lifestyle is just as radically different from that of great apes, to which we often turn for inspiration about the life of early hominids. Extant great apes may thus show us the evolutionarily older core of our moral psychology, but not the derived outer layer.

This latter point hints at an inevitable consequence of adopting this evolutionary approach to morality. We must admit the possibility of other species' also having a morality that is adapted to their particular social system. We shall examine this issue in the discussion, where we attempt to draw up a chimpanzee moral psychology of cooperation. In all these attempts, of course, we must accept the limitation that we may not be able to identify the subjective character of the moral motivations these animals may have.

4.2 The Consequences of Being a Human Forager

4.2.1 Human Ecology

The humans we consider here are nomadic foragers, a lifestyle that is unique to our species. Extensive ethnographic and behavioral-ecological work over the last century has produced a consistent picture, explicated below, of the central features of extant

human nomadic foragers (Keeley 1988; Boehm 1999; Johnson and Earle 2000; Gurven and Hill 2009; Kaplan et al. 2009; Marlowe 2010; Hill et al. 2011). Foraging as a lifestyle gradually established itself after beginning in the early Pleistocene (around 1.8 million years ago [Mya]), during which early representatives of the genus *Homo* initially opportunistically, but gradually ever more systematically, engaged in hunting or scavenging large-game using stone tools to process and defend the meat of large mammals against carnivores, and perhaps to procure it as well (de Heinzelin et al. 1999; Pobiner et al. 2008). No doubt these large packages of meat were shared, but it is unknown when and in what order all the other elements characterizing human foraging arose.

Foragers today show a variety of characteristics: long-term, and to some extent, exclusive pair bonds (usually monogamous or mildly polygynous); a sexual division of labor; the presence of central places (camps) to which women and children always, and men usually, return for the night; intensive and obligatory food sharing and strong mutual support, including care for the temporarily sick and injured; and frequent collective action. They live in small camps of 20–50 people of shifting composition, which move multiple times per year. Camps consist largely of genetic (biological) and affinal (by marriage) kin of various degrees of closeness as well as various people not closely related to any others. Isolated nuclear families do not exist. There is no clear tendency for either sex to remain in their natal group once adult. Nomadic foragers live in highly egalitarian societies, largely free of dominance, that are not clearly segmented (i.e., without a hierarchy among camps or groups within a band) and are united by a shared set of customs and myths, as well as a common language.

Until the Neolithic, some 10,000 years ago, all people were foragers. Indeed, until the earliest settlements at highly productive places some time earlier, all people were nomadic foragers, and thus marked by strong interdependence and an absence of food storage (Keeley 1988). Since this lifestyle has been around for anywhere between 100 and 1,000 times longer than our settled lifestyle if measured in number of generations, we must assume (1) that it has had a profound impact on our psychology, and (2) that the basic architecture of the human mind has not changed fundamentally in the relatively short amount of time since our species adopted a sedentary life style, involving agriculture and more recently life in cities (cf. Tooby and Cosmides 2005).

Some may doubt that our foraging history has had such a pervasive effect on our psychology. Indeed, the hypothesis presented here is based on plausibility and not directly testable. Nonetheless, there is good indirect evidence that human psychology still reflects its evolutionary origin in egalitarian societies, in which systematic sharing prevented the buildup of major inequities, and all individuals knew that help would be provided when they needed it. First, the development of prosocial behavior suggests an innate component (see Sect. 4.4). Second, many features of modern behavior reflect this heritage. For instance, present-day people in more egalitarian societies have fewer health problems and better education outcomes, and have lower rates of drug abuse, mental illness, suicide, homicide and teenage pregnancy (Wilkinson and Pickett 2010); these effects continue to hold true when the overall level of affluence or education of a given society or group within a

society is controlled for. This persistent effect of social inequality strongly suggests that the fear of being uncared for in times of need creates chronic stress in the individuals in more despotic societies (Wilkinson and Pickett 2010). Finally, this history of egalitarianism is also reflected in the use of a reassurance signal by dominants, the smile, that was historically a signal of subordination (Preuschoft and van Hooff 1997); our 'cooperative eyes' (Tomasello et al. 2007), which serve as an indicator of shared intentionality by facilitating joint attention; our crying with tears, which elicits empathy, even from strangers; and our 'blushing,' which serves as an honest signal of shame.

We link human moral emotions to the proximate rules undergirding human cooperation. To build up our picture of these emotions, we therefore start by describing human cooperation. We then describe the known proximate rules regulating our cooperation, and subsequently interpret these rules in subjectively experienced, emotional terms.

4.2.2 Human Cooperation and Its Function

Human foragers cooperate, but so do nonhuman primates. The fundamental difference is that a solitary human forager, in stark contrast to a solitary chimpanzee or orangutan, is always far worse off than a cooperative forager. This situation made humans utterly dependent on intensive cooperation with fellow group members. Consequently, human cooperation differs from that in nonhuman primates in several important ways. The relative homogeneity of all great apes in this respect strongly suggests that these differences reflect features that are derived in humans, i.e. evolved anew in the human lineage. Figure 4.1 lists the major elements of human cooperation, as well as their proximate underpinnings discussed in the next section.

First and foremost, human foragers cooperate in the context of child rearing: we are cooperative breeders (Hrdy 2009; Sear and Mace 2008; Hill and Hurtado 2009), because men, older immatures and grandmothers make a major contribution to childcare, both in terms of energy and protein inputs through provisioning and as babysitters and teachers (Hawkes et al. 1998; Marlowe 2003). In contrast, all great ape females are essentially single mothers. Humans also tend to cooperate more with non-kin than other primates. This may be partly due to our tendency to form strong pair bonds, which are then extended to the kin of our pair partners (Chapais 2008).

Second, human foragers are cooperative hunters and gatherers. These activities involve two important cooperative actions: collective action and sharing. Much hunting and gathering involves group-level cooperation, whereas most cooperation in nonhuman primates is at the level of dyads. Cooperative hunting itself is now less important than it was before the invention of truly projective weapons (bows and arrows and spear throwers) and poison perhaps as long as 100,000 years ago (Marlowe 2005). Even so, hunters still often collaborate in joint pursuit, and other elements, such as obligate food sharing, were also retained. Sometimes, the collective action involves most of the group (Hill 2002; Hrdy 2009), as when foragers

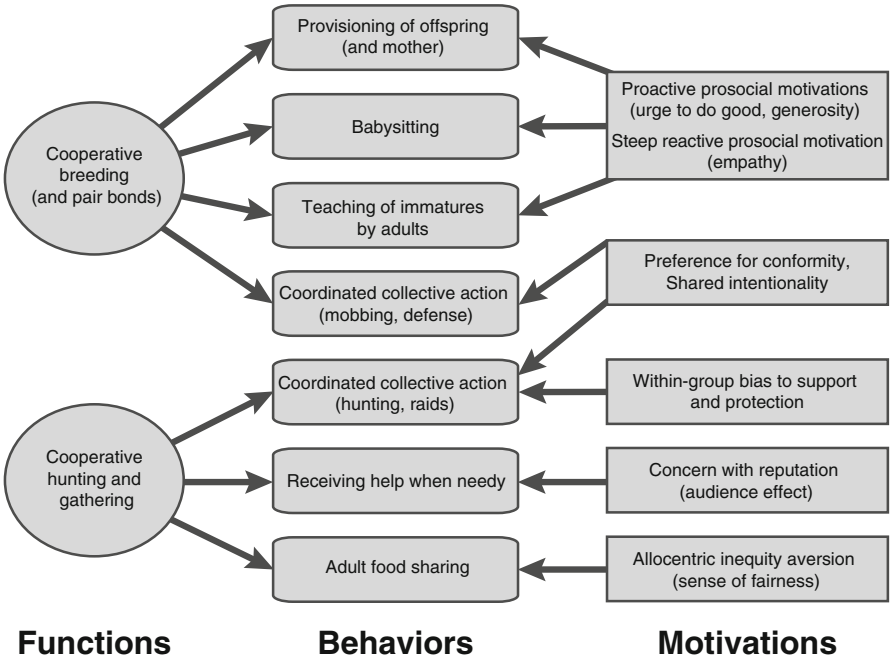

Fig. 4.1 The two major functional components of human cooperation, the behaviors that serve these functions, and their underlying proximate control mechanisms

collaborate in fish harvesting or moving camp. With regard to hunting, the yields were so uncertain that humans came to rely on food sharing, to the point that most nutrient-dense foods such as meat and honey became systematically shared.

This leads directly to another major difference with great apes: the fact that humans readily share food or help people they do not know directly, but only by reputation, such as when helping starving foragers who moved to their camp to ask for food. Reputation decides whether one is supported in times of need (Alexander 1987). A monkey or ape only needs to worry about its reputation with its direct interaction partners or those that can observe her interactions with these partners. In humans, language enables third parties who are physically elsewhere to have information about one's interactions with others, so people's reputations can literally reach a place before they get there in person.

Some other differences are probably not universal, in that they became most pronounced after foragers settled down and societies became larger and no longer largely face-to-face. The most relevant difference concerns the punishment of non-cooperators (de Quervain et al. 2004). Among foragers, people who are not willing to share are shunned by others (Marlowe 2009), and thus face reduced survival. Punishment is rare (Marlowe 2009), and when it occurs is highly collective (Boehm 1999). People in larger societies, in contrast, are often willing to incur some cost to punish free riders in group-level kind of cooperation in which individuals

contribute to common goals, and free riders risk the breakdown of all cooperative effort (Henrich et al. 2010).

Other features of human cooperation that have had an impact on our psychology may not be that different from those of chimpanzees, and thus perhaps, from the last common ancestor we shared with the African *Pan* apes. Like chimpanzee males, humans tend to engage in high-risk cooperation. Until the invention of long-distance weapons, hunting was necessarily cooperative, and hunting large game was dangerous. Hence, injuries were not uncommon (Boyd and Silk 2009). Joint hostile action against neighboring groups, as in raids, was even more dangerous, although great care was usually taken to create surprise and maintain a great, if temporary, imbalance of power (Wrangham 1999).

The key differences can be summed up as follows: humans cooperate in rearing young, in hunting and gathering, and this cooperation often involves collective action and proactive food sharing with less lucky foragers. We cooperate more with non-kin, partly because of our pair bonds, even with relative strangers (as long as they are perceived as within-group). Widely publicized recent empirical research with humans has served to confirm that our cooperative tendencies in small-group settings still largely reflect those of our forager ancestors (Fehr and Rockenbach 2004), albeit with clear-cut cultural variation (Henrich et al. 2005).

This is the broad picture. But it is also important to point out exceptions and limitations. First, humans are not exclusively prosocial: we also have selfish tendencies, and they may be tenacious. A tension between selfish and prosocial motivations is inevitable, since without the selfish motivations we would not survive very long. This tension is also reflected in individual variation. In the experiments of behavioral economists, a certain proportion of the subjects always behave as free riders, parasitizing on the generous dyadic donations or contributions to the public good by the majority (Gächter and Hermann 2006). Second, cooperation has its limits, namely at the boundary of the society. For instance, in interactions with individuals from a different community ('out-group'), different behavior patterns are observed, as reflected in the extensive research on in-group vs. out-group differentiation (Rabbie 1992), widespread reports in cultural anthropology (reviewed in de Waal 2006), and recent experiments on punishment of selfish individuals (Bernhard et al. 2006).

Let us now turn to the functional basis for these cooperative tendencies. The key feature from our perspective of forager ecology is that the foraging lifestyle has led to a strong interdependence among group members, on two levels. First, on a day-to-day basis, returns are uncertain, especially those of meat from large game, for which virtually all foragers have a strong preference (some languages have a special word for "meat hunger"). Without widespread and obligate sharing of meat, meat intake would be highly erratic (Gurven 2004; Hill and Hurtado 2009). Second, the foraging lifestyle leads to frequent injuries, and some of these may prevent a person from producing for weeks or sometimes months. People seek insurance against such periods of incapacitation by trying to be associated with good hunters (Sugiyama and Chacon 2000; Hill and Hurtado 2009). The benefits for the latter come in the form of being helped in turn when they are in need, good treatment for them and

their family in general, and perhaps in the form of sexual favors. In sum, then, each individual forager is part of a web of mutual sharing obligations. Each individual is therefore strongly aware that he or she must have a positive reputation in case of future need, and painfully guards it. As Hrdy (2009) put it: "their most important resources were their reputations and the stored goodwill of others." Because they cannot store food, foragers must store social obligations.

This situation gives everyone in a forager camp a stake in each other's (and thus the group's) wellbeing. Such a 'cooperate-or-die' situation is not common in nature, but it is found among various species, in particular cooperative breeders (Clutton-Brock 2002).

4.2.3 The Proximate Control of Human Cooperation

When closely related species have diverged in the kinds of cooperative behaviors in which they routinely engage, especially when the differences are as dramatic as between humans and great apes, we should expect that they have evolved differences in the underlying set of proximate motivations and response predispositions as well. Thus, we expect that humans have evolved some novel proximate mechanisms relative to the ancestral state that was probably shared with the extant great apes, none of whom have become cooperative breeders (Jaeggi et al. 2010a). Nonhuman primates are therefore useful as background, but not the best model organisms to identify the proximate control of human cooperation.

The best way to identify the set of underlying rules is to do experiments, and to check whether these experiments provide results consistent with the observed natural history. Numerous such experiments have been done with humans for decades, but in recent years these have been done to specifically test social preferences (as in economic games), and even more recently have been repeated as much as possible with great apes (and other primates) to identify the shared and derived aspects of our cooperative psychology. We have recently reviewed these experiments in detail (Jaeggi et al. 2010a), and here we briefly summarize the conclusions for humans (see also Fig. 4.1).

Perhaps the most basic distinction in cooperative motivations is between reactive and proactive prosocial motivations. Reactive prosocial behaviors are elicited by signals (or mere signs) of need, of social proximity (such as being a relative or a friend) and the presence or size of an audience. Numerous experiments and everyday experience show the existence of these reactive prosocial acts in humans. Friends and relatives generally elicit stronger support than strangers. Humans also find it hard to ignore expressed signals of need (crying, begging) even when these emanate from total strangers. These responses tend to be stronger in humans than in apes.

Human responses to prosociality-eliciting stimuli are stronger in the presence of an audience, apparent or real. Experiments have shown that respondents are not consciously aware of the effects of an audience on their prosociality (Haley and Fessler 2005; Bateson et al. 2006). This hypersensitivity to audiences functions to

ensure that selfish acts are only performed when no audience is present to spread gossip, i.e. to guard the actor's reputation. There is, as yet, no evidence for such sensitivity among nonhumans.

Food-sharing experiments allow a direct comparison and show an interesting contrast. Whereas among human children, food is often shared because owners spontaneously offer food to others or respond to requests by actively handing over food, most sharing among great apes is by passively allowing a beggar to take food or by having it stolen; refusals to share are also common (see Jaeggi et al. 2010b). We see similar differences in the sharing of information (as in teaching: Burkart and van Schaik 2010) and tools (Meulman et al. 2012).

Proactive prosocial behaviors are performed in the absence of any of these external stimuli, and are thus spontaneous. When dealing with animals, it is of course impossible to be sure that all external stimuli have been excluded, but in the case of humans we can check this directly. In one-shot, anonymous dictator games, a player is given a certain amount of money and given the option to share it with another player, who is a stranger. In such games, many people act prosocially in the absence of any obvious stimuli of need, social proximity or audience (Camerer 2003). This tendency also explains why humans in collective-action experiments generally begin by cooperating, only to start defecting when they notice others are free-riding and being unable to curb that through punishment or selective shunning of free riders (Fehr and Gächter 2002; Milinski et al. 2002).

The experiments have also revealed the presence of an additional regulator of the degree of prosociality, be it elicited or spontaneous. People tend to act to ensure that the degree of inequity that is created in these experiments remains within limits, a mechanism known as inequity aversion (Fehr and Fischbacher 2003). This is especially pronounced in the experiments in which windfall rewards are given, such as the commonly played dictator or ultimatum games, where no work is done for the rewards (an ultimatum game is like a dictator game, but now the second player must accept the outcome, and if she refuses, neither player gets any money; because refusals of low offers actually do happen, the proportion of the reward that is offered to the second player is generally higher than in the dictator game). Whereas animals may also show inequity aversion, it is predominantly of the egocentric kind, in which actors show an aversion to receiving smaller rewards than others (Brosnan and de Waal 2003). Humans show inequity aversion of both the egocentric and allocentric kind, and thus also show an aversion toward receiving bigger rewards than others (Fehr and Fischbacher 2003).

Human foragers have a motivation to conform to the majority and synchronize their actions with those of others, which may well be the collective expression of the shared intentionality, or "we" intentionality (Tomasello 2009). This requires that we comply with the norms of our society. Social norms are "socially agreed-upon and mutually known expectations bearing social force, monitored and enforced by third parties" (Tomasello 2009), or "standards of behavior based on widely shared beliefs how individual group members ought to behave in a given situation" (Fehr and Fischbacher 2003). Social norms are ubiquitous, but their content is not always easy to identify. Unless they are verbally explicated, norms are best recognized by

responses to their violation, and these reactions among foragers may be subtle, largely by way of shunning the violator. Nonetheless, foragers, like people elsewhere, have social norms, and have strong opinions about norm violators (Hrdy 2009; Marlowe 2009), which sometimes get expressed in violent responses (Boehm 1999). Moreover, experiments in modern societies have shown the underlying rules: people value the opinion of the majority, even if that seems counter-intuitive, as shown decades ago by the work of Asch (1955). Likewise, children preferentially follow the majority in social learning experiments (Berndt 1979), indicating the presence of active conformity. Below we offer a separate discussion on the evolution of social norms.

We can sum up the proximate rules underlying human cooperation as follows (see also Fig. 4.1). The first set is linked to cooperative breeding. Thus, we have highly prosocial tendencies, which are both proactive (spontaneous, unsolicited) and reactive, i.e. elicited by stimuli (need) that trigger our empathy and modulated by social proximity (friendship, kinship). These motivations ensure that we care for others. A motivation to conform to the majority and to synchronize plans is also expected in cooperative breeders, because they frequently engage in collective action. The second set of motivations may be more tightly linked to cooperative hunting and foraging, with the obligate food sharing imposed by the uncertainties of foraging returns. Thus, our prosocial tendencies are modulated by the perception of a fair balance among the rewards to the various players. They are also strengthened when the actor perceives the presence of an audience, reflecting the importance of reputation in our lives. Finally, like cooperative breeding, cooperative hunting and foraging have favored the strong tendency to have shared goals with others, and thus to internalize the local social norms.

4.3 From Proximate Rules of Cooperation to Morality

So far, we have shown that humans have evolved a derived psychology to support the intensive cooperation that characterized us as cooperatively breeding, nomadic foragers. We now move from the descriptive level of motivations, which are estimated by the degree to which these hypothetical constructs can predict behavior, to that of the experienced, subjective feelings (emotions) associated with these motivations to the acting individual human.

This is of course a difficult transition to make. First, it is not easy to formulate empirical tests, so we must rely on parsimony and introspection. This seems warranted because most of us immediately recognize the subjective emotions that accompany these rules as familiar. A second problem is that delineating individual emotions is problematic, perhaps because the underlying neuroendocrine mechanisms may create a large number of discrete or overlapping mental states (Christen 2010). As a result, there is no agreement among experts about which emotions can be recognized (but see Haidt 2007), and we will therefore refrain from naming and listing them (see Christen 2010). Nonetheless, some basic statements are possible, in that

all of us know about their existence introspectively. Peculiar to the moral emotions is that they have a strongly felt element of obligation ("ought"), suggesting that there is a high priority to fulfill them relative to other motivations, such as hunger, physical comfort and sexual desire, giving them a virtually privileged status.

Foundational to the prosociality rule is sympathy, which is empathy accompanied by a prosocial motivation, where empathy is the ability to imagine another individual's mental state. In humans, empathy tends to be readable from facial expressions, gestures and vocalizations (de Waal 2006, 2009). We can roughly translate sympathy as liking others and wanting to do good to them. We actually feel good about doing good, as suggested by both verbal reports and physiological indicators (Harbaugh et al. 2007; Steger et al. 2008).

The second element is our obsession with reputation, with how other people perceive our actions. This concern is only partially explicit, but much of the social anxiety felt by people, such as whether or not they are accepted by others, can be functionally linked to concern about reputation as well. Concern with the reputation of self is mirrored by a rabid interest in updating the reputation of others. Gossip is the main means of updating, and many people spend an inordinate amount of time engaged in it (Foster 2004), and more importantly, gossip arouses strong emotion at both the sending and receiving ends.

The third element is our urge to belong to the group and comply with its norms, the group-level pendant of the dyadically expressed shared intentionality of human children and adults (Tomasello 2009). People often report feeling happy when part of a synchronized group effort. Only humans engage in team-based play, often in a competitive context as team against team, in which the thrill is to engage in synchronized and coordinated joint action. We can add that we feel very guilty about violating group norms, even if unobserved, suggesting that we have internalized them and made them part of our conscience. Functionally, this may simply be about avoiding shunning or punishment, but our emotions suggest that we wish to abide by the norms.

In addition to this intrinsic component of conformity, there is usually also an extrinsic one that operates simultaneously. Compliance with local norms arises in part from sensitivity to peer pressure (the "moral community"). There is an implicit representation of this pressure in that one is more prosocial when one has been reminded of the presence of the moral community, as in the experiments using eyes as a subconscious cue of the presence of an audience mentioned above. This external influence on our moral emotions amounts to Adam Smith's (1759/2009) inner spectator or David Hume's (1739–1740/2003) homunculus, and is based on the well-developed human Theory of Mind, the ability to take the perspective of others. If we take this to its extreme and fully internalize the perspectives of others, such external motivators are no longer needed and represented, and we only possess emotions such as conscience (to avoid norm violations) and guilt and remorse (to repair norm violations).

The final element is founded on a sense of justice or a sense of fairness, which obviously restates some of the egalitarian social norms. We deploy our empathy in favor of the weaker individual when that person is treated unfairly by the stronger

person. Depending on the opportunities to express these concerns, we quietly disapprove, we protest and try to recruit others in our disapproval, or we punish them. In the latter case, we actually feel good about this (de Quervain et al. 2004). Punishment has the goal of restoring equitable distributions.

The last step in the argument, which is admittedly untestable, is the claim that these four elements (sympathy, concern with reputation, the wish to conform, and a sense of fairness) are the major components of human moral psychology, upon which our reflective morality is built. Additional elements have also been proposed (Haidt 2007), but we suspect that they have been co-opted into the moral psychology more recently, post-foraging.

4.4 Discussion

4.4.1 The Key Features of Human Morality

Our main claim is that the moral emotions that underlie the snap moral judgments people make in everyday situations are in fact the subjective side of the evolved proximate regulators of our uniquely intensive cooperation. This claim thus refers both to the capacity to have moral emotions and to the contexts in which they evolved. Human cooperation differs in some major ways from that of our closest primate relatives, especially due to our interdependence, which has favored the evolution of proximate rules that prioritize the wellbeing of fellow family and group members and guard one's reputation as a sympathetic person worthy of support in times of need. Obviously, we expect the moral emotions of other species to be adjusted to different, only partly overlapping contexts.

The Darwinian hypothesis about morality presented above is congruent with many known features of practical human morality. First of all, we now know that many every-day moral decisions are not entirely built on conscious deliberation but rather on intuitive, and rapidly executed responses, too fast for mental calculation to have affected them (cf. Greene and Haidt 2002), showing there is an intuitive, emotional (non-cognitive) core (Haidt 2008). Unreflected morality is thus more about emotion than about cognition. Second, practical morality has a tribal element (Bernhard et al. 2006), strongly distinguishing between in-groups and out-groups. We need reason based upon conscious reflection, to go beyond 'tribal morality' (larger social units are more recent than small tribal groups and are linked to higher densities and sedentism). This in-group bias needs to be overcome through reasoning and redefinition of in-group membership, reflecting its origin as an adaptation for life in small-scale societies that had few contacts with the outside. Third, the hypothesis explains why all societies have moral codes (i.e., morality is a "human universal"), even though the content of these codes often varies in space and time. Thus, moral emotions were neither bequeathed to us from the outside by some benign force, nor invented by us at some time during our history, but are instead

the product of an evolutionary process functionally linked to our uniquely intensive cooperation. Although some philosophers may be willing to admit that morality has some biological basis (e.g., Kitcher 1998), many are not really interested in this fact, or in the small moral acts that we perform on a daily basis, and in fact leave the ultimate explanation of our moral faculty unexplained.

A secondary, but perhaps in the end more important claim is that the hypothesis can in fact explain biases in the content of moral judgments. Perhaps the majority view among philosophers is that the enormous variety of actual moral systems shows the irrelevance of a biological core and the supremacy of cultural influences (Prinz 2007b; Chap. 6, this volume). The challenge to the position that argues that morality reflects a biological adaptation is to account for this variability. It does so by arguing that natural selection put in place the set of basic principles by which to regulate cooperation in a foraging system, not to adjudicate on the major moral dilemmas facing the modern world, such as abortion, animal suffering, the production of transgenic organisms, using nuclear energy, etc. The fact that the majority of social norms have a cultural influence (Rudolf von Rohr et al. 2011) can be explained by the diversity of social conditions that arose subsequent to the onset of food production. Indeed, among different foragers, rules are largely similar (Hill 2009). The evolutionary hypothesis can use the same argument to account for the fact that philosophers cannot agree on a universally accepted system to derive moral rules (think of utilitarianism, deontology and contract theory). Selection simply did not provide rules that were flexible enough to deal with modern problems. The cultural variability is then seen as the product of the interaction between innate moral preferences and a preference to conform with social norms, once they have been established. The major distinction, then, between the evolutionary approach and the purely cultural approach is that the former predicts that norms generally show modal tendencies that reflect a biologically determined set of innate moral preferences.

Here, we will argue in favor of the view that there is a biological core to our morality that reflects a set of innate moral preferences. This view can be reconciled with the enormous variety of actual moral systems, a variety that has been argued (Prinz 2007b) to show the irrelevance of a biological core and the supremacy of cultural influences. We suggest that the approach that sees morality as an adaptation to life in small-scale egalitarian societies may reveal a basic set of moral preferences that underlie all moral judgments, along the lines of those presented in Fig. 4.1. Their presence places constraints on the contents of moral judgments and social norms, at the very least arguing that their contents are not biologically arbitrary, and may even lead to some universal social norms (Rudolph von Rohr et al. 2011). For instance, it is unimaginable that moral philosophers will ever develop rules stating that practices involving mistreating infants or brother-sister marriage are morally desirable. In other words, there is a strong bias toward prosocial content, with a heavy emphasis on concepts like fairness and equitability, and on avoiding harm to the defenseless. In the end, despite philosophical reasoning, moral judgment sometimes comes down to ‘gut feelings’, i.e. intuitions with a strong emotional basis, which are usually heavily prosocial in nature. Thus, we suggest that biology may have constrained the contents of human morality.

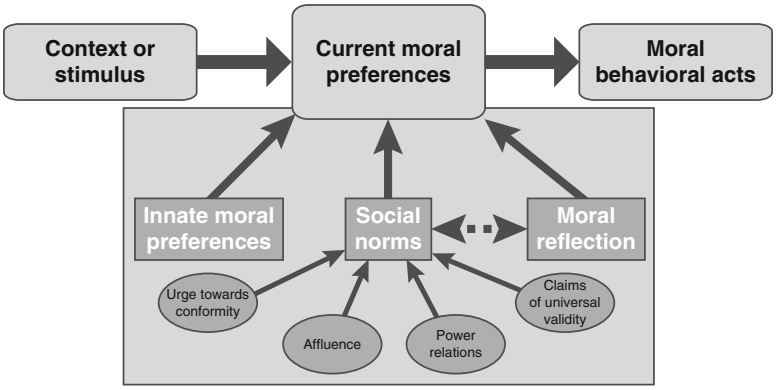

Fig. 4.2 The factors molding individual moral preferences in humans, and thus the way in which an individual responds with moral behavior in a particular context. These factors include innate moral preferences, social norms, and (potentially) individual moral reflection

This position requires that we flesh out the explanation for the geographic and cultural variation in social norms and the products of moral reflection in terms of interactions among, or suppression of, these basic moral emotions. Figure 4.2 summarizes this explanation. First, the moral emotions we postulated above based on forager ecology may produce conflicting results. Thus, our prosociality-based empathy may be in conflict with fairness. For instance, when assigning patients eligible to receive an organ transplant when suitable organs are scarce, we may be in favor of a set of rules that assign these organs based on a list of objective criteria that identify the most eligible candidates, but are simultaneously inclined to help a good friend or relative in need of such an organ (e.g., Batson 1991). In addition, both empathy and fairness may often be in conflict with our urge to conform or with our strong in-group bias. This conflict must be resolved by rational argument fed by moral reflection, and rules are derived (although this is when ethical systems produce conflicting conclusions). Moreover, most of us feel that moral judgments must have a universality claim: They must lead to conclusions that ought to be universally applicable.

Second, resource limitations force individuals to set priorities, caring first for the family, and then the tribe, the nation, the world's population, and finally other organisms, as affluence increases. Thus, the 'expanding circle' (Singer 1981) is better seen as a pyramid (de Waal 2006) floating on the resource base: the broader the base, the wider the circle. Third, the variation may reflect lack of knowledge, as when suffering elsewhere in the world was not visible until modern communication techniques made it so. Finally, variation may be due to coercion or manipulation by powerful elites (e.g., by convincing us that others are monsters not worthy of receiving moral treatment), preventing the expression of the basic moral emotions. For all these reasons, the emotionally based innate moral preferences may be molded into current moral preferences to fit current conditions.

We stress that this implementation of the evolutionary approach to morality claims that morality is universal among humans and that there are biological constraints on the contents of social norms or moral judgments. However, we also want to stress that we do not claim that their content is entirely determined by the adaptive set of moral preferences found in humans evolved to guarantee success in a system of hunting and gathering. We accept that moral reflection is unique to humans, and stress that we do not wish to replace moral reflection or deny its importance.

4.4.2 Development and Morality

Tinbergen (1963) recognized another aspect of the proximate causation of behavior, namely its ontogeny or development. Although in theory an adaptive behavior pattern can be either fully learned or strongly canalized (i.e., have a close correlation with its genetic basis) in its development, adaptive behaviors commonly have some innate component. If moral behavior reflects such an adaptation, it would be likely to appear at an early age, before getting modified in all possible directions by social learning or individual reflection.

The development of moral behavior has recently undergone a revolution based on new empirical results, which end up supporting the perspective developed here. Rather than having morality drilled into them by rational adults against their natural tendencies, recent findings suggest that proactive prosociality, and indeed even the distinction between conventions and morality, arise so early in life that they are likely to have some innate core (Tomasello 2009; Vaish et al. 2009). As young as 6 months of age, human infants already show signs of empathy (e.g., Draghi-Lorenz et al. 2001), and equally young infants already show moral evaluation of social events (Hamlin et al. 2007). Thus, the major elements of our moral emotions arise very early, suggesting an innate core.

4.4.3 The Phylogeny of Morality

On the ultimate side, in addition to function or adaptive significance, Tinbergen (1963) also distinguished the phylogeny or evolutionary history of the behavior in question, which in our case leads to asking about the taxonomic distribution of the components of morality. Knowledge of the taxonomic distribution of a behavior often can be used in testing hypotheses about its function through the comparative method. In the case of morality and moral emotions, this would be difficult since we have no yardstick to measure it yet, but one can speculate about what the content of morality should be in a given species.

If we assume that other, especially closely related species also have subjective experiences underlying their proximate rules, then we must expect a high overlap in the emotions between humans and great apes. Indeed, this is what most research

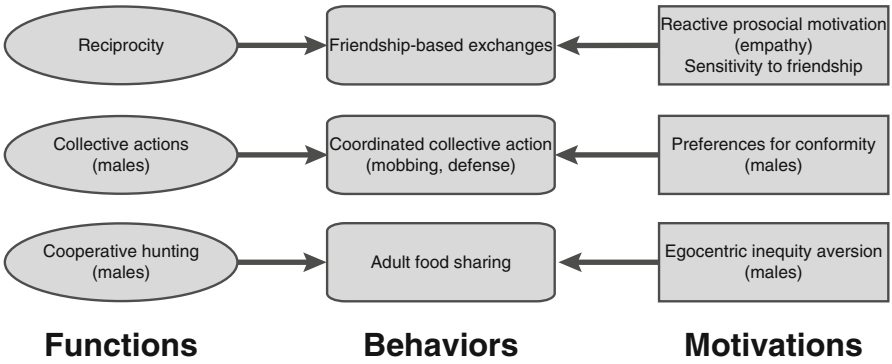

Fig. 4.3 The contexts of the major functional components of chimpanzee cooperation and their underlying proximate control mechanisms

suggests (de Waal 2006, 2009; Flack and de Waal 2000; Rudolf von Rohr et al. 2011). This means that what we subjectively experience as moral emotions or intuitions may also exist among nonhuman animals. However, we have also argued that humans have evolved a different style of cooperation, and therefore a different set of proximate rules; we should therefore also expect differences in the set of emotions, and therefore the moral emotions of chimpanzees, say, should be different from ours. The best way to find out is to examine the style of cooperation, identify the proximate rules and infer the emotions from that.

We expect differences in motivations to directly correspond to differences in the natural history of cooperation (see Jaeggi et al. 2010a). Chimpanzees differ from humans mainly in having less intensive cooperation and lacking the mutual interdependence of human foragers, as well as having less frequent collective action. Although male chimpanzees do engage in collective action to some extent when hunting and more clearly during violent confrontations with neighboring communities (Wrangham 1999), usually only part of the male cohort participates in these activities (e.g. Mitani 2006). Figure 4.3 for chimpanzees follows the organization of Fig. 4.1 on humans, and obviously only refers to the moral preferences surrounding cooperation (like humans, there is possibly a morality concerned with the avoidance of harm). We thus expect clear overlap, in that there should be reactive prosocial emotions and some preference for conformity (expected to be more strongly expressed in males). We also expect a sense of fairness, but mainly of the egocentric variety. However, we also expect clear differences, such as the absence of strong proactive prosocial motivations or strong concern with reputation. These absent components lead us to expect a difference in the nature of the innate moral preferences.

This approach assumes that chimpanzees have moral emotions, but is there any evidence for this? Flack and de Waal (2000) and de Waal (2006) refer to mechanisms to keep up dyadic prosocial behavior, but also to community concern. Rudolf von Rohr et al. (2012) have provided new evidence for community concern in chimpanzees. However, a more direct approach is to assess whether chimpanzees

meet the preconditions for moral emotions, in that they evaluate situations differently when in the role of uninvolved bystanders. A recent experiment by Rudolf von Rohr et al. (unpublished) tried to investigate exactly this. The results showed that chimpanzees respond differently to video scenes containing severe aggression against infants, including infanticide, compared to control video scenes that also contained social aggression but lacked the putative moral transgression of attacking an infant. Although they expressed their reactions merely in increased looking times, rather than visible outbursts of emotional reactions, this might be explained by the use of video stimuli rather than realistic events. More work is needed to experimentally show the existence of moral emotions in chimpanzees.

4.4.4 The Evolution of Social Norms

Finally, we briefly address a topic that must be developed more fully to understand the evolution of morality, namely the relationship between individual moral preferences and social norms. All norms (and thus morality) reflect attempts to resolve fundamental conflicts of interest. In the absence of conflict, no norms would be necessary because all actors would be in perfect harmony. In the absence of overlapping interests, however, they would also not be necessary because individuals would not live together in a society. The presence of social norms further requires that the majority can exert enough power against powerful individuals, a requirement called subordinate leverage in the animal behavior literature (Preuschoft and van Schaik 2000).

Individuals in all social species have expectations about how others should behave toward themselves; one might call these expectations private norms. If they are violated, other individuals should expect avoidance, protest or aggression on the part of ego (Rudolf von Rohr et al. 2011). Much aggression among animals can be seen as a response to the trespassing by others. For instance, a subordinate monkey attacked by a higher-ranking one may respond with special screams that attract allies with a higher probability than regular screams (de Waal et al. 1976). Normally, such attacks are unprovoked and occur without the subordinate having come too close or attempting to steal the dominant's food. These screams can be interpreted as protests at having one's egocentric, private norm violated.

Norms become majority norms when they also apply to expectations concerning the behavior of third parties, i.e. dyadic interactions that do not involve the owner of the norm. A possible example of such a majority rule, shared by chimpanzees, humans, and probably many other nonhuman primates, is "do not attack infants." In many nonhuman primates, males that attack small infants are greeted by a sheer-collective outcry and fast approaches that may involve mobbing and even outright physical attacks by individuals not related to the infant and thus not directly affected (Hrdy 1977; review: van Schaik 2000). It is likely that the protesting individuals are highly aroused emotionally, because these reactions are highly intense. It is clear that the majority (most group members) benefits from this rule, whereas a small minority (infanticidal males, infanticidal females) stands to gain

from harassing or killing particular infants. Since some individuals do commit infanticide under some conditions in both humans and nonhumans (van Schaik and Janson 2000; Daly and Wilson 1999), they must have a temporarily high motivation to attack and kill particular defenseless babies. Such majority norms are therefore upheld because violators are punished by others, rather than by intrinsic motives shared by all group members. Trespassers actually benefit from violating the norms and only abide by them to avoid the effective coalitionary attacks (punishment) by the majority of group members. Majority rules may be widespread, but they do require sufficiently strong subordinate leverage to allow the majority to punish dominant norm violators.

Given that norms reflect the presence of a conflict of interest, it may be hard to explain the existence of collective norms (norms that are complied with because all individuals benefit from adhering to them). Yet mobile foragers show what look like collective norms, perhaps because individuals manage to control temptations to violate them. The reason for this may be that any short-term gain from not sharing food or stealing from others would generally be offset by the loss of one's reputation, jeopardizing receiving future help from others when in need. Thus, the functional reason why the temptation to trespass still exists may well be that if the violation is not detected by anyone, then the violator gains (subtle cheating: Trivers 1971), whereas the functional reason to internalize the norm as private psychological goal (thus making it collective) is the high future cost of being detected (relative to the short-term gain). But even among foragers, potential dominants may refrain from taking more than their fair share in part out of fear of being ridiculed or ostracized by the "moral community" (Boehm 1999). Majority norms and collective norms may thus be on a continuum, where we expect humans to be closer to the collective end and nonhuman primates such as chimpanzees closer to the majority end, on average.

4.5 Conclusions

The moral preferences described in recent research correspond remarkably well with the proximate machinery required to maintain cooperation in small groups of human foragers, suggesting that they represent the subjectively experienced dimension of the motivational rules underlying this cooperation. Other species may therefore have their own moral emotions and their own social norms, adjusted to their specific form of cooperation. We suggest that chimpanzees in particular have overlapping moral emotions, but largely lack specific human ones, such as the desire to do good without being asked (generosity), a concern with reputation, and an allocentric concern with fairness. In species that lack the strong interdependence that humans have evolved, prosocial behavior may mainly reflect the fear of punishment, and it is more the interests of the group's majority that are expressed in moral emotions. Thus, cooperative breeders and hunters, such as wolves or wild dogs, may possess the closest analog to human morality.

To sum up, our core claim was that the moral emotions that underlie the snap moral judgments people make in everyday situations are the subjective dimension of the evolved proximate regulation of cooperation among our forager ancestors. Cooperation among foragers is characterized by generosity and sharing, reflecting their extreme mutual interdependence. Social norms reflect a conflict of interest between individuals, resolved in favor of the majority, either because the potential violator has a long-term interest in complying with the norm (collective norm) or because she can prevent being punished this way (majority norm). Some norms and moral emotions may be present in other animals, although they probably differ in the degree to which social norms are internalized and represented as their own psychological goals.

Chapter 5
Precursors of Morality – Evidence for Moral Behaviors in Non-human Primates

Sarah F. Brosnan

5.1 The Evolution of Moral Behaviors

Is moral behavior unique to humans? Although moral behavior is primarily discussed in relation to humans, if a function of moral behavior is to promote social cohesion and harmony within a social group, there is no *a priori* reason not to expect a similar set of behaviors in other social species (Bonnie and de Waal 2004; Flack and de Waal 2000; Haidt 2003). Although this certainly does not mean that what we see in other species need be identical to humans' behavior, there may be a suite of related behaviors that have evolved for the same purposes in other species. Of course, this idea is not new. In *The Descent of Man and Selection in Relation to Sex* (Darwin 1871/1981: 71–72), Darwin argued that sociality is innate, rather than created by humankind, and provided a framework for the development of morality in any species. The question is, then, from which precursor behaviors did morality evolve, and how can we study this in other species?

Moral behavior itself encompasses many different aspects that can be empirically studied. Moreover, some of these features may exist in non-human species. These features in and of themselves are not moral per se, but may represent the kinds of raw material from which evolved the moral behaviors we see in humans. For instance, we know that chimpanzees will help experimenters, for instance to obtain objects that are out of reach, which may be due to the expectation of a reward rather than any reason within the realm of morality. Humans may also help others because they get something out of it, not because it is the "right" thing to do. In both cases, the behavior still serves to benefit another. Once a behavior such as this exists, natural selection may then shape it in to a moral behavior. We can think of such behaviors as evolutionary "building blocks" in the sense that they provide the

S.F. Brosnan (✉)
Departments of Psychology and Philosophy, Georgia State University, Atlanta, GA, USA
e-mail: sbrosnan@gsu.edu

M. Christen et al. (eds.), *Empirically Informed Ethics: Morality between Facts and Norms*, Library of Ethics and Applied Philosophy 32, DOI 10.1007/978-3-319-01369-5_5,

foundation upon which moral behavior was "built."[1] These building blocks may differ in function, mechanism, outcome, or breadth from human moral behaviors, but they are nonetheless informative. They can reveal much about how the behaviors originated, were first used, and subsequently were shaped by environmental or selective forces. Using the comparative approach to investigate the evolution of our behaviors, we have the opportunity to learn more about from whence our human moral behaviors came, which can give us both a better understanding of them and a better ability to adjust the environment to encourage (or discourage) them.

5.1.1 The Empirical Study of Moral Behavior

With this in mind, then, it may be more appropriate to break down the study of "moral behavior" to specific components that can be investigated empirically. There are, of course, myriad ways in which moral behaviors and their constituent components can be considered. Below I consider two that may be particularly important when investigating the evolution of human moral behavior, but also are feasible to study in other non-verbal species.

First, one can study whether there are behaviors that may contribute to social regularity, or the smooth functioning of the social group. This is a key function of moral behavior, and so such behaviors may be related to those seen in moral behavior in humans. A side note here is essential. Behaviors that evolved because they provided fitness benefits, such as a well-ordered society, need not have been understood as such by the individuals involved. Behaviors can exist even when the individuals themselves do not understand why. As a case in point, humans exhibit behaviors we might consider as moral at a very young age (e.g., LoBue et al. 2009), when it is unlikely that they understand these behaviors as 'moral'. In fact, it is likely that at this point their behavior is the result of basic associative learning. They are simply responding to the feedback, positive and negative, that they have previously been given for their actions. Other species, too, respond to associative feedback and it may well be that the reactions of conspecifics lead to a series of behaviors which are generally followed by the group without any understanding or conscious intent on the part of the actors. This process of conditioning by one's group mates has been documented to cause large-scale changes in group demeanor that can outlast any of the initiators (Sapolsky and Share 2004).

Second, one can study mechanisms related to moral behavior, such as moral emotions. Moral emotions have been described as those emotions that are

[1] Although these words imply foresight and intentionality, note that evolution neither plans for the future nor aims for an optimum. Those traits that provide the most benefit, in terms of reproductive output, in the current environment are said to be selected because they are more likely than alternate traits to be passed on to subsequent generations. Once the environment changes, the trait that is favored by selection will change as well.

other-focused, are elicited by stimuli which do not affect the self (e.g., inequality eliciting sympathy), and result in an outcome which helps others or society (Haidt 2003). Moral emotions are both positive (e.g., elevation, gratitude, or pride) and negative (e.g., shame, guilt, and embarrassment; Tangney et al. 2007). There is growing evidence (Aureli and Schaffner 2002; Darwin 1872/1998) in favor of the presence of emotions in other species (although none as yet for specifically moral emotions). This suggests that emotions represent a phylogenetically ancient mechanism for affecting behavior. Thus, the study of emotions may provide insight into mechanisms that shaped species' behavior in ways relevant to moral behavior.

5.1.2 What the Evolution of Behavior Is Not

What is meant when we say a behavior evolved from an innate feature? Importantly, asserting that there is an innate substrate to moral behavior does not suggest an expectation that the evolved moral behaviors will be invariant. Virtually no biological feature is invariant despite innate underpinnings, and hardly any evolved behaviors are either universal (that is, present across all individuals of a species or group) or wholly unaffected by the environment or learning. To this end, asserting that there is no innate component to morality is a misrepresentation of the argument. As shown throughout the 'nature vs. nurture' debates, virtually any trait, behavioral or other, has both a genetic and an environmental component, and denying one of these is as fallacious as denying the other. The most appropriate approach is to consider both what biological factors underlie a trait (i.e., moral behavior) and how these factors interact with the environment (including culture) to create the variation we see both within and between species and groups.

Second, while this should not be necessary, given that it was first articulated in the eighteenth century by Hume (1739–1740/2003), there is often still confusion between the scientific goal of understanding *what is* and the ethicist's goal of determining *what ought to be*. These goals are different. Science is descriptive, not normative, and a scientific discussion of how behaviors emerged and developed may not tell us anything about the value of these behaviors, even relative to each other, or the goals to which we, as humans, should aspire. In the context of evolutionary biology in particular this misunderstanding may be due to the early tendency to organize the species according to a hierarchy topped by man (e.g., the *Scala naturae*). Nonetheless, evolution does not make 'progress' in the sense of getting better, ethically or otherwise. This tendency towards assuming normativity may be doubly fraught when considering moral behavior, which inherently expresses normativity. In the following paper I discuss only the ways in which moral behaviors may emerge, with no normative goal of prescribing behavior. If a behavior is referred to as 'benefitting' I mean it only in the sense of a cost-benefit analysis based on long term survival and reproduction (the currency of natural selection) and not as an ethically desirable behavior.

5.1.3 Homology vs. Convergence

Studying the evolution of these behaviors begs another question: if behaviors are similar between or among species, what does that tell us about their evolutionary development? There are two possibilities. First is that the behavior is shared between the species due to common descent. This pathway, known as homology, means that a common ancestor had already evolved the behavior and that it was then passed down to all of the descendent species. Homology explains, for instance, why almost all bird species can fly. Once the behavior emerged in the common ancestor it was inherited by all subsequent species.

The second possible pathway is that the behavior may be shared among species due to common environmental pressures, constraints, or opportunities. In this case, otherwise known as convergence or homoplasy, a trait or behavior is beneficial in a certain set of environmental conditions. Any species that experiences these particular conditions and happens to have some of its members undergo the right mutation or recombination event will subsequently develop the beneficial trait. Convergence explains why many insects, birds, and even a mammal (the bat) developed the ability to fly. Presumably an open environmental niche, the air, which allowed both increased opportunities to escape from (some) predators and access to additional food sources, led these disparate species to develop the same trait, without an intervening common ancestor providing it to both.

Thus, when studying the evolution of moral behavior, one approach is to break down the question and consider different aspects independently, all the while bearing in mind the ways in which similarities between species are and are not informative. As discussed above, it is possible to look for those behaviors which may be related to social norms. These may be behaviors such as third party interventions, in which one individual intervenes in a conflict in which she or he has no personal interest. This is currently an approach that has yielded a great number of insights across a large number of topics related to moral behavior. A second approach is to consider potential mechanisms for moral behavior, such as empathy. I discuss each of these possible strategies below.

5.2 Behaviors That Uphold Social Norms

Although there are few, if any, studies of morality in other species, research done on other behaviors may point to situations in which standards or norms may exist, representing situations from which moral behavior may have emerged. Below, I summarize this research across several topics related to social behavior, discussing what is known and how it may link to moral behavior. I focus on studies involving primates for the reason that humans are primates as well. This does not mean, however, that similar moral precursors are absent in species beyond the Family Primates.

5.2.1 Social Interactions

Given that moral behavior is fundamentally social, one obvious area in which to look for moral behavior is that of social interactions. There are many basic social rules which seem to be enforced in primate groups. Without social rules that restrict dominant or stronger individuals from behaving in any way or taking anything they like, there is no basis for moral behavior.

Possession rules are often maintained, at least among some species of primates (Brosnan 2011; Kummer and Cords 1990; Sigg and Falett 1985). In these experimental studies, primates typically do not take desirable objects that are in the possession of another, although this respect for possession ceases the moment the object is no longer in an individual's immediate control. Interestingly, what counts as 'possession' seems to vary by species in ways that might have been predicted based on the species' social environment. Hamadryas baboons maintain harems, so males have 'possessions,' or females, which are not under their direct physical control. In experimental studies with objects, these males treat others' possessions as deserving respect even when they are not under their immediate control, as long as they are within a (rather small) close radius (Sigg and Falett 1985).

With social norms comes the possibility of individuals supporting one another to uphold these norms, even when they are not affected. This behavior is seen in species other than humans. For instance, among primates, there is evidence of third party interventions in a variety of contexts (de Waal and Luttrell 1988; Nishida and Hosaka 1996). Such behavior is not restricted to primates, either; other species, such as ravens, are known to support third parties to maintain social norms, such as possession rights (Heinrich 1999). Such third party support is among the behaviors consistent with moral behavior, as it supports the social rules and norms in the group.

Individuals may also support one another during fights. Typically the pattern is for individuals to support winners, that is, enter in to a fight on the side of the individual who is currently winning (Machida 2006; Watanabe 2001). This has obvious benefits in terms of gaining the later support of these winners and being viewed as a part of the winning side, which might result in reputational benefits, but in some situations, loser support is also seen. Dominant male Japanese macaques are more likely to support losers than winners (Watanabe 2001), and among chimpanzees, loser support may be the predominant approach (de Waal 1978). However, even among these chimpanzees, males engaged in a dominance contest supported winners until they had established their rank. Nevertheless, the presence of loser support shows the presence of a behavior that benefits another and hence could be a precursor to moral behavior among these species.

Finally, policing behavior, in which one or a few individuals mediate disputes and settle conflicts, at the risk of cost to themselves, has also been documented in primates (Flack et al. 2005). Typically the individual doing the policing is a high ranking male in the group, who clearly has much to gain from group stability. The presence of policing appears to stabilize social networks, potentially making it

easier for behaviors that may be critical for the emergence of a moral system, such as cooperation and culture, to emerge (Flack et al. 2006). Such behavior represents an important step in the development of moral behavior, as individuals may act against their own immediate best interest to support group harmony and stability.

5.2.2 Reciprocity

Reciprocity is the exchange of goods or services between individuals (Brosnan and de Waal 2002). Although reciprocity presumably evolved due to the benefits of cooperation (Dugatkin 1997), reciprocity is also fundamental to moral systems. Consider that among humans, reciprocity is often lionized as the "Golden rule" (i.e., Do unto others as you would have them do unto you.). The presence of reciprocity in other species provides a base upon which moral behaviors may develop. Note that in this context, reciprocity need not be calculated, but may be based upon emotions or rules of thumb (Brosnan and Bshary 2010; Brosnan and de Waal 2002).

One area in which reciprocity has been argued to exist is in food sharing. Food sharing in primates is fairly rare outside of the parent/offspring bond (Feistner and McGrew 1989; Feistner and Price 1991). However, among one of our two closest relatives, the bonobo, some studies have found that food sharing is fairly common (Wobber et al. 2010) and social tolerance is high (Hare et al. 2007; although see also Jaeggi et al. 2010b for evidence that chimpanzees may be more tolerant than are bonobos). In our other closest relative, the chimpanzee, food sharing occurs primarily between adults and infants (Silk 1979) or after a group hunt (Boesch 1994). There are only a handful of documented cases of sharing a commodity other than meat between adults in the wild (Bethell et al. 2000; Hockings et al. 2007; Nakamura and Itoh 2001; Slocombe and Newton-Fisher 2005). In most of these cases, sharing was either noted as an anomaly or explained as a mechanism for gaining either status or mating opportunities (or both; Hockings et al. 2007). Beyond food, chimpanzees exchange meat, grooming, sexual access, and support for each other (Duffy et al. 2007; Gomes and Boesch 2009; Gomes et al. 2008; Mitani 2006; Nishida et al. 1992; Watts 2002).

Outside of these field observations, contingent reciprocity between chimpanzees has been extremely difficult to elicit (Brosnan et al. 2009a; Melis et al. 2008; Yamamoto and Tanaka 2009a). One of the issues may be that reciprocity typically takes place over the course of a minimum of several hours (de Waal 1997) and, often, weeks (Gomes and Boesch 2009; Gomes et al. 2008). On the other hand, most experimental studies require reciprocity on the order of minutes (e.g., Brosnan et al. 2009a). A second issue is partner choice, which is freely determined in most field studies, but is determined by the experimenters in most lab studies. It may be that reciprocity occurs most easily amongst individuals who choose to interact with one another, possibly because reciprocity is mediated by familiarity and positive emotions (Brosnan et al. 2010b; Schino and Aureli 2010). Finally, it may be that the majority of reciprocity is not contingent upon the immediate preceding behavior but

based, again, on relationships. This subtlety is captured in the field, but not in short-term experiments.

Despite chimpanzees and bonobos being our closest relatives, other species are nonetheless important to consider in order to more fully understand the ecological pressures that lead to convergence on reciprocal behavior. Amongst the other primates, reciprocity also occurs, primarily in as the contexts of support and grooming (e.g., de Waal 1992; Perry 1997; Schino and Aureli 2008). Reciprocity also occurs outside of the primates (e.g. Hart and Hart 1992; Milinski et al. 1990; Romero and Aureli 2008; Rutte and Taborsky 2007). In fact, perhaps the most famous example of reciprocity concerns vampire bats, which give each other blood meals when required to avoid starvation (Wilkinson 1984). Reciprocity appears to be phylogenetically widespread among mammals, as well as some non-mammalian vertebrates, giving us the opportunity to use the comparative approach to determine which social or ecological factors are most closely associated with the evolution of reciprocity, and possibly indicating that social norms and social rules are generally critical to group living species.

5.2.3 Inequity

Among humans, behavior changes when outcomes are not the same (Walster et al. 1978). Even if both individuals are better off for having completed an interaction, humans typically respond negatively (e.g., express displeasure or change their behavior; Walster et al. 1978) if someone else got more than they did. The extensive literature on economic games, particularly games such as the Ultimatum Game (e.g., Guth et al. 1982), show the degree to which people will go to equalize outcomes.

On the surface, of course, disliking getting less than a partner has little to do with morality and everything to do with self-interest (morality implies disliking getting *more* than a partner, or disliking when two other peoples' outcomes are not equal). However, it may be that moral behavior cannot develop until individuals have a concept of inappropriate behavior, possibly gained through one's own experiences. If this is the case, then evolving an ability to judge one's own outcomes relative to another's may be a critical step in the evolution of a more nuanced and other-oriented moral system. After all, it is much easier to explain how natural selection would have increased the frequency of behaviors that ameliorate one's own negative outcomes, which has an immediate positive benefit to the individual actor, before selecting for behaviors which ameliorate those bad outcomes in others (Brosnan 2006).

There is evidence for negative responses to inequity[2] in some non-human species. Capuchin monkeys and chimpanzees show negative reactions to receiving a less

[2] Note that it is not clear whether the monkeys perceive rewards that differ as *inequitable* or *unequal*. Although the rewards in the experiment were objectively unequal, differences in rank, hunger level, etc. may render even equal rewards inequitable. While I only refer to inequity in the context of rewards that differ in objective value, I prefer the term inequity in this context as we do not know whether these other factors influence expectations.

desirable reward than a social partner in some situations (Brosnan and de Waal 2003; Brosnan et al. 2005, 2010a; Fletcher 2008; van Wolkenten et al. 2007), as do domestic dogs (Range et al. 2008). All of these species refuse to participate in interactions when a social partner receives a better reward for completing the same task. This demonstrates that these species are capable of recognizing when their outcomes differ from those of others. This is expected given that these same species can learn socially, which should require similar cognitive abilities (Brosnan 2006). This recognition of inequity may pave the way for moral behaviors.

There is also evidence that some responses to inequity may drift closer to our understanding of human moral behaviors. In a situation in which rewards for a cooperative task (with a mutualistic payoff) differed, capuchin monkeys were much more likely to cooperate if they received their 'fair' share of the better reward (Brosnan et al. 2006). The monkeys had to work together to pull in a tray, but one of them got a better reward for doing so. Thus on an individual trial, one individual had the opportunity to receive a better outcome than the other (e.g., short-term inequity). However, the monkeys could alternate such that over the long run, both got approximately the same number and kind of rewards (e.g., long-term equity). In fact, this is precisely what they did; pairs in which receipt of the better reward alternated were more than twice as likely to keep cooperating as compared to those pairs in which one individual consistently claimed the better reward (leading to both short- and long-term inequity). Again, this behavior is not necessarily moral. By 'sharing' opportunities to get the better reward in the short term (a short-term cost), the dominant individual is essentially increasing his or her own outcomes in the long term (a long-term gain). Analysis showed that although dominants in these pairs got *relatively* fewer good rewards as compared to their partner, overall they received more rewards (of both higher and lower quality) than did dominants that monopolized the preferred rewards. This is evidence that individuals in some non-human species are willing to change their behavior to affect their partners' outcomes. I discuss this possibility in more detail in the next section.

5.2.4 Prosocial Behavior

Once an individual develops the capacity to respond negatively to personal inequity, she or he may develop a response against inequity directed at others (Brosnan 2006). This behavior, often referred to as prosocial behavior, may result in benefits to the partner at no cost whatsoever to the actor. Even in these seemingly simple cases, however, prosocial behavior may not emerge.

Much research has been done of late investigating how non-human species respond when given the opportunity to benefit partners at little or no cost. These studies make it clear that, even in the non-human primates, the behavior is surprisingly rare. With minor variations, the most common paradigm used to test this predilection is to present one individual, an actor, with a dichotomous choice. Both outcomes reward the actor to the same degree, but one outcome brings rewards to a social partner while the other does not. Most of these studies have involved

chimpanzees, who have not adjusted their behavior to take their partner's outcomes into account. Instead, they typically chose between the prosocial and selfish option at chance levels (Jensen et al. 2006; Silk et al. 2005; Vonk et al. 2008, although see Horner et al. 2011). Chimpanzees may not be the best species to use in studies of prosocial behavior, though. Three studies on capuchin monkeys have found that capuchins do take their partner's outcomes in to account (de Waal et al. 2008), even rewarding their partners with better foods than they received themselves (Lakshminarayanan and Santos 2008), at least if the inequity is not too great (Brosnan et al. 2010c). Finally, two callitrichids (a family of new world monkeys) have been studied, with different results. Marmosets do show an interest in their partners outcomes (Burkart et al. 2007), while tamarins do not (Cronin et al. 2009; Stevens 2010; although see Cronin et al. 2010).

The above are a very limited subset of prosocial situations, in which individuals are noted as prosocial only if they act to equalize their and their partners' rewards. This may be misleading, since failing to bring benefits to one's partners is not the same as failing to notice that one could do so. In one study of inequity, chimpanzees were more likely to refuse a better reward (a grape) if their partner got a less good reward (a piece of carrot) than if their partner also got the more preferred grape (although they did not give their partners the better grape; Brosnan et al. 2010b). This does not necessarily mean that they were 'protesting' to make the outcomes more equitable (they may have refused to avoid later repercussions), but it does indicate that they were at least recognizing situations in which their partners got lesser value rewards. Additionally, not all prosocial behavior occurs in the context of food sharing; chimpanzees also provide assistance to others. There are numerous anecdotes of chimpanzees assisting other individuals, including members of other species (de Waal 2006). In experimental studies, chimpanzees are documented to assist others in small tasks, such as returning a dropped item or giving another individual access to a desired locale. This is often called 'helping behavior,' whether those others are human experimenters or other chimps (Warneken et al. 2007; Warneken and Tomasello 2006; Melis et al. 2010). As chimpanzees both help others in these non-food contexts and clearly recognize when their partner receives a less good outcome than themselves, the failure of chimpanzees to bring food rewards to their partners may not reflect the full extent of their tendency towards prosocial behavior. Future tasks using different scenarios and outcomes are needed to more fully untangle the degree and extent of chimpanzee prosocial behavior.

Overall, primates apparently have the potential to be prosocial, but not all species engage in these sorts of behaviors. Even those non-human primates that do show similar actions do not do so consistently across situations. Regarding species differences, these primate species are all under different selective pressures, which presumably accentuated the behaviors to greater or lesser degrees. Regarding individual variation, there may be social (e.g., relationship) or contextual (e.g., the presence of food) cues that affect prosocial behavior in those individuals who show the propensity (see Burkart et al. 2007; Yamamoto and Tanaka 2009b for several hypotheses). As I will discuss below, what is critical for our understanding of the origins of human moral behavior is that there is the possibility for other-regarding behavior in non-human species.

5.3 Potential Mechanisms for Moral Behavior: Moral Emotions

Although the above examples provide evidence for behaviors that may be indicative of moral systems, in none of these situations do we know the underlying mechanisms. While neither behavioral outcomes nor mechanisms are individually the most important (they are equally important), both need to be considered for a full understanding of any phenomenon. Mechanisms provide addition insight in two ways.

First, there may be behaviors that result in the same outcome in multiple different species, but that have different underlying mechanisms. For instance, one species may show behavior that functions to equalize their outcomes with those of social partners in order to avoid negative repercussions, while another species may do so out of empathy, even if there are no potential negative repercussions. Although both actions resulted in the same outcome, most of us would consider the latter to be more moral than the former due to its focus on benefitting the other individual rather than avoiding a negative repercussion. Thus, understanding the underlying mechanisms may help to elucidate the degree to which a behavior is "truly" moral.

Second, it may be that a similar underlying mechanism is reflected differently in different species' behaviors. Two species may have an equal tendency toward empathy, but one instantiates this more frequently in food sharing situations, while another does so more frequently in helping situations. Although a consideration of only outcomes would imply that these species are quite different, the tendency towards empathy may indicate a high degree of similarity.

One mechanism related to moral behavior that has been proposed as common between humans and other species is the moral emotions. For instance, it has been argued that non-human primates, in particular, may show gratitude (Bonnie and de Waal 2004). Of the moral emotions, empathy may be the most important. Although it is possible to act in a moral way without understanding why, typically some degree of empathy or sympathy is considered essential. However, empathy in non-human species is an extremely contentious issue.

The vast majority of examples of empathetic behavior in non-humans are anecdotal (reviewed in de Waal 2006). This is problematic because without carefully controlled studies, it impossible to determine whether the underlying mechanism was truly empathy. A good case in point is the example of Binti Jua, the female gorilla at Brookfield Zoo in Chicago, IL. She picked up a young boy who fell in to her enclosure and took him to her keepers without harming him. Some see empathy in this example; Binti was herself a mother and may have understood that the boy was hurt and required help (de Waal 2006). But others point to the fact that Binti had been trained earlier, using a doll, to bring her baby to the keepers for husbandry purposes in exchange for food rewards. Thus, it is possible that Binti's only motivation for returning the boy was her previous experience of gaining a reward for doing so, without her understanding anything about the boy's needs. Of course, these explanations are not mutually exclusive. It is possible that Binti both understood the

boy's needs *and* was motivated by the prospect of a reward, or that she understood that the human boy required assistance *but* only knew what to do because of her prior training. This example highlights the problem when functional outcomes are the same, but potentially different motivations would reflect different underlying mechanisms.

On the other hand, there are behaviors that are more difficult to "explain away." There are examples of non-human individuals who seemed to recognize the need of another individual and help him or her, even when there was no obvious reward to the helping individual, either immediately, or in the future (e.g., the helped individual was not high ranking). For instance, Jakie, a young male chimpanzee, retrieved a tire full of water for an older female in his group who had been trying unsuccessfully to obtain it (de Waal 2006). Not only did he bring her the tire of water, without taking any himself (excluding selfish motivations), but he carried the tire carefully to avoid spilling its contents. Given her age and physical state, it was unlikely that the female would be much assistance to Jakie (e.g., there was little chance of a long-term gain through reciprocity), although there may have been reputational benefits to him for doing so (Subiaul et al. 2008; Russell et al. 2008). Note that even among humans there are few, if any, times when it is possible to fully rule out all alternative explanations for an apparently altruistic act (Batson 1991).

One area in which empathy has been studied in a more controlled fashion is consolation behavior. Uninvolved individuals will sometimes come up to former combatants and groom or sit next to them. Such behavior is argued to cement social bonds and soothe distress, particularly with regard to the one losing the contest (de Waal and van Roosmalen 1979). Intriguingly, this behavior has not been found in monkeys (de Waal and Aureli 1996), potentially indicating that it requires greater cognitive capacity than other related social behaviors, such as reconciliation, which are distributed more widely across non-human species (Palagi et al. 2008; Schino 1998; Wahaj et al. 2001). However, even among apes, the interpretation of the behavior's meaning is contentious. A recent study indicates that 'consolation' behavior may provide benefits to the consoler, reducing her or his stress as well (Koski and Sterck 2007; this is similar to the argument against true altruism put forth by Batson 1991). In this case, the behavior may have emerged because it provided a direct benefit to the consoler, but later the existing behavior was selected for another purpose, namely to encompass situations in which the motivation was benefitting the victim, bringing it in to the realm of a moral behavior. Monkeys are also very sensitive to each other's distress, and may respond in ways that appear to be aimed at alleviating it (Masserman et al. 1964; Colman et al. 1969).

Finally, it is again worth noting that empathy is not restricted to the primates. There are anecdotal examples of empathy among other large-brained mammals, such as elephants and dolphins (Bekoff 2000). More intriguingly, there is a series of studies showing that rodents are apparently attuned to each other's pain, indicating that a negative response to a group mate's distress may have ancient evolutionary roots. In particular, rats respond emotionally to others' distress (Church 1959), and will actively relieve this distress, even at the cost of sharing a preferred food reward (e.g., chocolate; Bartal et al. 2011). Among mice, the presence of a conspecific in

distress apparently increases the individual's experience of its own pain (Langford et al. 2006). The social relationship is important; this effect occurs more strongly between cage mates than strangers. There is also a genetic component (Chen et al. 2009), providing some of the first evidence that empathy can be directly affected by natural selection. As research extends to additional species, we may find that the primates, or even mammals, are not alone in their ability to show at least some forms of empathy to others.

5.4 Moral Behaviors in Non-human Primates

Humans have the most advanced moral system of any species. Nonetheless, there are behaviors present in other species that appear to be related to the moral behavior seen in humans. Such behaviors may show us the pathway through which moral behaviors evolved. While this does not necessarily mean that any species with empathy has a system of moral behavior, it opens the door to the possibility, and provides a better understanding of the conditions that may have led to the evolution of these behaviors.

At the most basic level, many other species besides humans must successfully negotiate social relationships in order to continue maintaining the multiple benefits of group living. Although clearly all of these interactions must ultimately function to benefit the individual, or else they would not have evolved, and in many cases are motivated by self-oriented concerns, in other cases other-oriented concerns may also play a role. For instance, individuals have to inhibit always taking the best available option, even if they are the dominant actor, in order to keep other group members around them. While this is self-serving for most individuals in most species, at some point the actor may inhibit taking the best outcome—even when he or she could do so—because it is in the best interests of another. Another example is policing behavior, in which dominants intervene on behalf of other members of their group, or loser support, in which individuals assist those who are losing a fight. Again, while these behaviors may originate because of the benefits to the dominants of maintaining group harmony or reputational benefits, they may later be used in situations that are not self-serving, and thus may qualify as moral or represent the precursors to moral behavior.

Another example of this is in the context of inequity, where subjects may come to prefer benefits to their partners. Several species have been shown to respond negatively when they get less than their partners (see above). While this does not match the model of a perfectly self-interested actor (as they are turning down absolute gains), it is also not an other-oriented behavior as the individuals are responding to their own less good outcomes. On the other hand, this sort of reaction may be critical for the development of a potentially moral behavior; responding when one's partner gets *less* than one's self. In such a case, the benefitted individual may turn down a reward that is better than his or her partner's. There are many possible explanations for this behavior. At the most self-interested end, subjects may respond

because they are afraid of future retribution from the less well benefitted partner. Subjects may also respond because they recognize that their behavior now may influence the willingness of their partner to continue interacting, hence influencing the subject's own long-term payoffs (e.g., Brosnan and de Waal 2012). Finally, subjects may respond because they perceive the outcome as "unfair" and are working to instantiate equity or equality, a behavior for which there is only currently evidence in humans. At all of these levels of explanation, the fact that animals are behaving in ways that functionally benefit others indicates that these may represent evolutionary precursors to human moral behavior.

A good experimental example of a potentially other-oriented behavior in a non-human is the capuchin monkey study in which subordinate capuchins refused to cooperate—even though they were guaranteed rewards, and those rewards were the same as their partners—if they were working with partners that consistently claimed the better reward when rewards were unequal (Brosnan et al. 2006). In this case, the subordinates' refusals cannot be attributed to self-interest, as a purely self-interested actor should always cooperate, since some rewards are better than none (particularly in the trials in which the rewards were the same). Thus, the only explanation seems to be that they were not willing to work with a partner who didn't share. Of course, this has two important implications. First, the subordinates were sensitive to *relative* outcomes that, by definition, take the partner's outcome into account. Second, any dominant that wanted to maximize its own outcomes needed to take its partners' lesser outcomes into account as well. Indeed, for pairs in which dominants did refrain from always claiming the better rewards, cooperation levels were quite high and these dominants received a higher number of rewards, both preferred and less preferred, than their non-sharing counterparts. The dominants' behavior was in their self-interest, but nonetheless implies that they understood the result of their behavior on their partner's actions and were adjusting their behavior accordingly, albeit to maximize their own rewards. This recognition of partners' outcomes and ability to adjust one's own behavior accordingly seems critical for truly moral behavior, and may pave the way for behaviors that are increasingly other-regarding.

Finally, of course, there are moral emotions that regulate such behaviors, and these may appear in other species as well. While quite difficult to prove empirically, one could argue that the strongest evidence of moral behavior in other species would be if subjects showed the same form of empathetic behavior toward their social partners as do humans. This is challenging to demonstrate, and most would agree that we do not yet have a smoking gun for empathy in other species, but circumstantial evidence is beginning to accrue. In controlled experiments, rats and mice behave in ways that appear to recognize the needs of their partners, and anecdotal data from the primates indicate the same. Despite the difficulty inherent in running such studies (most require the experimenter to induce pain or suffering, which is typically not considered ethical in non-human primate work) this will clearly be a fruitful avenue for additional research investigating such behaviors in other species.

Almost 150 years ago, Darwin first proposed that sociality was innate and might provide a framework for studying morality in any species. Recent data support his case; accruing evidence indicates that other species besides humans have complex

social behaviors that may provide evidence about the evolution of our own human moral systems. Humans' moral behaviors are distinct from those of other species, of course, but by understanding other species, we better understand the selective pressures which shaped the behaviors we see in ourselves today. While it is always important to keep in mind the limits of any research program, the evolutionary study of morality promises to open the door to a better understanding of morality, both in humans and other species.

Acknowledgements Funding to the author was provided by a National Science Foundation Human and Social Dynamics Grant (SES 0729244) and an NSF CAREER Award (SES 0847351).

Chapter 6
Where Do Morals Come From? – A Plea for a Cultural Approach

Jesse J. Prinz

In recent years, there has been an empirical turn in ethics. Using the methods of psychology, neuroscience, behavioral economics, and evolutionary modeling, we have been able to make progress on old philosophical questions about the nature of morality. For example, much recent research has lent support to the view that emotions are integral to moral judgment. Unsurprisingly, empirical research in ethics has tended to be reductionist: the loftiest aspects of human behavior have been related to simple mechanisms that can be identified in the brain. The implicated mechanisms, most notably emotion circuits, are also known to have homologues in other creatures. This fact, together with evolutionary theory and behavioral ethology, has helped promote the idea that there is an innate moral sense. Nativist accounts have always been popular in cognitive science, so this outcome can hardly be surprising. But we should be cautious about importing that approach into the moral domain. Moral diversity within human populations suggests that, at the very least, culture is an important variable in shaping morality, and it is a variable that we cannot afford to overlook.

My goal here is to make a plea for a cultural approach to empirical ethics. I will begin by reviewing what I take to be the main empirical lessons about how we make moral judgments. Then I will argue that judgments, so understood, are not universal in content. This will lead to a discussion of where moral judgments originate. The brief answer is that cultural factors, unfolding across time, are crucial for understanding the content of morality. This has implications for how to think about the biological contributions to morality and the processes by which moral values are acquired.

J.J. Prinz (✉)
Philosophy Program at the Graduate Center,
City University of New York, New York, USA
e-mail: jesse@subcortex.com

M. Christen et al. (eds.), *Empirically Informed Ethics: Morality between Facts and Norms*, Library of Ethics and Applied Philosophy 32, DOI 10.1007/978-3-319-01369-5_6,

6.1 What Is Morality?

6.1.1 Emotion and Moral Judgment

In order to understand from where morals arise, we need to know what morals are. By morals, I mean moral values. Values are long-standing evaluative attitudes or beliefs about what is good and bad. We evaluate many kinds of things: art, attire, wine, food, friends, manners, athletic performances, and so on. Typically, we evaluate things against standards, which include ideal features or exemplars, on the positive side, and objectionable features or exemplars on the other. To call something good or bad is usually to comment on its distance from a stored conceptualization of good-making or bad-making criteria or cases. For example, a wine might be judged as good if it has a balance of acidity and sweetness. In this respect, evaluative classification is like categorization more generally; it involves some kind of matching process. But there is also a crucial difference between evaluation and categorization.

To see this, notice that a person could taste a glass of wine and recognize it as such, without having any view about whether it is good. One can even discern a balance of acidity and sweetness without judging that this balance is good. Judging that such balance is good requires a *response* to it. To qualify as a positive evaluation, the response has to have a motivational force; it has to promote consumption of the wine. When we evaluate things positively, we are usually thereby attracted to them. Negative evaluation, in contrast, motivates avoidance, cessation, or withdrawal.

If evaluations are responses to recognized features, and those responses have motivational force, then it is natural to suppose that evaluations are *emotional* in nature. Emotions are responses to things that go beyond recognition, and emotions promote various forms of approach and avoidance. To evaluate a wine as good, it is plausible that the wine causes a positive emotion in us: a kind of pleasure. Alternatively, we might say that a wine is good without experiencing such pleasure. For example, we might suppose that a wine is good because the sommelier recommended it. But in such cases, our evaluations are deferential, or parasitic on another evaluator. The sommelier, we can presume, takes pleasure in good wine, or has at least mastered a list of preferences from someone whose pleasure is regarded as authoritative.

I think such emotional responses are the mark of the evaluative. Without emotional reactions, we can categorize, but we cannot appraise things as good or bad. A dispassionate appraisal is possible only by deference to a passionate judge. In philosophical jargon, such an appraisal would be a case of "mentioning" rather than "use."

Against this background, it is plausible to suppose that moral evaluations are also emotional. To judge that infanticide is bad is not just to say that it involves a certain activity (the intentional killing of a baby), but also to find the activity abhorrent. This simple observation lies behind a philosophical tradition called sentimentalism according to which moral values are sentiments (prominent defenders include Hume

1739; Smith 1759; Ayer 1952; Blackburn 1984). A sentiment can be defined as a disposition to have an emotional response. Thus, to have the value that infanticide is bad is to have the disposition to have an emotional response (of a kind to be described below) towards killing babies. Moral judgments are occurrent emotional states towards actions, and moral values are dispositions to make such emotion-laden judgments.

The sentimentalist tradition in philosophy has gained renewed support from cognitive science. Over the last 15 years, there have been numerous empirical studies investigating what goes on when people make moral judgments. These studies have varied tremendously in design and methodology, but they have converged on the conclusion that emotions are centrally involved in moral judgment (for a recent overview see: *Emotion Review*, 2011, volume 3). Neuroimaging studies have shown that emotion centers of the brain are active when people consider moral dilemmas (Greene et al. 2001), read sentences describing moral violations (Moll 2002a), view morally significant pictures (Moll et al. 2002b), or encounter morally questionable playing partners in economic games (Sanfey et al. 2003). Behavioral studies have shown that emotion induction causally influences moral judgments. For example, people make more severe judgments of wrongness when situated at a dirty desk or when smelling noxious odors (Schnall et al. 2008), and when they experience hypnotically induced disgust. Induction of anger through films or autobiographical recall can also lead to harsher judgments (Lerner et al. 1998), and induction of happiness can lead people to be more utilitarian in orientation, approving the violent sacrifice of one innocent person to save five people in danger (Valdesolo and DeSteno 2006). Working with collaborators, I have sought to replicate and extend these findings. We have shown that disgusting beverages make moral judgments harsher (Eskine et al. 2011), and that irritating music increases negative moral judgments, and uplifting music increases positive moral judgments (Seidel and Prinz 2013). All this suggests that people use emotions as information when they decide whether something is right or wrong: when asked to make a moral evaluation, people introspect and report the intensity of their feelings.

It also has been shown that emotions can lead people to make moral evaluations even when they can't produce reasons to justify those evaluations (Haidt 2001). In a pilot study on this theme, I was able to show that people harshly judge a child molester even when his victim is unharmed and has no way of recalling or being traumatized by the incident (Prinz in press). Such findings suggest that we report our moral values by introspecting on our emotional states. The degree of negative emotionality determines our assessment that something is morally bad, even in the absence of supporting reasons. This suggests that emotions are *sufficient* for evaluating something as bad.

Emotions may also be *necessary*. Individuals who have impairments in emotional responsiveness show corresponding impairments in morality. Criminal psychopaths, for example, show deficits in negative emotions, and also seem to treat moral rules as mere social conventions (Campagna and Harter 1975; Blair 1995). Individuals with frontotemporal dementia suffer from a diminished capacity to evoke emotional states and show a corresponding tendency to see morals as

conventional (Mendez et al. 2005). Such findings suggest that, absent certain emotions, we lose the capacity to make moral judgments. The personal *evaluation* that something is morally bad gets replaced by the social *categorization* that something is prohibited by the community.

These empirical results can be systematized by the sentimentalist theory of morality. Emotions seem to be sufficient and necessary for moral judgments, and that can be explained by assuming they are component parts of such judgments. The judgment that something is wrong *consists in* a negative feeling toward it. If negative feelings are introduced extraneously (e.g., by noxious smells), we will feel more intense emotions and report that we think things are more wrong that we would report under other conditions. If emotional responsiveness is diminished, things seem less wrong than they otherwise would.

This story about moral judgments can be extended to other kinds of evaluations. For example, recent neuroimaging studies suggest that emotions are involved in aesthetic judgments (Kawabata and Zeki 2004; Vartanian and Goel 2004) and that reduced emotionality promotes aesthetic indifference (Chapman et al. 1976). This raises a question: what distinguishes moral judgments from other kinds of evaluative judgments?

The answer I favor is that moral judgments involve a distinctive class of emotions. It has been shown that other-directed moral judgments characteristically involve anger, contempt, or disgust and these are tuned to different kinds of transgressions (Rozin et al. 1999). We become angry about crimes against persons, contemptuous of crimes against the community, and disgusted by crimes against nature. There are also self-directed moral emotions, which may also have different functional roles. Guilt seems to arise when we harm another person, and shame arises when we do something that others might regard as unnatural or grotesque (Prinz, unpublished data). I have proposed that moral values are constituted by sentiments that dispose us to feel anger, contempt, or disgust towards others and guilt or shame towards oneself. To have a moral value requires the disposition to feel both these other-directed emotions and self-directed emotions. Sentiments involving different emotions, or lacking in both the other- and self-directed dispositions do not qualify as moral judgments. Aesthetic values, for example, involve different emotions, and drinking bad wine may cause disgust, but it won't cause guilt or shame.

The picture so far can be summarized by saying that moral values are sentiments, and sentiments are dispositions to feel both the self- and other-directed emotions of a certain kind. The emotions I have been discussing can be classified as emotions of *blame*, since they are socially directed and punitive in nature. A person who is the target of anger, contempt, or disgust will feel punished in virtue of being regarded in these negative ways, and each emotion will also motivate behaviors (such as aggression, in the case of anger, or avoidance in the case of disgust) that are tantamount to forms of punishment. Legally proscribed forms of punishment, such as torture, execution, banishment, and incarceration, can be seen as social inventions that institutionalize the kinds of actions we might be inclined to carry out given our emotions of blame. In equating moral values with sentiments, I mean to suggest that

they are sentiments and nothing more. Thus, beyond the cognitive representation needed to represent a certain type of action (e.g., stealing), sentiments are sufficient for regarding that action type as wrong; to think stealing is wrong consists in our negative sentiment towards it. One might come to have many cognitively represented beliefs about wrongdoing (e.g., that stealing decreases social stability or impedes with autonomy), but there are best described as contingent theories about what makes things wrong, which, unlike sentiments, are neither necessary nor sufficient for having moral values.

In addition to the punitive attitudes that I have been discussing, there are positive moral values that revolve around praise, rather than blame. Praise and blame play asymmetric roles in morality. For example, we rarely praise people for conforming to moral rules, but we do blame people for deviating. Praise is usually reserved for supererogatory acts, such as charity, or other forms of self-sacrifice. Positive emotions, such as gratitude and esteem, are likely to underwrite the values that lead us to appraise such acts as good, but I will not survey those emotions here. My focus will be on moral prohibitions since, given the asymmetry, these are the mainstay of moral life.

6.1.2 The Content of Morality

I have characterized moral norms in terms of the emotions that arise when we make moral judgments. It is by means of these emotions that we can identify when someone is moralizing, even if their values differ from our own. To that extent, the characterization is content neutral. It does not define morality by its subject matter. This is important because, as we will see, people moralize different things. Indeed, almost anything could be moralized. We moralize interpersonal actions, thoughts, character traits, personal habits, self-presentation, and so on. Even things outside our control can be regarded as morally wrong; consider the Christian doctrine of original sin. That said, the sentimentalist framework presented here can be used to make some broad generalizations about the content of morality. Such generalizations have already been hinted at with the taxonomy of other-directed emotions.

Recall that anger, contempt, and disgust arise in response to different kinds of transgressions. In particular, they vary as a function of who is victimized by a transgression: anger is a response to crimes against persons; contempt arises in response to crimes against community; and disgust responds to crimes against nature. These broad categories can be further refined by reflecting on ways that persons, community, and nature can be assailed against. Consider, first, crimes against persons. This category includes physical harm, as when a person is hurt, mutilated, or killed. But the category also includes violations of individual rights. Rights, in the Western tradition, are usually regarded as entitlements: the right to own property, to free speech, to education, to choose a religion, and so on. Preventing someone from having something to which she or he is entitled is usually regarded as a moral wrong; entitlement itself is usually understood as a moral, not just legal, construct.

When this happens, anger is the dominant emotional response. Anger can also arise in response to violations of distributive justice. If a distribution is unfair, those who get less than their share have been victimized. Thus, unfairness is a crime against persons, and it incites anger.

Contempt arises when a transgression is construed as an assault against the community. This happens, for example, when someone disrespects authority. Disrespect to authority can threaten to undermine the structure of the community, even if no one is directly harmed. The community can also be threatened when someone fails to conform to a social status hierarchy. Stepping out of line (e.g., looking down on one's parents or the elderly) is viewed with contempt. In addition, each social class tends to view the others with a degree of contempt, and this may serve to keep classes in their place. Contempt is also the emotion that arises when there is a transgression against public goods, such as vandalism or cases where a politician embezzles public funds or violates public trust. Here, again, the community as a whole is harmed.

Disgust is the response to unnatural acts. In non-secular societies, such acts are usually construed as crimes against God or gods (Shweder et al. 1997). Within a religious framework, supernatural agents are the authors and regulators of nature, so crimes against nature are forms of sacrilege. Secular societies continue to regard certain acts as unnatural, even if there is no obvious human victim. This is especially true of acts that involve the body. For example, some sexual behavior is considered immoral in many societies, such as bestiality, incest (even if consensual), and exhibitionism. There are also norms governing appropriate appearance (e.g., gender specific attire, broad conformity to current clothing styles, appropriately groomed hair, and cleanliness). Minor violations may provoke ridicule, but more extreme cases are likely to provoke disgust. In addition, there are norms governing diet. Kosher laws are a non-secular example, but secular dietary norms are also easy to find: some cultures prohibit consumption of horses, animals that have been domesticated as pets, and insects, for example. The consumption of human flesh, even if the person died naturally, is also widely condemned, and, in all these cases, the emotion of condemnation is disgust.

These examples illustrate two things. First, the content of morality is highly varied. Many moral values have little to do with harm, and every aspect of human life can be subject to moral rules. Second, in some broad, metaphorical sense, negative moral values can be regarded as concerning actions that are directed against one of three categories: persons, community, or nature. These categories may turn out to exhaust the moral domain (e.g., can there by crimes against abstract objects?). Each category is governed by a different moral emotion. We also have moral values pertaining to things other than actions, such as sinful thoughts or vicious character traits, but these attitudes may depend on a connection to actions: thoughts and traits potentially affect behavior. Thus negative moral values can be captured by the schema:

An agent A's doing/having/being X is bad iff by X, A (potentially or actually) has an effect on victim V, where V is construed as a person, a community, or nature, and, depending on that construal, an evaluator E who so construes A's X-ing will

have the corresponding emotion of blame (anger, contempt, or disgust, if E is a third party, and guilt or shame if E = A)

This schema summarizes the foregoing discussion. It gives us an account of what moral values are, and we can now reflect on where they come from.

6.2 Where Do Moral Values Come From?

6.2.1 Is Morality Innate?

Psychologists and cognitive neuroscientists typically assume that they are studying universal facts about human nature. Studies of memory, attention, and reasoning are presented as revealing the laws of thought, akin to natural laws in other sciences. A typical study of memory span, for example, rarely begins with the qualification that this is how memory works among American college students, or whoever makes up the subject pool. The demography of the subjects is (roughly) indicated, but it presumed that demography has little impact on the results. The presumption rests on the view that these basic faculties of the mind are innate, and relatively unaffected by learning. There is, in other words, an implicit nativist bias in the way the sciences of the mind are typically pursued.

The nativist bias is also implicit in some of the empirical work on morality. Psychological and neuroimaging studies of moral cognition rarely look at culture as a variable (consider the citations in Sect. 6.1.1). This implicitly assumes that moralizing is part of the universal human bioprogram. Many of these studies say little about the *content* of our moral values and focus more on the processes involved in moralization. To that extent, they are neutral about the origins of our specific values, even if they are implicitly nativist about the mechanisms that allow moralization. Some other research, however, takes a stance on questions of content.

We can see that there are three basic positions one can take with respect to the innateness of morality:

- *Strong Nativism*: The content of our moral values in innately determined or strongly constrained.
- *Weak Nativism*: We have an innate faculty for acquiring moral values, but the content of those values is not strongly constrained.
- *Anti-Nativism*: We have no innate faculty dedicated to morality.

As I read the literature, Weak Nativism is often implicitly presumed, and Strong Nativism is sometimes explicitly defended. Anti-Nativism is a minority position, which is rarely implicitly or explicitly endorsed. I myself am a methodological anti-nativist, which means I think we should assume that a faculty is not innate until evidence leads us to say otherwise. In the case of morality, some researchers think the evidence supports Strong or Weak Nativism. I am not convinced. Here I will briefly consider some of the evidence (see also Prinz 2007a).

Let me begin with Strong Nativism. One research program that has a Strong Nativist orientation is the so-called moral grammar approach, which pursues an analogy between morality and language (Mikhail 2000; Hauser 2006a). It is ironic that defenders of this approach tend toward Strong Nativist positions; given that language is generally regarded as weakly innate (languages vary hugely in phonology and vocabulary). Officially, defenders of moral grammar say morality can vary too, but much of their research is designed to establish universal moral content. Notably, Mikhail and Hauser have acquired evidence that most people respond in predictable ways to a range of "trolley dilemmas," in which an agent performs an action that leads to one person's death in order to save five others. For example, most people think it is wrong to push someone into a runaway trolley's path in order to save five people further down on the tack, but it is permissible to divert a runaway trolley onto a track where it will hit one person instead of five. Responses to such dilemmas are cross-culturally robust, but people have great difficulty articulating the principles on which they are relying. Nativists interpret this as evidence for unconscious rules, analogous to those used in language processing.

Trolley experiments, however, can also be interpreted in other ways. The fact that people in different cultures give similar responses might be explained by prototype effects. When people learn the concept *murder*, the paradigm cases involve direct intentional physical assault, not indirect harms. The reason for may have nothing to do with innateness. All cultures must have rules to stop people from directly and intentionally aggressing against each other, on pain of societal collapse. Rules against indirect harms, however, are less prevalent, because there are fewer circumstances within a society when indirect actions will result in someone's death, and a society that failed to have such rules might be relatively stable. The pushing scenario conforms most closely to the kind of actions that every society is likely to condemn. It is more clearly an instance of murder than the scenario in which a person is killed as the side-effect of diverting the trolley. In the "diversion" scenario, the death is also less salient and the cause of death for the one person is rendered comparable to the cause of death for the five, making the comparison between the two outcomes vivid. So there need not be any unconscious rules at work here. People are taught that murder is wrong by means of prototypical cases, and they tolerate killing more readily when it departs from the prototype, lacks salience, or is rendered comparable to an alternative action that involves the same kind of killing but greater losses.

Another research program that is committed to some degree of strong nativism is the moral domains theory of Turiel (1983). Turiel argues that genuine moral rules involve harms, and that other kinds of rules are mere conventions. In comparison to conventional rules, rules pertaining to harms are treated as more serious and less dependent on authority. Turiel believes that this pattern of conceptualization is innate. But there are five reasons for rejecting this position. First, harm norms are judged to be authority dependent in some studies (Kelly et al. 2007). Second, norms pertaining to diet, sexuality, and hierarchy are treated as equally serious by some groups (e.g., Vasquez et al. 2001, on Filipinos; Nisan 1987, on Palestinians). Third, there is a simple learning story available to explain why moral norms are treated

differently than conventional norms. Moral norms are taught by emotional conditioning, and once emotional attitudes have been internalized, the norms feel serious (i.e., emotionally evocative) and somewhat independent of authority (i.e., we are conditioned to feel emotions towards these acts even if we are in a community where others don't have such emotional dispositions). Fourth, there is massive cultural variation in attitudes towards harm. Many societies have practiced slavery, corporal punishment, judicial torture, agonizing body modification, blood sports, animal cruelty, spouse beatings, and virtually unconstrained brutality against out-groups; hardly evidence for an innate prohibition against harm. Finally, the fact that many societies do have moral norms against some forms of harm (notably gratuitous harm against the in-group) can be explained by the fact that we devise such prohibitions as a condition on societal cohesion. It does not take innate mechanisms to realize that tolerated killing will lead to social unrest. The fact that such norms have a highly moral status worldwide may also reflect the fact that anger is a natural response to aggression in the first-person case. Given that we are all disposed to get mad when others try to harm us, it is not surprising that the more general stricture against harm, which extends to third parties, is grounded in anger. This grounding helps give harm norms their moral cast.

Another research program that has a Strong Nativist flavor is evolutionary ethics. Evolutionary ethicists admit that nativism is compatible with moral diversity (e.g., Krebs 2008), but they tend to offer evolutionary models that emphasize highly predictable behaviors, suggesting that morality may be strongly constrained. Most of this work focuses on altruistic behaviors, in which individuals incur costs to benefit others. Models that use iterated economic games have shown that cooperative strategies, such as reciprocal exchanges, increase fitness, suggesting that cooperation may be an evolved response. The evolutionary interpretation gains support from the fact that general purpose reasoning, together with hyperbolic discounting, does not predict cooperation. Reasoning would lead people to see the value of defection, yet we do, in fact cooperate. Other prosocial behaviors, such as helping people in need and sharing resources, are also widely documented. Like cooperation, these behaviors are hard to explain by appeal to reasoning, which suggests that they may be innate. The evolutionary approach is bolstered by ethological research on non-human primates. Monkeys and apes are known to reciprocate, share, and help (de Waal 1996; Brosnan and de Waal 2003; Hauser et al. 2003). It is presumed that these behaviors are unlearned in our primate relatives and may reflect hardwired precursors to our own prosocial tendencies.

There are several reasons to resist the evolutionary approach to morality. First, most of the work concerns moral behaviors, not moral judgment. By that, I mean behaviors that we now happen to regard as morally praiseworthy (cf. Joyce 2006, on this distinction). In principle, a species could evolve to act in ways we find praiseworthy without evolving a capacity to praise. That is, there can be moral conduct without moral judgments. This point is especially problematic when it comes to extrapolating from animal research, since most of that work concerns "altruistic behaviors," and not moral judgments per se. Moral judgments have two features that are unlikely to be found in many other species. First, they require a disposition for

self-directed emotions. Evidence for guilt and shame in non-human primates is scant at best. If apes get angry when conspecifics trespass against them, it does not follow that they would feel guilty for trespassing themselves. Reactive aggression is not the same as forming a moral judgment; self-directed dispositions are needed as well. Second, there is only a little anecdotal evidence that non-human primates have concern for third parties. Moral rules quantify over agents and action types. They are not restricted to the second-person. If apes get angry when conspecifics trespass against them, it does not follow that they would get angry if one conspecific trespassed against another, especially a non-relative. If they do not do this, then their anger reactions don't stem from values that have the schema indicated above.

A second problem with animal models is that there are profound differences between apes and humans. Chimps often fail to share with long-time companions, even when there is no cost (Vonk et al. 2008), and they are often highly aggressive in the wild. Goodall (1986) documents cases of chimpanzee warfare, calculated murder, infanticide, and cannibalism. Wrangham et al. (2006) report that chimpanzees are alarmingly violent; comparing several wild populations to a small-scale human group known for aggression, the found male chimps were 384 times more likely to engage in a violent attack than were their human counterparts. One might reply that apes simply having a different morality than ours, but given these differences, the burden is on the nativist to say why ape behavior must be interpreted as based on moral judgments, as opposed to some other kind of motivations. After all, not every kind human act is a result of morality (threat of punishment, instrumental gain, friendship are among other motivators). This is not to deny that some forms of ape altruism might have biological roots in common with our own, but only to emphasize that we must be cautious about over-attributing human-style moral tendencies to apes. There may be important discontinuities.

Moving beyond comparative research, evolutional theorizing suffers from another limitation with respect to Strong Nativism. Evolutionary models have shown that it is difficult for altruistic behaviors towards non-kin to evolve through individual selection. If I mutate to reciprocate, but you do not, I will suffer a profound decrease in fitness. This has led to a widespread endorsement of group selection models. But group selection raises the possibility that widespread reciprocity evolves culturally, rather than biologically. Of course, nativists can offer alternative explanations that avoid group selection, but once such models are shown to be viable the pressure to explain altruism biologically decreases. More generally, there is something suspicious about any argument that moves from a demonstration of fitness enhancement to a conclusion about innateness. Many behaviors that would enhance fitness are not evolved; over generations, groups can learn to perform actions that are beneficial and avoid actions that are harmful. To show that morality is innate, models are not enough. Evidence must also show that specific moral rules are *universal* and *unlearnable*.

With respect to universality, evolutionary approaches tend to suffer from a dearth of empirical support. The models might be taken to suggest that all people are equally altruistic, but, in reality, there is considerable cultural variation. Sharing, for example, varies with respect to competing principles of distribution. In America, the

preferred principle is equity (distribution as a function of achievement), in China there is a preference for equality, and in India there is a preference for distribution as a function of need (Leung and Bond 1984; Berman et al. 1985). It is hard to think of sharing beyond one's kin as a biological norm given the rise of global capitalism, widespread opposition to taxation, and staggering discrepancies in wealth. Similar conclusions can be drawn about helping. Trivial, low-cost, helping behaviors, like picking up a pen that some has dropped, differ dramatically from place to place, with Rio residence coming out on top and New Yorkers bringing up the rear (Levine et al. 2001). Cultures also vary in the degree to which helping the needy is seen as a cultural requirement. In the United States, helping strangers with moderate neediness is considered entirely optional, but it is morally mandated in India (Miller et al. 1990). Americans, unlike Indians, also seem to think the obligation to help someone in moderate need depends on whether we like that person (Miller and Bersoff 1998). In general, we do amazingly little to help the needy. Preventable diseases claim about nine million lives a year, as does starvation, suggesting an annual toll that dwarfs the holocaust, and nearly universal crimes of omission.

Finally consider learnability. Evolutionary ethics presumes that we would not engage in prosocial behavior if we relied on domain-general resources such as reasoning. Given the human tendency to discount the future, we would behave unethically to reap short-term rewards. The fact that we are generally pretty good to each other is taken as evidence that morality is innate. Here again, one wants to distinguish moral behavior and moral attitudes. After all, squirrels are pretty good to each other, but no one thinks they have innate morality. But putting this issue aside, one can also deny the premise that domain general resources would not lead to cooperative behavior. It is true that reasoning might not be up to the task, but emotions are well suited to this purpose. Suppose I fail to cooperate with you and you get mad. I may be frightened of punishment or sad about losing you as a partner. Thus, your anger can condition me to associate negative emotions with defection. Suppose now there is an opportunity for me to defect without you finding out. Reason might lead me to do so, but emotions operate somewhat independently of reason, and my negative associations may promote cooperation even in this situation where free-riding is an option. Notice that this appeal to emotions as mechanisms of cooperation is also central to evolutionary models (Trivers 1971; Frank 1988). The point here is that once we recognize that emotions are the glue that promotes prosociality, there is actually less pressure to assume that morality is innate, because emotional dispositions can be easily learned through conditioning. Emotions may be evolved for selfish purposes (anger protects us against threats, and sadness makes us withdraw in times of loss), but selfless dispositions can arise when these selfish patterns are conditioned by interactions with others. Your rage becomes my loss, so I learn to avoid making you angry.

Expanding this last point, the acquisition of prosocial behavior needs two ingredients. First, if I defect in my dealings with you, you will get mad. That's not a moral response; it's just reactive aggression. Second, if you get mad, I feel bad and associate this with defection, leading to increased tendency to cooperate. These two steps could even be realized in non-human primates. Human beings may go on to a

third step: we generalize moral rules and apply them in cases where we have no direct involvement. This might be explained by the fact that human beings have two capacities that are underdeveloped in primates: imitation and abstract thought. Imitation leads us to mimic the reactive aggression of those who get mad at us. Abstraction leads us to internalize emotional dispositions in a way that can generalize across individuals, because we can represent actions abstractly rather than merely first-personally, as something I do. Thus, if you get mad at me for defecting, I might come to have bad feelings about defecting in general, whoever does it, and I might adopt your anger response when I encounter the defection of another. I don't want to suggest that this is the whole story. There may be innate behavioral tendencies that contribute to the moral rules with which we end up. But these simple observations suggest that the acquisition of moral rules need not involve any highly specialized mechanisms.

This last point allows us to move from Strong Nativism to Weak Nativism. Strong Nativists claim that the content of morality is innately determined or strongly constrained. I have tried to cast doubt on that conclusion by briefly reviewing some of the leading research programs that emphasize innate content. The content of moral rules is variable, and convergence can be explained without innateness. Now, with this simple story about psychological prerequisites to morality, we can see that even Weak Nativism may be mistaken. The acquisition of moral rules may not depend on any kind of morality acquisition device (Sripada and Stich 2005), but may instead derive from cognitive resources that evolved for other purposes (emotions, imitation, abstraction). Far more would need to be said to firmly establish that domain general resources are up to the task. For present purposes, I am content with the conclusion that we should be open to this possibility. Just as religion may arise in all cultures without a religion module, morality may be a byproduct of capacities that are not specific to the moral domain. As a methodological anti-nativist, I'd like to see more evidence for domain specificity before concluding that morality is even weakly innate.

6.2.2 Morality, Culture, and History

I just reviewed evidence for moral nativism and found it wanting. I also indicated some of the proximate psychological mechanisms that may be involved in the acquisition of moral rules. But what about more distal factors? Why do we have the rules that we do? If I am right, the answer to this question cannot be given solely by evolutionary theory, but must recruit the resources of cultural anthropology and history. The factors that give rise to moral rules include our social circumstances, some of which are widely shared across human groups, and some of which are more particular.

The inclusion of history in the study of morals is not new. Philosophers have long speculated about how historical factors have shaped moral values, and many leading ethicists have offered historical accounts. Prominent examples include Hobbes,

Rousseau, and Hume. The stories we find in these authors' works are in some sense fanciful, however. They offer highly speculative accounts of why values might emerge from an initial state of nature, in which moral values as we know them do not exist. No evidence for these stories is offered; they are inferred from specific views about how people act in their natural state. In the *Leviathan*, for example, Hobbes tells us that human beings are naturally selfish and violent, but relatively equal in strength, which means the state of nature is a war of all against all. Morality emerges as a solution to this unhappy form of life. Taken as an empirical hypothesis, the Hobbesian account might be investigated by analyzing our natural tendencies towards aggression (a psychological thesis), and the role of the state in reducing interpersonal conflict (a historical thesis). Some empirical evidence sits well with Hobbes. For example, Wrangham (2004) documents extreme violence in small scale societies, and Pinker (2007) argues that violence has been on a steady decline. On the other hand, the Hobbesian idea of a state of nature may be a fiction. Our species is social and has always lived with socially negotiated norms and Hobbes may also exaggerate our tendency toward violence, which is counterbalanced by a tendency to look out for members of the in-group. The claim that states have served to reduce violence is hard to reconcile with mass-scale war, imperialism, and slavery, even if recent times have seen a significant decline in mortality rates. In any case, it should be clear that empirical evidence could be brought to bear on this and other historical accounts within philosophy.

Hobbes, Hume, and Rousseau are interested in how we arrived at morality from a pre-moral position. That is an interesting question, but one which hinges on a confusion if humans form social groups by nature: There may be no pre-moral position. These approaches also pose the historical question at a high level of generality, asking about the origin of cooperation, justice, or morality in general, rather than specific norms. As such they offer little insight into why cultures have different moral values, values that can even be diametrically opposed. The philosopher most famous for addressing this question is Nietzsche, whose *On The Genealogy of Morals* (1887/2009) offers a historical conjecture to explain why Christian morals differ from values documented in ancient Rome. Nietzsche offers philological evidence for his thesis that Christians inverted the Roman value system, and he relies on basic historical facts and psychological conjecture in supposing that this inversion might have occurred because the Christians had been enslaved by the Romans. When the Christians gained power, their resentment towards their former oppressors led to a moral inversion in which Roman ideals of the good, such as flourishing, were replaced by a conception of the good that includes asceticism and guilt. Again, these are empirical claims. Is Christian morality driven by resentment? Were Christians serving as Roman slaves? There is some evidence that Nietzsche got it wrong (Prinz 2007b). The Christian revolution might have been driven by middle-class Roman converts, who were predominantly female and wanted to achieve a better life.

In any case, Nietzsche's "genealogical" approach points to an under-developed resource in studying morality. Some philosophers, most notably Michel Foucault, have offered genealogical analysis to explain contemporary values and moral

variations across time and place. But there has otherwise been little uptake of the Nietzschean approach within philosophy. Within cognitive science, the story is similar, with disproportionate resources funneled into evolutionary accounts, which do better at explaining moral universals than moral differences.

One reason for this resistance to genealogical approaches is that they may appear to be unscientific in an important sense. Science specializes in generalization, and many historical developments seem to depend on one-off events, rather than repeatable laws. For example, the specific styles of art that emerged in Europe during the course of the twentieth century reflect non-repeatable historical events and innovations by individual artists. Cubism arose, in part, because the invention of the camera freed the artist from the fetters of realism; futurism arose in part because of the rapid rise of technologies of speed; Dadaism emerged in the wake of the first world war; and so on. Some moral rules are like this, including Nietzsche's case study of Christian values. But, in many cases, the factors that influence moral values are repeatable and repeated in different historical contexts. In those cases, we can see that there is room for a cultural science of moral norms. To illustrate, let's consider some examples.

Cannibalism: Cannibalism is now reviled as the most evil activity that a human being can engage in, but is has been practiced by many societies across the globe throughout history. In one sample, more than a third of historically documented societies engaged in some form of cannibalism (Sanday 1986). Even the Christian Eucharist can be seen as a residue of a practice that was once more widespread. Given this variation, it would be nice to explain why some cultures engage in cannibalism and others do not. Harris (1977/1991) offers an explanation that appeals to three factors: size, subsistence, and resource availability. Hunter-gatherer societies who compete with neighbors over resources often end up in violent conflicts (Wrangham 2004). Victors in those conflicts end up with dead bodies and prisoners. From a cost benefit analysis, it makes sense to eat dead bodies, since they are a source of good meat and meat is hard come by. It also makes sense to kill the prisoners since it is too costly to enslave them. That means more dead bodies, which should also be consumed. Harris argues that cannibalism disappears with the rise of state scale societies. States have the power to form armies, which can collect taxes or tribute money from neighbors. States also tend to engage in trade relations, and have agriculture and domesticated animals, which minimizes resource competition and the need for hunted meats. Eating your neighbors is no longer advisable when they are trade partners and tax payers, so cannibalism tends to disappear with societal development.

Marriage: Marriage is a moralized institution. We consider some kinds of relationships acceptable and others unnatural or morally dubious. In contemporary Western societies, monogamy in morally preferred. When politician or golf stars stray, they lose votes and commercial sponsors. But, when we look beyond the West, more than 80 % of societies allow polygyny (Murdock and White 1969), so our moral attitudes toward indiscretion make us cultural outliers. Monogamy in Western Europe may result largely from a historical accident. Under the early Christian Church, there

was a sweeping set of reforms, which had the net effect of reducing the number of sexual partners by curtailing premarital sex, divorce, concubines, and polygyny. These policies reduced family size and led to increased heirlessness, which meant more money was donated to the Church, allowing it to spread its reforms farther and farther (Goody 1983). But monogamy is unusual because many common factors promote polygyny (see White and Burton 1988): Male-centered living arrangements favor male control over resources (e.g., patrilocal households), giving men opportunities to control women's lives; female contributions to subsistence, especially domestic contributions, make women a "commodity" worth collecting for men; room for territorial expansion promotes families with a large number of offspring, which again favors polygyny; warfare, which increases male fatalities and increases the female to male gender ratio promoting many-to-one marriages; warfare for plunder, which includes capture of wives can affect gender ratios and allow young men to avoid paying for brides, promoting a further increase in polygyny; restrictions on female property ownership and competition in open labor markets makes women depend on men, creating a gender asymmetry that compels women to accept plural marriages. Given widespread male dominance, it is not surprising that polygyny is the norm. But the degree of polygyny diminishes as these factors decline. For example, polygyny tends to decline with lifestyles that are less conducing to expansion, including fishing, some forms of farming, and urbanization. The Romans who were highly urbanized made monogamy the law. In settings where expansion is particularly limited, polyandry may even arise, as in traditional Tibet and Nepal. In contemporary Western culture, there is no a widespread move to allow gay marriage, which may stem from the fact that contemporary economic systems make it profitable, for the first time, to have fewer children (Werner 1979). Heterosexual couples are also marrying later, and wealthy families are having fewer offspring than the poor. Gay marriage may be part of this same syndrome.

Incest: Cultures also vary in the degree to which they permit marriage within the family. There is probably a biological predisposition to avoid some forms of incest, but only 44 % of societies have explicit incest taboos (Thornhill 1991). The presence of these taboos and the severity of the punishment correlate with social stratification, suggesting that moral sanctions against incest arise to prevent families from consolidating wealth and moving up the social ladder. There is also cultural variation in what counts as incest. The Christian Church prohibited cousin marriage up to the seventh degree, but in the Islamic world cousin marriage is encourages. In contemporary Saudi Arabia and Pakistan over 50 % of married couples are cousins (Bittles 1990). This may have to do with the fact that power is distributed across clans in such societies, rather than centralized, as under the Christian Church. There are also conditions that favor sibling incest. This is well documented in royal families, who want to retain wealth and avoid forming obligations to other families and groups. In Ptolemaic Egypt, Greco-Roman citizens had sibling incest rates up to 30 %, presumably to avoid having to intermarry with the Egyptians whom they had conquered (Shaw 1992).

Slavery: Many societies allowed slavery, and the anti-slavery movements of the eighteenth and nineteenth centuries were virtually unprecedented historically,

especially when considering large-scale societies. Large-scale societies often placed restrictions on who could be enslaved (outgroups, rather than ingroups), but, until recently, there has been widespread consensus within such societies that slavery in some form was permissible. Small-scale societies tend not to have slaves because they cannot feed or police slaves effectively. But when state-scale societies emerge, usually though the innovation of agriculture and food storage technologies, surplus resources and power differentials arise, and labor demands increase. This makes slavery cost-effective. Goody (1980) reports that only 3 % of hunter-gatherer societies have slaves, as compared to 43 % of societies with advanced agriculture and 73 % of pastoral societies. Economic advances gave rise to new needs (e.g., a need for a large class of laborers who lack upward mobility), new opportunities for the powerful to pursue self-interested desires (e.g., obtaining fully submissive sexual partners), and the technological and human resources needed to wage war against weaker neighbors, resulting in a class of conquered captives. Given this pattern, slavery is a likely outcome of economic growth. It is surprising, then, that slavery was ultimately banned in many parts of the world, and the primary cause may have been the industrial revolution. Proponents of the anti-slavery movement in England, which helped spark reforms elsewhere, argued that an economy based on wage labor would be more profitable. In the end they were probably right. The argument was harder to sell in the United States, where slave cotton constituted up to 30 % of the U.S. economy (Davis 1984), but Northern manufacturers who had an opportunity to change the balance of power from the agricultural South had some incentive to end slavery, and that may have contributed to the American Civil War.

Torture: Judicial torture was once widely practiced in Europe. Torture was often horrifically cruel and sometimes observed by the public. It was used to extract confessions, and, less frequently, as a form of punishment. Torture is still practiced in some countries today, and Western nations occasionally debate whether certain forms of torture should be legally permitted, but there is a wide consensus now that torture is wrong. In the eighteenth century, torture came under heavy criticism and mostly disappeared (Beccaria 1764). There had been critics of torture before Beccaria, because it was often administered at the whim of lay judges, but the eighteenth century brought a more dramatic shift in thinking. Slavery was not just something that had to be carefully regulated; it came to be regarded as fundamentally wrong.

No one knows exactly what caused this shift, but several factors may be relevant. As one example, Europe endured massive losses during the 30 Years War (almost 10 % of the population died), and people were weary of violence. That, and subsequent brutal wars, helped fuel contempt for governmental use of violence, sowing the seed for an anti-torture sentiment. In the following century, there also was a shift from monarchy to more democratic forms of government. This meant that governments were, for the first time, "of the people." When a state is led by a monarchy, it needs to establish authority, and violence is one method of doing so. When a state is led by the people, there is less need to establish authority, because the people have no difficulty granting authority to themselves. Thus, democracy may have bolstered negative attitudes towards torture.

Another variable is the perception of a foreign threat that has penetrated the sanctity of the state (Thurston 2000). European torture was often directed at people accused of heresy or witchcraft, which was regarded as a kind of supernatural invasion from within. In more recent times, torture was used during the Soviet Terror of the 1930s, under paranoid suspicion that counter-revolutionaries were secretly operating from within to undermine the state. Torture was practiced during Argentina's Dirty War, which was fuelled by fear of an internal communists threat. As part of the War on Terror, the U.S. used torture techniques against alleged foreign enemies who allegedly conspired to commit violent acts on American soil.

6.2.3 Implications

Examples of the foregoing kind are easy to multiply. They illustrate several important points. First, there is a tremendous amount of moral variation. Each value endorsed by one culture is rejected by others. This shows that morality is plastic. There are dramatic cultural differences concerning who is deemed morally worthy and in the appropriate treatment for those designated as unworthy. Thus, we must move beyond nativist and evolutionary approaches if we are to understand the beginnings of morality.

Second, moral values are essentially historical. Each has a genealogy. Thus, history is an important tool in explaining morality. Third, though many cross-cultural differences result from specific historical events, others can be explained by appeal to variables that re-appear across time and space.

For these reasons, there can be a cultural science of morals, tracing factors that can lead to the emergence and retention of some values and disappearance of others. Research on cultural evolution has moved in this direction. Cultural evolution refers to the idea that cultural items are subject to pressures similar to natural selection. Cultural items, including moral norms, vary in their degree of fitness (i.e., their likelihood to be passed on to the next generation). Fitness here can include biological fitness because some norms lead to greater reproductive success. But it can also include psychological fitness since some standards are easier to learn or more catchy. Norms that increase the power of norm-disseminators can also be said have a high degree of cultural fitness, such as norms that increased the coffers of the church. Given this broad notion of fitness, it is important to see that cultural evolution differs from biological evolution, but both forms of evolution illustrate how historical processes might be characterized by general principles, and are thus amenable to scientific inquiry.

It does not follow from this that human plasticity is open-ended. Perhaps some moral rules are easier to learn than others and some might even be impossible to sustain. Morality is no doubt constrained by our biological endowment. The emotions we have, our capacity to attribute mental states, and our care for kin all serve as building blocks that help shape the outcome of norm construction. The anti-nativist does not postulate a blank slate. But the biological constraints should not be

mistaken for a moral sense. They may constrain morality the way human visual capacities and emotions constrain the arts.

Thus, the scientific study of morality should not be limited to psychology, neuroscience, ethology, and biological evolution. It should expand to include anthropology, history, sociology, and other fields that track sources of cultural variation. A complete science of morality will work at multiple levels. Material factors will influence cultures, cultures will affect moral education, moral education will tune emotions, and emotions are implemented by circuits in the brain. Evolved human biology will contribute to this story, by shaping behavioral predispositions and the affective and cognitive faculties that allow us to internalize moral values. But this should not lead us to adopt the kind of reductionism that construes the moral faculty as a historical. To do so would be to overlook the most distinctive aspect of human psychology: how we think is affected by institutions that we create and transmit socially. Moral variation over time and the conflicts that divide the world today can be understood only if we overcome nativist biases and look at morality through a cultural lens.[1]

[1] I am deeply indebted to Markus Christen, Carel van Schaik, and an anonymous referee for enormously helpful comments.

Part III
Assessing the Moral Agent

Chapter 7
Moral Intelligence – A Framework for Understanding Moral Competences

Carmen Tanner and Markus Christen

While virtues, moral values and concerns have always been an inherent theme of philosophy, moral concerns in society appear to pace up and down. Mostly, there are particular events (such as military interventions, terror attacks, natural catastrophes, business scandals) or the development of new methods and technologies (such as cloning, stem cell research, and biotechnology) that lead to publicly recognized moral crises or moral hazards. As such, they can induce "moral revolutions" that result in changes in social practices (as e.g., the abolition of Atlantic slavery; Appiah 2010). No doubt, what has given rise to a new wave of moral crisis more recently are the corporate ethical scandals and the financial crisis that have shocked the business world. Business practices are again heavily scrutinized and many people are asking what can be done to promote moral behavior and to prevent similar transgressions in the future.

When discussing interventions, promoters of moral change typically refer to the content of moral standards or values. They often advertise new moral guidelines, codes of conduct or a set of virtues that individuals (e.g., business leaders) or institutions should adopt to enhance moral behavior. Indeed, moral change sometimes simply results from a change in the meaning of behaviors or practices during history. Some practices that were non-moral became heavily moralized (as with the example of slavery, Appiah 2010), whereas other behaviors that were considered "bad" lost their moral blemish (e.g., homosexuality). Some authors also argue that expanding the "moral circle" (Lecky 1869), i.e., the domain of entities or creatures

C. Tanner (✉)
Department of Banking and Finance, Center for Responsibility in Finance,
University of Zurich, Plattenstrasse 32, 8032 Zurich, Switzerland
e-mail: carmen.tanner@bf.uzh.ch

M. Christen
Institute of Biomedical Ethics, University of Zurich,
Pestalozzistrasse 24, 8032 Zurich, Switzerland
e-mail: christen@ethik.uzh.ch

M. Christen et al. (eds.), *Empirically Informed Ethics: Morality between Facts and Norms*, Library of Ethics and Applied Philosophy 32, DOI 10.1007/978-3-319-01369-5_7,

that should become subject to moral consideration, is a prerequisite of moral progress (Singer 1981).

Such content-based approaches that rely on the semantics of moral terms rarely suffice to encourage moral transitions. Changes do not just require new moral content, they also require agents who are skilled in how to deal with moral issues, once identified, and how to turn moral standards into actions. Of course, individuals are embedded in complex socio-cultural structures which facilitate or inhibit some developments. But humans are neither totally autonomous, nor passive in responding to the environment (Bandura 1991). They are active moral agents endowed with some capacity to control themselves and the environment. Scholars and practitioners alike have therefore agreed on the view that improvements in the propensities and abilities of moral agents to cope with moral contents are crucial in fostering moral transitions (Dane and Pratt 2007; Narvaez 2005; Pedersen 2009; Reynolds 2006; Treviño and Brown 2004). Hence, efforts to which abilities are important and how to explain and measure individual differences in those abilities are essential.

Drawing from current literature and research, the goal of the present work is to specify the abilities that facilitate moral functioning. In doing this, we refer to the concept of Moral Intelligence. Moral Intelligence (MI) refers to the agent's capacity to process and manage moral problems. To our knowledge, Lennick and Kiel (2005) were the first to introduce this term. They referred to the business world and, based on case studies, concluded that mere strategic thinking is not sufficient for being a successful business leader. In addition, even though researchers and practitioners alike recognized in the past emotional intelligence as an encompassing, useful and advantageous capability, MI puts an emphasis on moral skills and heralds the examination of a new facet of intelligence. Recent approaches have provided compelling arguments that moral agents do require several abilities, but the approaches differ in terms of which skills and subskills are considered as relevant (Lennick and Kiel 2005; van Luijk and Dubbink 2011; Narvaez 2010a; Rest 1986). Building on this work and our own perspectives, we will highlight a small but essential set of moral abilities.

In this chapter, we put forth a theoretical framework of MI that integrates moral decision-making with concepts and topics of social cognition and self-regulation theory. We start our work with defining MI and then present a moral process model that provides the foundation of the MI framework. Afterward, we introduce the elements and moral competences that we deem as essential for moral agents. Finally, we briefly present some ideas for how to enhance MI.

7.1 Defining Moral Intelligence

We define Moral Intelligence as the capability to process moral information and to manage self-regulation in any way that desirable moral ends can be attained. Our picture of a morally intelligent person is someone who is endowed with a desire to strive for moral goals and to use moral principles and self-regulatory skills to do

what is good for society, other human or nonhuman beings. This definition expands Lennick and Kiel's (2005) initial conception, according to which MI refers to the capacity to apply universal moral standards to one's values, goals and actions. Despite Lennick and Kiel's seminal effort in stimulating attention to MI, their framework does not specify underlying processes and mechanisms. If we want to understand, teach and encourage MI, however, we need an understanding of the basic mechanisms involved in moral functioning.

To explore MI, we make use of social cognition and self-regulation theory, which provide a theoretical basis for understanding individual differences in moral decision-making and conduct. Social cognitive theory adopts an interactionist view to moral phenomena, whereby personal and environmental factors operate interactively in determining behavior (Bandura 1991). In addition, it acknowledges that human information processing is highly flexible and can be based on automatic and/or deliberate processes (e.g., Chaiken and Trope 1999; Epstein 1991; Sloman 2002). Self-regulation perspectives provide means of acknowledging that moral conduct is motivated and regulated by self-regulatory mechanisms, which are closely intertwined with cognitive and affective processes.

Although not stated explicitly, Lennick and Kiel's interest seems to be primarily focused on actions, such as whether leaders are able to align their actions with moral beliefs. More specifically, of interest is whether leaders exhibit integrity, responsibility, compassion and forgiveness. Putting moral values into action is certainly one important skill. Yet, research and daily experiences alike suggest that more aspects have to be taken into account. Before acting on what is right, agents have first to recognize that a moral issue is at stake when it arises, and then to decide which course of action may be right (Narvaez 2005; Rest 1986; Reynolds 2008; Treviño and Brown 2004). Given that moral problems are often complex and involve conflicting values, identifying the best moral option is often far from simple (Treviño and Brown 2004). Apparently, individuals vary in their attentiveness to moral matters (Reynolds 2008) and in their reasoning and problem solving capacities. Thus, a MI framework should account for a more complete set of moral abilities.

Several researchers have proposed that individuals are agentic operators (moral agents) in their moral life course (e.g., Bandura 1991). Moral agency is based on multiple abilities, which have an evolutionary basis, but develop with individual and cultural experiences (Chambers 2011; Narvaez 2010b; Nichols 2004; Prinz 2007b; Rest 1986; see also Part II in this volume). A rich and detailed approach of moral expertise development has been provided by Narvaez (2005) that is grounded on Rest's (1986) multi-stage model of moral decision making. Narvaez suggests that moral agents need to develop distinct competences in moral sensitivity (paying attention to moral issues and being responsive to other needs), moral judgment (being skilled at moral reasoning and selecting which actions are most moral), moral motivation (prioritizing moral values and goals over other goals) and moral action (implementing behavior). A critical part of our model, which is clearly related to the framework set forth by Narvaez, is the idea that moral commitment is the central competence. It is governed by an appraisal of moral standards and values and affects all other stages.

In addition, our approach highlights the importance of agents referring to some (pre-established or newly constructed) moral standards, based upon which events or options can be evaluated and behavior regulated (Carver and Scheier 1990; Lennick and Kiel 2005). Such comparison processes between current states ("what is") and desired states ("what should be") are built in psychological mechanisms and involved in each of the proposed competences. Along with Lennick and Kiel, we will call this moral reference system the moral compass. We view the moral compass as an important element of MI—not in the sense that a *specific* set of norms and values is required to be morally intelligent, but in the sense that a moral agent needs to have *some* moral standards available and accessible. Overall, building on previous work and our own perspectives, our MI framework will consist of the following five competences.

1. *Moral Compass:* The reference system containing one's (either existing or newly formulated) moral standards, values or convictions which provide the basis for moral evaluation and regulation.
2. *Moral Commitment:* The willingness and ability to prioritize and strive for moral goals.
3. *Moral Sensitivity:* The ability to recognize and identify a moral issue.
4. *Moral Problem Solving:* The ability to develop and determine a morally satisfactory course of action that resolves conflicting tendencies.
5. *Moral Resoluteness:* The ability to build up moral behaviors by acting consistently and courageously upon moral standards, despite barriers.

Our main goal is to set forth essential moral competences. We refrain, however, from taking a position on which specific moral norms, values, judgments and actions are normatively right or wrong in a defined context. For the following considerations, we define "morality" very broadly as a set of norms, principles, values, and virtues that are governed by an orientation towards the good. As such, they reflect concerns for oneself and for other entities (persons, animals, environment) and are embedded in a justification structure. We are aware that understanding one's moral decision-making and behavior requires an analysis of the agent's lay understanding of morality and on what he or she considers as right or wrong. Yet, we do not mean to suggest that grounding moral intelligence in moral psychology makes normative reflection redundant. On the contrary, moral agents can and do use reflective, deliberate analysis for justifying which moral standards and judgments can reach normative authority (Kennett and Fine 2009). Deliberative reasoning is one element that is involved in constructing the moral compass of an agent, but not the only one.

7.2 Basic Mechanisms of Moral Functioning

7.2.1 Multi-stage Model of Moral Decision Making

Contemporary models of moral decision-making reflect Rest's (1986) multi-stage model, whereby individuals move through a series of four interrelated steps: recognition of the moral issue (moral awareness), making a judgment (moral judgment),

establishing an intention to act (moral motivation), and finally engaging in behavior (moral action). Moral motivation—described as reflecting a person's degree of commitment to take out a moral course of action—has been shown to mediate between moral choice and action (Blasi 1980; Hardy and Carlo 2005) after research has found only disappointing correspondence between moral judgment and behavior.

This view, however, has two essential limitations. First, the extent that moral motivation is equated with setting up an intention to act, which is usually grounded on thoughtful reasoning, this perspective does not address the possibility of morality being based on intuitive judgments and routinized responses (Haidt 2001). As Blasi posited, moral desires can sometimes be so strong that moral actions follow from a "kind of spontaneous necessity" (Blasi 2005: 85). That is, the distinction between both judgment and motivation or motivation and behavior is often blurred. Second, positioning moral motivation only between choice and behavior, neglects the possible impetus of motivation on the other stages. Yet, moral desire may not only serve as a driving force for moral action, but is also likely to affect (consciously or non-consciously) moral perception and moral choice. More precisely, we expect individuals with a strong moral motivation also to be more attentive to moral topics (an aspect which refers to moral perception), to be more likely to engage in reflection and to prioritize moral values when faced with conflicts (aspects which refer to moral decision-making), or to act persistently and courageous (aspects which refer to moral action).

We therefore advocate a model of moral functioning that differs from previous accounts by suggesting that moral motivation is an overarching component (see Fig. 7.1). By moral motivation, we generally mean the desire to bring current state of affairs into line with some valued moral standpoints. This view, with motivation linked to all three other components, acknowledges a) that attempts to meet moral goals do apply to overt behavior, moral perception and judgment alike, and b) that the related processes in each step can be both controlled or automatic. Because motivation works through its use of norms or values, it is also closely tied to the moral reference system (moral compass) which serves to direct our responses.[1]

[1] This construction is somewhat related to the philosophical discussion with respect to moral externalism and internalism (Brink 1997; Simpson 1999)—i.e. the question whether a specific judgment, in order to be called a "moral judgment", motivates the corresponding action *necessarily* or only *contingently*. In our model (Fig. 7.1), motivation mediates between the content (of the moral compass) and the three stages that turn a specified moral stimulus into a moral behavior. This demonstrates a close connection between a moral term and its motivational force, *whenever* the term may play a role in moral behavior. Our model is neutral towards the conceptual question with respect to internalism and externalism in moral philosophy, but it assigns motivation a distinguished role compared to the other components.

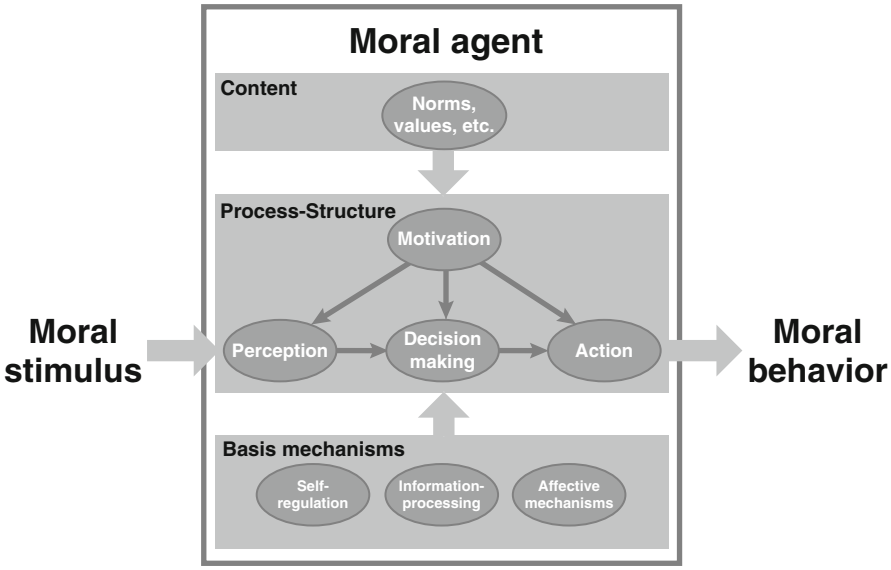

Fig. 7.1 Overview of the of moral functioning and the influence of content and basic psychological mechanisms

7.2.2 Basic Mechanisms

Our theoretical model is grounded in self-regulation and social cognition theory which provides a basis for understanding individual differences in MI (Bandura 1991; Reynolds 2008). We briefly sketch the elementary concepts and mechanisms (see also Fig. 7.1).

Self-regulation: A premise of our framework is that self-regulation is an important feature of moral agency (Bandura 1991; Baumeister 1998; Carver and Scheier 1981). Self-regulation is a highly adaptive process by which people control their attention, thoughts, feelings, impulses and performance so as to live up to social and moral standards in concert with situational factors (Baumeister et al. 2006).

Classic models consider self-regulation usually as a conscious and controlled process, whereby people typically monitor themselves and the environmental circumstances through a feedback loop. They compare and judge their actions in relation to their standards and goals. If they become aware of discrepancies between the current and desired end-states, they can then exert conscious self-control to reduce the discrepancies (Carver and Scheier 1981). In this cybernetic system, emotions do also play a crucial role in that positive affect functions to sustain and negative affect functions to discourage specific goal strivings (Bandura 1991; Carver and Scheier 1990). Another prominent approach emphasizes the role of self-regulation to resist immediate temptations and undesired impulses (such as selfish tendencies) (Baumeister and Exline 1999). Since such forms of conscious self-control require

mental resources, a state of mental fatigue or resource depletion can result in impaired self-control (Mead et al. 2009).

Although reflection and controlled processing play an important role in self-regulation, researchers have also started to emphasize that regulation also critically depends on non-conscious, automatic processes. It is argued that characteristics of the social environment can directly activate schemas and goals which in turn exert non-conscious effects on self-regulation (e.g., Fitzsimons and Bargh 2004). Repeated practice and goal pursuits are also likely to promote automatic self-regulation, while decreasing involvement of controlled processes. Fitzsimons and Bargh (2004: 152) propose that "due to the apparently quite limited capacity of conscious self-regulatory abilities…much of self-regulation has to occur nonconsciously to be successful". Our framework advocated in this chapter sympathizes with this view that moral self-regulation operations are governed both by automatic and controlled processes. While conscious moral self-regulation occurs through willful application of moral standards to moral processing, automatic regulation occurs as a result of learned orientations and responses (see also Sekerka and Bagozzi 2007).

Information Processing: This conception of self-regulation is closely related to dual process or dual system models that have been advanced in cognitive and social psychology to account for the fact that human information processing is highly flexible (for reviews see: Lapsley and Hill 2008; Smith and DeCoster 2000). Virtually all models assume two systems which work interactively (e.g., Chaiken 1980; Epstein 1991; Petty and Cacioppo 1986). The operations of System 1 are usually described as automatic, intuitive, implicit, fast, effortless, often emotionally charged, evolving from associative learning, and working on a preconscious level (Bargh 1997). This system has been referred to as performing pattern-matching and pattern-completion functions (Smith and DeCoster 2000; Reynolds 2006). The operations of System 2, in comparison, are usually described as deliberate, controlled, explicit, slow, effortful, based on propositional thinking, and conscious. It enables individuals to monitor the quality of mental operations and overt conduct and to engage in reflection, reasoning and conscious self-control.

Though most dual-process models assume that both systems interact, there is a rich literature indicating that the prevalence of automatic or controlled processes is affected by situational and personal factors (Chaiken 1980; Fazio 1990). For instance, research has shown that expenditure of cognitive effort is more likely under conditions of high personal accountability (i.e., conditions were people need to justify one's decisions and actions to others; Lerner and Tetlock 1999), or among people who enjoy to engage in effortful analytic activity (high in need for cognition; Cacioppo et al. 1996). Opposingly, in conditions of low accountability, lack of motivation for extended reflection or lack of situational opportunities (such as time pressure, high mental workload) individuals are more likely to foster spontaneous processing (Fazio 1990). Obviously, it is of paramount importance to take into account this variability in processing when examining moral functioning in professional settings and daily life to better understand and support moral functioning.

One important (personal) factor that is proposed to facilitate automatic processing has to do with the accessibility of moral concepts. As Kahneman (2003) asserted, a core feature of intuition is that moral concepts pop up very easily and effortlessly. In order to explain why some ideas come to mind more easily, while others demand work, some authors have adopted the term "accessibility" from memory and social cognition research (e.g., Higgins 1996). It is proposed that mental representations vary in their activation potential, i.e., in terms of how easily they can be activated. Once activated, they guide information processing and allow the individual to interpret situations through the lens of the activated elements. Particular mental representations, such as strong attitudes, deeply held values or principles, beliefs or traits which are central to one's identity or culture, are said to be "chronically accessible" in that they become habitually activated (Higgins 1996). Consistent with other researchers in the moral domain, we conceive moral standards and values as moral schemas that vary in their accessibility (Jordan 2009; Lapsley and Narvaez 2004; Narvaez et al. 2006). Hence, chronic accessible moral schemas are considered to foster automatic moral self-regulation.

Affective Mechanisms: Automatic and deliberate processes go along with emotions which also affect self-regulation. Moral theory and research has traditionally focused on the conscious and deliberate aspects of moral judgment (e.g., Kohlberg 1969). Meanwhile, many authors assert that emotions are important cues that provide information and motivational resources for judgment and decision making (e.g., Loewenstein and Lerner 2003), moral regulation (e.g. Bandura 1991) and the development of moral functioning (Narvaez 2010b). For example, it has been argued that our emotions reflect an inherent "moral sense" (see the contribution of Prinz (Chap. 6) in this volume), or that moral judgments are sometimes influenced, if not dominated, by "gut feelings", which tell us that something is right or wrong (Monin et al. 2007; Wheatley and Haidt 2005). That is, emotion or affect is seen to play role in the intuition process itself, resulting in affect-laden judgments (Epstein 1991; Haidt 2001).

Generally, emotions are expected to disrupt cognitive control and deliberative processes when their arousal level is high (Janis and Mann 1977; Luce et al. 1997). However, when emotions are on a moderate level, they serve informational and motivational functions. In terms of informational functions, emotions are considered to impact cognitive processing as they signal where to focus attention (Forgas 1995) or help to evaluate and select options, as they provide vital information about aspects of the current situation or about past experiences with similar situations (Damasio 1994; Schwarz and Clore 1983; Slovic et al. 2002).

As to the motivational functions, some approaches consider affective self-reactions in the form of anticipatory self-satisfaction (e.g., pride) or self-sanctions (e.g., guilt) to provide the mechanisms by which standards motivate and regulate moral conduct (Bandura 1991; Carver and Scheier 1990), and by which people's commitment to moral values are reinforced (Tangney et al. 2007). According to Hoffman (2000), emotions transform abstract moral principles and "cool" reasons into hot cognitions, thereby energizing moral goals. Frank (1988) has argued that moral emotions (such as guilt, shame) work as "commitment devices" that help individuals to overcome immediate rewards in order to pursue long-term strategies.

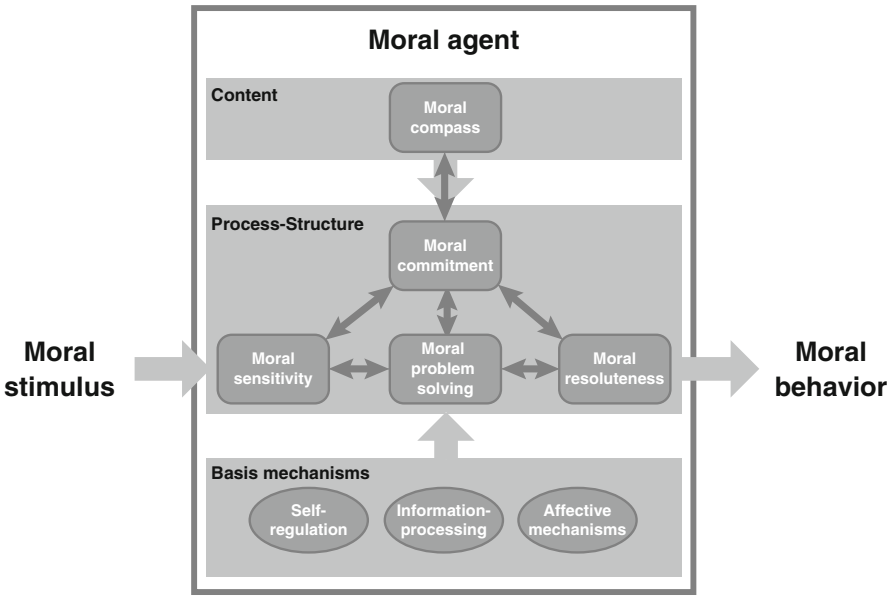

Fig. 7.2 The five building blocks of moral intelligence in relation to the multi-stage model of moral functioning

7.3 The Framework of Moral Intelligence

We now turn to the prerequisites of MI. As depicted in Fig. 7.2, we suggest that effective moral regulation depends on having a moral compass and a set of four specific moral abilities.

7.3.1 The Moral Compass

The first prerequisite of MI is to have—as Lennick and Kiel (2005) posited—a "navigation tool" or a "moral compass" for one's life. The moral compass refers to some pre-established or newly formulated moral standards and norms, which direct the agent's reactions. It serves as a reference, based upon which events, options and conduct are cognitively and affectively evaluated and regulated (Carver and Scheier 1981; Baumeister and Exline 1999), and it sets the occasion for affective self-reactive influences (Bandura 1991).

The content of the moral compass is multifaceted. Moral values, moral convictions, ethical principles, religious beliefs, personal goals, self-related beliefs as well as behavioral scripts, etc., form such ingredients. In the following, we will exemplify how elements (or a structured set of elements) of the moral compass may interact with the abilities that constitute MI.

Formally, we conceive the single elements of the moral compass as moral schemas (Jordan 2009; Lapsely and Narvaez 2004; Narvaez 2005). Since such schemas are acquired by practice and shaped by iterative learning and social and cultural factors, the content and structure of the moral compass will vary across agents (Aquino and Reed 2002; Lapsely and Narvaez 2004). For moral schemas to become operative as standards of comparison in moral regulation, it is inevitable that they are accessible. As mentioned earlier, frequently or recently activated mental structures (e.g., through repeated practice or priming) are more accessible. Therefore, deeply held values and beliefs that are chronically accessible (Higgins 1996) are very likely to affect subsequent processes.

Our model suggests that the moral compass of individuals high in MI differs in at least two respects from individuals low in MI. First, they have more complex moral schemas. As Narvaez and other scholars posited, moral experts are similar to experts in other fields (but see the contribution of Musschenga (Chap. 11) in this volume). They differ from novices in that they have more complex, domain relevant and chronically accessible mental structures, which trigger effective responses (Dane and Pratt 2007; Narvaez 2005; Lapsely and Narvaez 2005). Second, individuals are likely to differ in how important moral values are for them, which is represented in the structure of the elements that form the moral compass (e.g., in the sense that they are more coherent; Thagard 2000). High-MI agents are likely to have strong internalized moral standards that penetrate their self-understanding. Since strong and central moral values represent highly accessible structures, high-MI individuals are more likely to make use of moral schemas in guiding responses.

Despite the relevance of the moral compass as a navigation tool, standards alone do not instigate action. In the following, moral commitment is proposed to represent the focal competence that invokes and enhances moral regulation.

7.3.2 Basic Moral Competences

Moral Commitment: Moral failures are often not the result of lack of knowledge about what should be done, but the result of a weak motivation to strive for moral goals (Monin et al. 2007). This observation justifies, from a psychological point-of-view, the specification of motivation as a distinct but overarching component in the multi-stage model of moral decision-making, with implications for all other components (see Sect. 7.2.1). Empirical data and daily observations also suggest that agents strongly vary in their adherence to moral goals and their desire to comply with moral standards. In our framework, moral commitment accounts for this variability. Narvaez has asserted that experts in moral motivation are capable of cultivating moral identity and moral regulation that lead them to prioritize moral goals and foster habituated moral concerns (Lapsley and Narvaez 2005; Narvaez 2005). Similarly, we posit that moral commitment consists of an implicit or explicit committing of oneself to moral goals that instigates an enduring striving for moral ends. We define moral commitment as the ability to selectively focus on moral goals

and strive for desirable ends. This is not always a simple task. As individuals proceed from perception to action, effective moral regulation often requires one to keep track of internal and external cues, reflect upon process and outcomes, and alter one's operations (e.g., Bandura 1991; Baumeister and Exline 1999). Hence, moral commitment requires self-monitoring, self-reflective and self-influencing capabilities.

Moral commitment has a privileged position in our framework (see Fig. 7.2). Being linked to the moral compass that helps to define which goals and ends may be desirable, moral commitment carries with it the urge to comply with those goals, which affects moral perception, choice and action. Since morally committed agents have a heightened sense of obligation and responsibility, they make morality part of their life and self-understanding, which in turn further contributes to the evolvement of (chronically accessible) moral schemas (Narvaez 2005; Schlenker 2008).

Strong expressions of moral commitment are, e.g., "protected values" and "moral identity". Protected values refer to non-instrumental values that involve strong moral convictions about the impermissibility of trading specific values in exchange for other good, in particular monetary benefits. For example, if people consider human life, nature, honor or honesty as protected values, empirical evidence indicates that those people are reluctant to sacrifice or to trade off such values (Atran et al. 2007; Baron and Spranca 1997; Skitka et al. 2005; Tanner 2008; Tanner et al. 2009; Tetlock et al. 2000). Moral identity, on the other hand, reflects the degree to which a set of moral beliefs and values are central to one's self-understanding (Aquino and Reed 2002; Blasi 1983; Colby and Damon 1992; Hardy and Carlo 2005).

Morally committed individuals are endowed with willpower (strength of self-control). In moral research, the study of willpower has only recently become more prominent. One influential model (Baumeister 1998) has advanced the idea that reality is filled with passion and selfish temptations (e.g., striving for short-term benefits instead of long-term collective benefits) which render moral behavior less likely. For example, a characteristic of many business situations is that people are provided with the opportunity to profit from dishonest acts (e.g., by deceiving or cheating on others). Such opportunities may present a conflict between taking selfish gains vs. acting in virtuous ways. Hence, one's moral strength relies on the ability to resist selfish temptations by exerting conscious self-control. In general, there is much evidence of individual differences in such self-control skills (Baumeister et al. 2006). Baumeister and colleagues also posited that self-control is a resource that fluctuates and can, like a muscle, be depleted (Baumeister and Exline 1999). Following this metaphor, it was hypothesized that people would be more likely to behave dishonestly when their self-control resources were depleted. Indeed, empirical studies have demonstrated that people were more likely to cheat under conditions of mental fatigue (Mead et al. 2009).

This approach of selfish temptations typically focuses on controlled exertion of willpower that is required to resist such temptations enabling a delay of reward. Such a view, however, tends to neglect the possibility of automatic regulatory processes and the possibility that not all individuals are tempted when faced with the

opportunity to profit from unethical behavior. Specifically, from morally highly committed individuals we would expect that their moral self-regulation is more automatized. Since they have strong internalized moral values, they are less tempted by opportunities for unethical gains. They therefore do not have to rely on conscious and active self-control. Consistent with this assumption, very recent studies referring to (dis)honest behaviors have revealed that people who routinely behave honestly, who endorse honesty as a protected value, or who consider morality central for their self-concept are less tempted and less likely to engage in controlled activities (Gino et al. 2011; Greene and Paxton 2009; Gibson et al. 2013). We generally believe that highly committed individuals, as long as the tasks are not demanding, will accomplish much of their moral self-regulation by automatic processes, since they can rely on highly accessible moral schemas and habits that maintain moral conduct. However, due to their heightened commitment to moral goals, they should also be more willing to mobilize willpower when faced with highly demanding tasks.

In sum, moral commitment is pivotal for the strength of moral regulation. We expect high-MI individuals to have a strong and enduring desire to strive for moral ends that leads them to engage in automatic or controlled self-regulatory processes (depending on task demands). Due to their strong moral motivation, they are more likely to monitor internal and external states in terms of how they meet moral standards, to reflect on the process and outcome, and to sanction their misconduct (by feeling shame or guilt).

Moral Sensitivity: Moral sensitivity refers to the key issue that individuals must first recognize that they may be facing a moral problem. If no moral issue is perceived, no moral judgment or decision-making process occurs (Clarkeburn 2002; Rest 1986; Sparks and Hunt 1998). Yet, moral aspects are rarely immediately obvious in daily life. Individuals are confronted with situations of great variety and complexity, making it necessary for people to attend to some stimuli while ignoring others (Fiske and Taylor 1991). While some individuals are endowed with an intuitive sense of concern for others, fairness or apprehension of what is right or wrong and rapidly detect that a moral standard, norm or code may be violated in a situation, others are "morally blind" (Pedersen 2009). Therefore, moral sensitivity refers to the ability to recognize and conceive of moral features when they arise in practice. This includes envisaging whether a given set of actions can harm or help other parties or, more generally, violate internalized moral standards or codes that govern professional conduct. It also entails the capacity to understand a situation from a number of different perspectives. As such, moral sensitivity involves empathy and perspective-taking skills (Narvaez 2005, 2010a).

A dual process conceptualization of moral sensitivity suggests that it includes automatic and controlled processes. As an inherently perceptional process, it involves non-conscious matching of patterns according to which individuals automatically compare their observations with their standards (e.g., Reynolds 2006). The outcome of such a comparison may be rapidly arising intuitions that the perceived situation or the behavior of another person is "wrong" in the moral sense

(e.g., other people might be harmed, human rights might be violated). Such reactions can be associated with more or less strong emotions which serve as additional signals that moral issues are at stake. For example, outrage or anger illuminates moral infractions of others, guilt or shame accompanies one's own wrongdoing. An individual, however, may also intentionally search and reflect on the potentially moral implications of an event.

Researchers have only recently begun to study the phenomenon of moral sensitivity (also referred to as moral awareness or ethical sensitivity). Jones (1991) pointed out that specific characteristics of the issue under consideration (such as the magnitude of the consequences, the immediacy or proximity of the moral issue) can attract attention and therefore affect moral sensitivity. Other research suggests that individuals largely differ in their ability to identify the moral implications of a given situation. In a recent study, Jordan (2009) compared business managers with academics. She argued that business managers have business schemas rather than moral schemas more dominant, because they have more experience with strategy- and industry-related problems (such as maintaining financial profitability) than with moral-related problems (such as protecting the interests of stakeholders, employees). Because schemas guide information processing and attention, it was expected that individuals with a dominant (i.e., chronic accessible) moral scheme would direct more attention to moral issues than an individual with other dominant schemas. In line with this, Jordan found that business managers were (compared to academics) less likely to detect moral-related issues than business-related issues in morally ambiguous vignettes.

There is also evidence that people holding or not holding protected values are attentive to different aspects. Some authors have claimed that people endorsing protected values are often prone to deontological thinking as opposed to consequentialism (Baron and Spranca 1997; Tanner and Medin 2004). That is, the focus is more on the inherent rightness and wrongness of actions themselves rather than on the magnitude of the consequences associated with the actions. One implication of a deontological focus is that it should make a difference whether outcomes derive from an act or an omission, whereas from a consequentialist perspective, this difference should be irrelevant. Consistent with this, Tanner and colleagues (Tanner 2009; Tanner and Medin 2004; Tanner et al. 2008) found that people endorsing protected values and a deontological orientation paid more attention to the distinction between acts and omissions, while for individuals not endorsing protected values and with a predominantly consequentialist focus it did not matter whether the consequences were an outcome of an act or omission.

Overall, these investigations demonstrate individual differences in the capability to identify moral issues, thereby individuals with dominant moral schemas (which should be especially the case for agents with strong moral commitments) show higher levels of moral alertness. From agents high in MI, we expect that they are more likely to detect moral aspects and that they are also quick and accurate in "reading" a moral situation. This follows from the idea that these individuals have highly accessible moral schemas, which in turn support automatic and fast detection of moral components. Yet, complex situations sometimes require deliberate processing. Because moral problems become only apparent to those who are interested

in them, we would furthermore expect that individuals high in MI be generally more motivated to detect the moral implications of an event which can involve both automatic and deliberate processes. Finally, as noted earlier, the capabilities of taking another's perspective as well as empathy are other elements that are seen to help in envisaging potential harm to other parties and thus support moral sensitivity (Narvaez 2005, 2010a). Agents high in MI should be more endowed with such skills than individuals low in MI.

Moral Problem Solving: Once a moral problem and the involved key parties have been identified, the next challenge consists of finding viable ways of coping with it. Moral decision-making is about finding out "what ought to be done", while dealing with competing pressures and generating and evaluating different options with moral and other (e.g., economic) consequences. Such problems can be emotionally distressing as they put fundamental issues at risk and involve trade-offs between conflicting values with unwanted or threatening consequences (e.g., other people may be harmed) (Hanselmann and Tanner 2008; Luce et al. 1997). Furthermore, moral problems are often complex and ill-defined, leaving the decision-maker uncertain about the range of alternatives and their short- and long-term consequences. Since such situations do hardly offer obvious solutions about which course of action is most ethical, a substantial part of the problem solving process consists of constructing options which are then evaluated.

Because many decisions are complex and ill-defined and individuals face limitations in cognitive capacity and time, researchers generally assert that decision-makers are not fully rational. Instead of considering all alternatives and consequences to identify the objectively "best possible" course of action, decision-makers cope with limited information by searching for options that are "good enough". That is, people can rarely "maximize", they have to "satisfice" (Gigerenzer 2010; Simon 1955). Consistent with this, our conception of MI posits that the goal is to create the solution that at best meets moral standards, while reconciling conflicting value systems. Such a search for morally viable solutions requires taking the various objections and divergent values into account without losing the moral direction. We define moral problem solving as the ability to generate morally satisfactory and reconciliary solutions.

Yet, the decisions that a person makes are shaped by their moral standards, external demands and his or her conscience. This process entails specific steps, such as (1) value clarification, (2) generating and evaluating different courses of actions, and (3) resolution. As noted earlier, decision-making does not only involve explicit reasoning but also current or anticipated emotions (e.g., regret, guilt, shame) that are used as inputs in the decision process (Haidt 2001; Loewenstein and Lerner 2003; Schwarz and Clore 1983). In addition, since agents may be required to generate or construct new options, effective moral problem solving also entails the capacity of creative imagination skills. We agree with Keeney (1992) that clarifying the values at the beginning rather than in subsequent steps of the choice process can help to promote creativity. He demonstrated that focusing early and deeply on the values, encourages people to search for new alternatives, which, in turn, may lead to more desirable outcomes. In a similar vein, we propose that individuals with an implicit

or explicit focus on moral values (moral commitment) search more extensively for integrative and morally acceptable solutions.

As with the other competences, moral problem solving can be deliberately or automatically directed. As noted above, while traditional paradigms have largely emphasized cognitive reasoning models (e.g., Kohlberg 1969; Piaget 1932/1965), recent research has started to recognize the importance of automatic processes which reflect implicit associations between concepts and normative and affective valences (Haidt 2001; Reynolds 2006). Whether decision-making involves more spontaneous or more deliberate processes is a function of various conditions. For example, empirical evidence suggests that intuitive processes take precedence over deliberate thinking under situational conditions of high time pressure, high mental workload or high levels of uncertainty (Fazio 1990).

No final conclusion can be made about the question of whether people should better rely on intuitions or deliberations when making decisions. However, a growing body of research suggests that intuitive processing is sometimes superior to analytical processing. Gigerenzer and colleagues, for example, have shown that with regard to measurable criteria (e.g., decision accuracy, success at stock market) simple, fast and frugal heuristics can perform as well or even better than rational decision-making models (Gigerenzer et al. 1999). Recent studies have demonstrated that under certain circumstances, choices based on unconscious processes outperform conscious decision-making (Dijksterhuis et al. 2006). Others suggest that intuitive judgments result in more accurate decisions when they are based on knowledge that reflects prior experiences with the same or similar tasks and match the demands of the given decision task (Plessner and Czenna 2008).

Of course, to the extent that the task forces automatic responses (e.g., under high time pressure conditions), the more important it is that people have—as some researchers pointed out—expert-like, well educated intuitions (Lapsley and Narvaez 2005; Narvaez 2005, 2010a). Experts differ from novices in that they have organized knowledge and highly automatic and effortless skills. Klein and colleagues have extensively studied professional's strategies (e.g., firefighters, military leaders, jurors or airline pilots), when they had to make rapid but tough decisions under difficult conditions (for a review see Klein 2008). They found that experienced decision-makers use their background knowledge to rapidly categorize the situation and to retrieve the most typical course of action. They evaluate this course of action by using mental simulations in order to analyze whether it will work or not. Interestingly, Klein and associates also found support for the hypothesis that the first option considered by experienced decision-maker is usually the most satisfactory one (Klein et al. 1995). In a similar vein, we argue that moral problem solving under conditions that promote automatic responses requires decision-makers to have proper and highly accessible moral structures which entails associations between moral standards, declarative knowledge and a repertoire of procedural patterns.

Clearly, individuals will vary in the extent to which they have proper moral schemas available and chronically accessible. From individuals high in MI, we generally expect that they will behave "more expert-like" and will quickly come up with integrative and morally satisfactory solutions (under conditions that trigger automatic

processes) or to engage in creative imagination to develop new options (in particular under conditions that better allow reflection and reasoning).

Moral resoluteness: Once a decision has been made, the next step is to implement moral goals into visible action. In their moral intelligence approach that they applied to the business context, Lennick and Kiel (2005) suggested that morally competent leaders are those who exhibit integrity (acting consistently with values, keeping promises), responsibility (e.g., being accountable for personal choices), compassion and forgiveness (e.g., caring about others; letting go of one's own and others mistakes). These may be examples of desirable and virtuous behaviors, yet, the straight path to virtue can be often very difficult. Acting upon moral standards that are considered as right can be hard, e.g., when individuals are faced with threats and dangers that are associated with moral behavior or when social norms are not congruent with moral actions. Other situational barriers, such as lack of money or time, risks of own career survival, social pressures, or the prevalence of unethical norms and practices in an organization (i.e., lack of an ethical work climate and culture; Treviño et al. 1998; Victor and Cullen 1988) are factors that may inhibit people from acting morally.

We propose moral resoluteness to be the competence that enables people to overcome external obstacles, to face dangers and threats to self, and to have stamina when pursuing moral actions. Moral resoluteness refers to the ability to act consistently and persistently upon moral standards, despite pressures. Agents with moral resoluteness are expected to stand up for their deeply held moral principles even in the face of adversity. They convey moral standards reliably through visible actions and consistently across time and situations. Moral resoluteness therefore entails resistance, courage, consistency and perseverance (Blasi 2005; Sekerka and Bagozzi 2007; Tanner et al. 2010).

As with the other components, moral resoluteness can be subject to automatic and controlled self-regulation. Moral behavior can be based on routinized, well-learned responses triggered in the situation. Because behavior in such situations is performed on an automatic level and under less self-control, proper moral reactions heavily depend on having the "right" mental schemas entailing strong associations between moral standards and procedural patterns. However, many other, more demanding and difficult situations require a more controlled, willful application of one's standard of behavior. Furthermore, moral resoluteness may be facilitated by emotional responses. Positive emotions in anticipation of performing moral acts, or negative emotions in anticipation of moral failures are likely to support moral courage and continued moral engagement (Carver and Scheier 1990; Sekerka and Bagozzi 2007; Tangney et al. 2007).

Individuals clearly differ in their courage and persistence to act upon their moral standards, even when it is costly. Milgram's famous experiments (1963) on the role of obedience to authority are examples of how easy it is to make people harm others. These studies demonstrated that many people were willing to give electric shocks to another person, simply because a scientific authority commanded them to do so. Nevertheless, there were also a few people who resisted the authority. Again, we deem moral commitment as an additional factor that functions to promote moral

resoluteness. Strong moral commitments have been found to be negatively correlated with corruption in business (Fine 2010), negatively with (self-reported) antisocial behaviors (such as lying, stealing or cheating), but positively with (self-reported) prosocial behaviors (such as helping or volunteering) (Schlenker 2008).

Strong moral convictions and moral identity which reflect strong moral commitments seem to also serve to promote application of moral standards into behavior. Gibson, Tanner and Wagner, for instance, have examined the role of individual's conviction that honesty is a value that is "not for sale" and therefore ought to be "protected" from trade-offs against monetary benefits. In experiments simulating realistic business settings, people were provided with the opportunity to gain (real) money by misleading others. The results confirmed that people with higher levels of protected values for honesty were more likely to sacrifice money to maintain honesty, some of them even displayed absolute resistance to trade off honesty for money (Gibson et al. 2013). Other empirical studies suggest that people with high levels of moral identity are more likely to engage in moral behavior (Hardy and Carlo 2005). Having a strong moral identity was also found to weaken the effect of moral disengagement (Aquino et al. 2007)—a common dissonance-reducing rationalization mechanism that allows people to shield themselves from moral self-condemnations when acting immorally (Bandura et al. 2001).

Consequently, we expect individuals high in MI to demonstrate moral resoluteness. This manifests in more consistency between words and deeds and more persistence and courage to overcome barriers. High-MI individuals are expected to be more efficacious and engaged in automatic or controlled self-regulation to act upon moral standards and principles.

7.4 Enhancing Moral Intelligence

We close this chapter with brief remarks about the practical value of the advocated model and the question of how to enhance MI. Our goal was to suggest a model that depicts the main elements and features of moral intelligence. In doing so, we considered having a moral compass (beliefs about what is the right thing do to) and a set of four main competences (moral commitment, moral sensibility, moral problem solving and moral resoluteness) as requirements of moral intelligence. Individuals are likely to vary with respect to each of those skills. Some may have excellent moral sensitivity, but may be poor in moral resoluteness. Some may have low moral commitment, but in the few situations where they indeed care about morality, they perform well in all three remaining competences. One of our next goals is to build upon this model to develop valid measurements of those competences that help to detect one's own moral strengths and weaknesses. In this vein, we hope that our framework can serve as an important platform for researchers and practitioners alike, for future research, intervention and education.

In line with a long tradition within moral philosophy and moral psychology, moral competences are acquired and enhanced by moral practice. Learning is

conceptualized as a reiterative cycle between moral competence and moral action (Narvaez 2010a; Pederson 2009). We note that moral performance is not just influenced by personal factors, but also by cultural and contextual factors (such as ethical climate, law or political structures, incentives, etc.). Early experiences establish trajectories for intuitions and reasoning but are then shaped by culture, education and experience (Narvaez 2010a). Therefore, promoting moral intelligence implies the creation of structures that allow people to cultivate and practice skills (Hogarth 2001). Narvaez and colleagues (e.g., Narvaez et al. 2006) describe the development of a moral personality as a construction of moral schemas. Applying a novice-to-expert approach, the education toward moral expertise is seen as a process of evolving moral schemas that is based on extensive practice and learning. Practice and repeated experience foster the development of percepts and concepts that become chronically accessible (Narvaez et al. 2006). In this vein, we see MI developing with practice and experience that continuously shapes the agent's mental structures. As with other capacities, moral competence is developed through explicit and implicit learning (i.e., conscious and non-conscious forms of knowledge acquisition), vicarious learning (i.e., learning by observing the behavior of others and its consequences), reflection and conscious self-regulation by which individuals instigate behavioral changes (Bandura 1965; Dane and Pratt 2007; Hogarth 2001; Pedersen 2009).

In conclusion, this contribution was designed to highlight main features and processes involved in moral functioning, and to discover the main abilities of Moral Intelligence. Certainly, more work is needed to further develop the framework, to define what the normatively proper standards, values and reactions should be, to develop useful and valid assessments of Moral Intelligence, or to ascertain how environment, education and training should be designed to facilitate and cultivate the development of Moral Intelligence.

Chapter 8
Moral Brains – Possibilities and Limits of the Neuroscience of Ethics

Kristin Prehn and Hauke R. Heekeren

8.1 Introduction

A "neuroscience of ethics" focuses on the question of what can be learned about morality or moral standards (standards that an individual or a group has about what is right and wrong or good and evil) through a growing understanding of how the human brain works.

Empirical research in the field of morality comprises two main questions: How people generally distinguish "right" from "wrong," and how people behave in a morally appropriate way, for instance, resisting the temptation to do wrong. Both questions, as well as the question regarding an interrelation of moral judgment and behavior, have been of recurring interest in many disciplines including philosophy, arts, religion, or law studies. In the field of psychology (the science aiming to understand and predict human behavior), in particular, a variety of theories and models have been developed to explain moral judgment. Interestingly, in most psychological approaches moral judgment is regarded as a precondition for moral behavior and defined as the evaluation of one's own or someone else's behavior with respect to social norms and values considered to be virtuous by a culture or subculture, such as not stealing or being an honest citizen (definition adapted from Haidt 2001: 817).

In recent years, advances in cognitive neuroscience have provided new technologies, such as functional magnetic resonance imaging (fMRI), that make it possible to investigate the neural substrates of moral judgment and behavior (to "localize the moral brain"). Since the advent of these methods, the question of how and where morality is located in the human brain has triggered much research. Research studies question which cognitive processes are involved, to what extent these processes are open to conscious deliberation, and whether human moral behavior is a product of education or a result of an innate mechanism activated during childhood.

K. Prehn (✉) • H.R. Heekeren
Cluster of Excellence "Languages of Emotion", Freie Universität Berlin, Berlin, Germany
e-mail: kristin.prehn@fu-berlin.de

M. Christen et al. (eds.), *Empirically Informed Ethics: Morality between Facts and Norms*, Library of Ethics and Applied Philosophy 32, DOI 10.1007/978-3-319-01369-5_8,

In particular, the question of whether moral judgments are caused by emotional or cognitive processes and whether emotional responses make moral judgments better or worse has caused much controversy and debate.

In the following chapter, we will, first, give a brief overview of traditional and recent psychological models of moral judgment and behavior (Sect. 8.2). Second, we will introduce the neuroscientific approach and the methods applied to the study of the "moral brain," including the examination of brain damaged patients, neuroimaging, and neurostimulation (Sect. 8.3). Then, we will present main lines of research and give a critical overview of some studies aiming to disentangle domain-specific and -general processes involved in moral judgment and behavior and, finally, present our own empirical findings based on a neuroimaging study investigating the influence of individual differences in moral judgment competence (according to the Dual Aspect Theory by Georg Lind; Sect. 8.4).

8.2 Psychological Models on Moral Judgment and Behavior

8.2.1 Moral Reasoning Investigated from a Cognitive-Developmental Perspective

Psychological research on morality has long been dominated by a cognitive-developmental approach, investigating the maturation of moral reasoning and its underlying moral orientations and principles as a precondition for moral behavior (Piaget 1965; Kohlberg 1969).

To investigate the maturation of moral reasoning, Lawrence Kohlberg presented children and adolescent participants with moral dilemmas and asked them to argue why it could be justified to choose a certain action. In one of his best known dilemmas ("Heinz dilemma"), for instance, a man named Heinz has to decide if he should break into a drugstore to steal a medicine that would save the life of his dying wife. Based on how children and adolescents argued, Kohlberg established his much cited six-stage model (three levels including two stages at each level) of cognitive development of moral reasoning. This model proposes that humans progress through six stages as their cognitive abilities mature. During this development, people acquire a more sophisticated understanding of social relationships and, in particular, come to see situations not only from their own perspective but also from the perspectives of all other people involved in the conflict.

According to Kohlberg, at the pre-conventional level, young children think a behavior is right when an authority says it is. Doing the right thing means obeying an authority and avoiding punishment (stage 1 = obedience and punishment orientation). At stage 2 (= self-interest and exchange orientation), children see that there can be different sides to an issue and each person is free to pursue his or her own interests. Additionally, children understand that it is often useful to do someone else a favor. Later, at the conventional level, young people think of themselves

as members of their society with its values, norms, and expectations. At stage 3 (= interpersonal accord and conformity orientation), they aim to be a "good boy or girl," which basically means being helpful to other people who are close to them. At stage 4 (= authority and social order maintaining orientation), the concern shifts toward obeying the laws to maintain society as a whole. At the post-conventional level, people start to think about the principles and values that constitute a good society. At stage 5 (= social contract orientation), laws are regarded as social contracts rather than rigid dictums. Those laws that do not promote the general welfare should be changed when necessary to meet the greatest good for the largest number of people (e.g., by democratic majority decisions). Finally at stage 6 (= universal ethical principles), moral reasoning is thought to be based on abstract reasoning using universal ethical principles of justice and of the reciprocity and equality of human rights with respect for the dignity of human beings as individuals (Kohlberg 1969).

For our purposes, in order to explore which processes are involved in moral judgment and behavior, the relevance of the cognitive-developmental theory could be seen as the idea that morality does not only rely on the acquisition of social norms and values held to be virtuous in a community (i.e., the acquisition of social knowledge), but also on the way individuals understand and think about social situations. Following Kohlberg, how people think about social situations qualitatively changes as result of an active interaction of the individual with his or her social environment.

It is also noteworthy that Kohlberg defined morality from his developmental perspective in terms of an ability, as "the capacity to make decisions and judgments which are moral" (i.e., based on internal moral principles and to act in accordance with such judgments; Kohlberg 1964: 425). Based on this notion of morality as an ability, Georg Lind in a current theoretical approach (Dual Aspect Theory) defines morality as consisting of two inseparable, yet distinguishable aspects: (a) a person's moral orientations and principles and (b) a person's competence to act accordingly. Following the Dual Aspect Theory, moral judgment competence is the ability to apply moral orientations and principles in a consistent and differentiated manner in varying social situations. Thus, social norms and values (represented in the Dual Aspect Theory as affect-laden moral orientations and principles) are linked with everyday behavior and decision making by means of "moral judgment competence." Moral judgment competence represents a cognitive component, regarded as an important condition for living together in a democracy. Moral judgment competence can be trained by interventions such as the Konstanz Method of Dilemma Discussion (KMDD), developed by Lind to improve pro-social behavior, learning and decision-making skills, affect regulation, and the prevention of antisocial behavior (Lind 2008).

Although the cognitive-developmental theory has strongly influenced the discourse about morality and the subsequent research on moral education, there is also some criticism. It has been criticized, for instance, that Kohlberg investigated only post-hoc justifications for moral judgments that already had occurred, rather than actual reasoning processes leading to moral judgments (see Sect. 8.2.2; Haidt 2001).

Moreover, the assumption of a universal and invariant sequence of developmental stages has been doubted (Snarey 1985). Another point of criticism is that Kohlberg's theory emphasizes justice to the exclusion of other values. Carol Gilligan, in particular, argues that Kohlberg's theory is mainly based on empirical research in male participants and thus does not adequately describe the concerns of women. Therefore, she developed an alternative theory of moral reasoning that is not based on justice but on the ethics of caring (Gilligan 1977; Gilligan and Attanucci 1988; for recent neuroscientific studies investigating differences between justice and care ethics, see Robertson et al. 2007; Cáceda et al. 2011).

8.2.2 The Role of Emotion and Intuition in Moral Judgment and Behavior

More recent theories and models question the importance of rational reasoning processes for morality and emphasize the impact of intuitive feelings and automatic emotional responses.

James Blair (1995), for instance, suggested that humans (similar to other animals) possess a mechanism which, when activated by the communication of distress, such as sad facial expressions or tears, mediates the suppression of aggression (a so-called violence inhibition mechanism, VIM). He claimed that the VIM is a precondition for the development of (1) moral emotions such as sympathy, guilt, and remorse, (2) non-violent behavior, and (3) the moral/conventional distinction during childhood. This latter distinction between moral and conventional transgressions found in the judgments of children and adults marks the ability to differentiate cases where harm is caused to a person (= moral transgressions) from cases where only socio-conventional norms are violated (= conventional transgressions) without necessarily causing harm (e.g., spitting in a glass of wine at a dinner party; see also Turiel 1983; Nichols 2002). Specifically, Blair proposes that a lack of the VIM would explain the core symptoms of psychopathy and his empirical study could demonstrate that psychopaths—which according to the diagnostic criteria show an early onset of extremely aggressive and violent behavior and a lack of moral emotions like sympathy, guilt, and remorse—also fail to differentiate between moral and conventional transgressions in contrast to healthy controls (Blair 1995).

The social intuitionist model by Jonathan Haidt (2001) is another theory suggesting that fast and automatic intuitions like gut feelings or aesthetic judgments are the primary source of moral judgments, whereas rational arguments as obtained in Kohlberg's interviews are only used to construct post hoc justifications for judgments that have already occurred. "Moral intuition" is defined as the sudden appearance of a moral judgment in consciousness including a strong affective valence (good vs. bad, like vs. dislike). This would mean that reasoning is less relevant to moral judgment and behavior than Kohlberg's theory suggests and implies that people often make judgments without weighing concerns such as fairness, law, human rights, or abstract ethical values. Haidt illustrates the alleged minor role of rational

reasoning in moral judgment provocatively as the "rational tail of the emotional dog" and provides some striking examples of "moral dumbfounding" in which participants were unable to generate adequate reasons for an intuitively given moral judgment. When presented with the case of consensual sex between adult siblings, for instance, almost everyone reports a strong emotional response and a feeling that it is wrong, even though he or she cannot articulate reasons for this opinion. Further highlighting the role of emotion, Haidt (2003) suggests some useful distinctions, sorting moral emotions (i.e., emotions in response to moral violations that motivate moral judgments and behavior) into other-condemning emotions (contempt, anger, and disgust), self-conscious emotions (shame, embarrassment, and guilt), the other-suffering family (sympathy and compassion), and the other-praising family (gratitude, awe, and elevation).

Similarly, the universal moral grammar theory proposes that the human mind is endowed with an innate moral grammar consisting of a domain-specific, complex set of rules, concepts, and principles that guide human social behavior in a community (by using concepts and models analogous to those used in the study of language; e.g., Hauser 2006b; Mikhail 2007). There is evidence, in fact, that people consistently judge harm caused by action as morally worse than the same harm caused by omission (action principle). Harm intended as means to an end is also judged as morally worse than the same harm foreseen as a side effect of reaching a goal (intention principle). Using physical contact to cause harm to a victim, moreover, is judged as morally worse than causing equivalent harm to a victim without using physical contact (contact principle).

Although there seems to be much evidence supporting the role of moral intuitions and principles in moral judgment and behavior (e.g., Haidt et al. 1993), other researchers qualify the strong assertions of the social intuitionist model and the universal grammar theory by pointing out that immediate intuitions and moral principles can also be informed and shaped by conscious reasoning (e.g., Pizarro and Bloom 2003; Takezawa et al. 2006). At least this is the case when participants have enough time to deliberate thoroughly (Suter and Hertwig 2011). Some principles, however (such as the intention principle with its distinction between intended and foreseen consequences) appear to be inaccessible to conscious reflection (see Cushman et al. 2006).

Incorporating psychological, developmental, and evolutionary perspectives, Haidt and Joseph recently proposed the moral foundation theory (MFT, Haidt and Joseph 2007). The MFT proposes that morality (perceived as a broad concept going beyond questions of harm and fairness) is built upon five innate and universally available "foundations" that have been selected through human evolution and are shaped during a person's individual development. The five foundations are: harm/care, fairness/reciprocity, in-group/loyalty, authority/respect, and purity/sanctity. Empirical studies aimed at verifying the MFT, for instance, showed that people with different cultural and political backgrounds (e.g., liberals vs. conservatives) differ in the degree to which they endorse each of the five moral systems (Graham et al. 2009). Glenn et al. (2009), moreover, found that higher scores in a measure of psychopathy predicted lower scores on the harm/care and fairness/reciprocity

subscales of a measure of the moral foundations, but showed no relationship with authority/respect, and only small correlations with in-group/loyalty and purity/sanctity. On a measure of "willingness to violate moral standards for money," psychopathy scores predicted greater willingness to violate moral concerns of any type. While the moral foundations approach enjoys a growing popularity (see e.g., www.moralfoundations.org), it must be stated that value und validity of this theory have already been put into question (see Suhler and Churchland 2011).

According to all three theoretical approaches highlighting the role of emotion (social intuitionist model, moral grammar and moral foundations theory), human morality relies at least to some degree on intuitive feelings and mechanisms, which are in part thought to be innate. Furthermore, it is stated that humans often have no conscious understanding of why they feel what they feel. The love felt toward one's own children and the anger felt toward someone who cheated on us can thus be considered as an adaptive mechanism of selective advantage that was shaped over the course of evolution (for further evolutionary considerations, see Prehn and Heekeren 2009).

8.3 The Neuroscientific Approach Investigating Morality

8.3.1 Lesion Studies Provide First Evidence of a Neurobiological Basis of Morality

A first hint indicating that morality (i.e., moral judgment and behavior) might have a neurobiological basis stems from the classic case of Phineas Gage, a railroad worker whose ventromedial prefrontal cortex (VMPFC) was damaged in an accidental explosion (Harlow 1848; Damasio et al. 1994). After his recovery, he showed preserved basic cognitive abilities and social knowledge (as indexed by IQ-tests and other measures) but an irresponsible and inappropriate social behavior, impaired decision making in everyday life, and a limited ability to experience emotions (a so-called "acquired sociopathy," Damasio et al. 1994).

More recent lesion studies report that damage to the prefrontal cortex (specifically, its ventromedial and orbitofrontal portions) leads to deficits in moral emotions, social behavior, and decision making (Saver and Damasio 1991; Barrash et al. 2000; Ciaramelli et al. 2007; Koenigs and Tranel 2007; Koenigs et al. 2007; Moretto et al. 2010; Thomas et al. 2011; Young et al. 2010b). For instance, it has been demonstrated that patients with lesions in the orbitofrontal cortex (OFC) display a defective ability with regard to anticipating negative consequences of one's choices during a gambling task, and they also do not experience regret afterwards (Camille et al. 2004). Notably, the age at which a brain injury occurred has been found to affect the degree and nature of the deficits. Anderson et al. (1999) showed that lesions in the VMPFC and OFC acquired in early childhood not only lead to impaired social and moral behavior but also seem to prevent the acquisition of factual knowledge about

the accepted standards of moral behavior in general (see also Eslinger and Biddle 2000). In sum, lesion studies provide evidence that at least some of the processes involved in moral judgment and behavior are dissociable (e.g., the distinction between acquisition and application of social rules mentioned above or identifying specific subcomponents such as the ability to anticipate punishment or to experience moral emotions).

Notably, lesion case studies have contributed significantly to theory evolvement in the field of moral cognition. Antonio Damasio, for instance, posited his "somatic marker hypothesis" based on his observations of patients with lesions of the VMPFC (Damasio et al. 1994; Damasio 1996). The somatic marker hypothesis suggests that emotional responses involving body function changes (labeled as "somatic markers"), such as an increase in heart rate or skin conductance, become associated over time with reward or punishment. After the repeated experience of certain bodily changes as response to the outcome of a certain action, such as a bad feeling when caught red-handed, the brain areas that monitor these bodily changes begin to respond whenever a similar situation with similar behavioral options arises. According to this theory, somatic markers are integrated in the VMPFC with other knowledge and planning functions and, thus, bias real life decision making in the future, especially in very complex situations with a high degree of uncertainty and ambiguity.

8.3.2 Neuroimaging Reveals a Distributed Functional Network Involved in Moral Cognition

It is important to keep in mind that lesion studies usually rely on a very limited number of cases with mostly very large and heterogeneous lesions. They give important hints (e.g., about single and double dissociations of cognitive processes) but cannot really reveal how the process of behaving appropriately or making moral judgments and ethical choices is organized in an intact human brain. In an attempt to overcome this limitation, cognitive neuroscientists have taken great advantage of the development of neuroimaging methods like functional magnetic resonance imaging (fMRI) which enables researchers to measure brain activity of healthy participants during a specific moral task such as judging a described behavior as being good or bad.

FMRI was first used in humans in 1991 (Belliveau et al. 1991). In its most popular variant, it measures cerebral changes of local hemoglobin oxygenation in response to a certain task (see Logothetis 2008 for a review). The method is based on the fact that the execution of a task leads to increased neuronal activity in the brain regions preoccupied with its processing. Increased neuronal activity is accompanied by a depolarization of neuron membrane potentials. Maintaining and re-establishing these potentials in groups of neurons requires an increased supply of energy and oxygen. This, in turn, leads to an increase in blood flow and blood volume in the capillaries of the activated brain tissue (commonly referred to as

"neurovascular coupling") resulting in both an increase of oxygenated hemoglobin, which overcompensates the actual supply of oxygen, and a concomitant decrease in deoxyhemoglobin concentration in this brain region. The changes of the local blood flow and blood volume as well as the relative change of deoxyhemoglobin in the blood concentration determine the so-called blood-oxygen level dependent signal (BOLD-signal) which can be detected due to the paramagnetic properties of deoxyhemoglobin by an MRI scanner with a powerful magnet (typically, 1.5 or 3.0 T). Although undoubtedly revolutionary for the study of mental phenomena, neuroimaging has some specifics that should be kept in mind when discussing its results.

First of all, it is important to know that during the performance of a task (e.g., when making a moral judgment or, in principle, at any time of wakeful activity) many if not all parts of the brain are activated to some degree. To identify brain regions that are specifically related to morality, most researchers are using "subtraction logic" in their experimental designs. Subtraction logic was pioneered by the Dutch physiologist Franciscus Cornelius Donders in reaction time experiments (see Donders 1969, translation of: Die Schnelligkeit psychischer Prozesse, first published in 1868, Archiv für Anatomie und Physiologie, 8, 657–681). The concept is based on the assumption of "pure insertion," which means that one cognitive process can be added to a pre-existing set of cognitive processes without affecting them and asserts that there are no interactions among the different components of a task. Although this assumption has not been validated in any physiological sense (Friston et al. 1996), it is applied in almost all fMRI studies mentioned in this chapter. In one of our recent studies (see Sect. 8.4.3; Prehn et al. 2008), for example, we compared neural activity during a moral judgment task with activity during a grammatical judgment task. The grammatical judgment task was designed to share almost all processes with the moral judgment task except the moral component: During both tasks, participants had to read sentences on a screen, to decide whether the actions described were "correct" or not (morally or grammatically), and then to respond with a button press. The grammatical judgment task, thus, controls for visual input, language processing, decision making, and motor output. In other words, colorful pictures of brains "lighting up" are actually artifacts of statistical analysis and selective presentation. They show those brain regions where a statistically significant level of increase or decrease in BOLD signal occurred during a task relative to a control state. In addition, results are mostly based on some kind of accumulation over a sample of only 20–30 subjects.

Following from the need to apply subtraction logic, data on the neural correlates of mental phenomena can only be as good as the underlying tasks and experimental paradigms. Experimental tasks have to be carefully designed so that they specifically activate the cognitive functions of interest and avoid the presence of other "confounding" factors that could possibly serve as an alternative explanation of the observed effects. The need to control for confounding factors in an experimental design prompts researchers to sometimes strip away real-life contexts and thereby undercut the "ecological validity" of a study.

FMRI studies investigating moral judgment and behavior have employed very different types of tasks and stimuli. As the study of morality has been traditionally

based in the domain of philosophy, many investigators have used complex moral dilemmas similar to those discussed by contemporary moral philosophers (e.g., Greene et al. 2001, 2004; Young et al. 2007; see Sect. 8.4). Other types of stimuli have been short sentences containing social norm violations (e.g., Heekeren et al. 2003, 2005; Prehn et al. 2008) as well as pictures or picture sequences with moral content (e.g., Bahnemann et al. 2010; Moll et al. 2002b). Some researchers invented innovative paradigms for the study of honest or dishonest "cheating" behavior (Greene and Paxton 2009), the making of charitable donations (Moll et al. 2006), or acts of reactive aggression and punishment (Buckholtz et al. 2008; Lotze et al. 2007). To study cooperative, altruistic, or self-interested behavior, economic decision-making tasks such as the Ultimatum Game or Reciprocal Trust Game have been used, during which two participants interact with each other via a computer interface and decide how to divide a given sum of money (Rilling et al. 2002; Sanfey et al. 2003; de Quervain et al. 2004; Spitzer et al. 2007). To investigate social interaction processes in more detail, a very interesting method is hyperscanning, by which two or even multiple subjects, each lying in a separate MRI scanner, can interact with one another while their brains are simultaneously scanned. Hyperscanning permits the study of brain responses that underlie processes during social interactions (for examples of scanning two participants to compute "between brain correlations," see Montague et al. 2002; Krueger et al. 2007).

Using these different tasks to investigate the variety of moral phenomena, neuroimaging studies have been remarkably consistent in revealing a functional network of brain regions involved. This network includes prefrontal brain regions, such as the ventromedial prefrontal cortex (VMPFC), the orbitofrontal cortex (OFC), and the dorsolateral prefrontal cortex (DLPFC), the temporal poles, the amygdala, the posterior cingulate cortex (PCC), the posterior superior temporal sulcus (pSTS), as well as the temporo-parietal junction (TPJ; for reviews, see Greene and Haidt 2002; Moll et al. 2003, 2005b, 2008a; Casebeer 2003; Lieberman 2007; Young and Dungan 2011).

The relative activation of a particular brain region during one experimental condition compared with another, however, does not tell us that much by itself. This information is only "spots on brains" until it is related to a hypothesis and to the developing picture of cognitive localization and integration in the brain. Complex tasks, such as judging whether a presented behavior is wrong in regard to moral conventions, comprise numerous cognitive and affective processes even when compared with a perfectly designed control task. These processes are represented by a distributed network of brain regions. Therefore, we cannot expect morality to be located in a specific and distinct brain area ("a moral center"). On top of that, different tasks often show highly overlapping neural networks. Processes thought to be different (such as emotion and cognition) are not necessarily subserved by separate and independent circuits (cf. Pessoa 2008). To be able to interpret a certain pattern of brain activity as a response to a specific task, one therefore needs very clear hypotheses about the involved mental processes. Such hypotheses can be derived from psychological theories or assumptions about the underlying neuronal mechanisms, for instance, resulting from lesion data or electrophysiological studies in monkeys and apes. A particularly nice way of linking imaging data with a targeted

function involves establishing some kind of intensity measure from the behavioral data (such as response times and post-hoc ratings), or the assessment of individual differences in personality traits, attitudes, and abilities. These measures can be used to demonstrate a corresponding change of intensity in the imaging data (on the question how to infer mental processes from imaging data, see Henson 2006; Poldrack 2006).

8.3.3 Neurostimulation Methods as an Attempt to Modulate Activity in the "Moral Brain"

In addition to the limitations already mentioned, it is important to understand that neuroimaging only allows us to see the changes in brain activity that are correlated with an experimental condition. Showing that one brain region is activated during a task does not show that this brain region is actually used or even necessary for the task. Neurostimulation methods like transcranial magnetic and direct current stimulation go beyond this correlational approach and can be used to demonstrate causality. If a subject performs worse on a task after a specific brain region was knocked out by an induced electric current for the time of task processing (also known as virtual lesion approach), this is much stronger evidence that this region is actually involved in performing the task.

Transcranial magnetic stimulation (TMS) is a non-invasive technique that is used to induce weak electric currents in the cortical tissue of the brain. TMS was first introduced for the investigation of the motor cortex by Barker and colleagues in 1985, who demonstrated that a magnetic pulse caused by a coil placed over the motor cortex produces an action potential (nerve impulse) which is transmitted from the cortex to the spinal cord and leads to a subsequent muscle contraction of the contralateral hand (Barker et al. 1985). The repetitive application of magnetic pulses (called repetitive transcranial magnetic stimulation = rTMS) in healthy participants is nowadays used to study a variety of cerebral functions either causing excitation or inhibition of neural activity (high-frequency rTMS leads to neuronal depolarization and enhanced neural firing, whereas low-frequency rTMS has been found to disrupt neural activity in a cortical area; see Guse et al. 2010).

As we will show in greater detail in the next sections, rTMS has been successfully used in the study of morality. For instance, it has been argued that the capacity to infer the actor's intentions and beliefs is central to moral judgment, which is associated with activity in the TPJ. Young et al. (2010a) showed that a disruption of neural activity in the right TPJ alters moral judgments insofar that participants in the TMS condition judged cases as less morally blameworthy when actors intended but failed to do harm than cases in which harm was caused accidentally. By using moral dilemmas that induce a conflict between emotion and reason (see Sect. 8.4.1), it was, moreover, found that a disruption of neural activity in the right DLPFC alters moral judgment and leads to an increase of utilitarian responses (Tassy et al. 2012).

Transcranial direct current stimulation (tDCS) is another non-invasive tool for modulating cortical excitability. Although already developed in the 1960s, this method has only recently been studied more extensively, with the advent of other brain activation techniques such as TMS (also with regard to its potential clinical application). TDCS protocols basically involve the application of two surface electrodes on the scalp of the participant, one serving as the anode and the other serving as the cathode. A 1–2 mA direct current then is applied for up to 20 min between the two electrodes, flows from the anode to the cathode, and leads to increases or decreases in cortical excitability dependent on the direction of the current. Anodal tDCS results in depolarization of the neurons underneath the electrode, hence causing an excitatory effect, whereas cathodal tDCS results in hyperpolarization and thus inhibition of cortical neurons (Been et al. 2007). In contrast to TMS, tDCS does not directly elicit action potentials (by means of suprathreshold resting membrane potential change) but renders neuronal populations more or less ready to fire in response to additional inputs. In other words, tDCS changes the likelihood that an incoming action potential will result in postsynaptic firing.

A number of studies has shown that tDCS applied to the prefrontal cortex has effects on cognition and mood. With regard to morality, it has been found that anodal stimulation over the right DLPFC (which results in an upregulation of neural activity) reduces risk-taking during decision making (Fecteau et al. 2007a, b). In contrast, an inhibition of the anterior prefrontal cortex through cathodal stimulation improves deceptive behavior; that is, it leads to better lying skills, reduced skin conductance responses, and feelings of guilt (Karim et al. 2010).

When following established safety protocols, both neurostimulation methods are safe and do not cause any side effects, apart from mild headache, discomfort because of unintended stimulation of nerves and muscles on the head, or itching underneath the electrodes (see Nitsche et al. 2003; Rossi et al. 2009). Although magnetic pulses and direct currents can only be administered on the surface of the cerebral cortex (only approximately 2 cm below the scalp), neurostimulation is not limited to cortical regions. Since the brain is an interconnected system, neuromodulation can also occur at distant but interconnected regions, such as deep brain structures like the amygdala (via its connections to the prefrontal cortex). In sum, neurostimulation methods offer an interesting perspective not only to the study of morality but also to a potential modulation of moral judgment and behavior (for instance, in therapeutic settings).

8.4 Studies Investigating the Role of Domain-Specific and General Capacities Contributing to Moral Judgment and Behavior

In the last decade, an increasing number of neuroimaging studies has been conducted to investigate the neural correlates of moral judgment. Some studies have focused on the neural correlates of moral judgment in general and in comparison to other non-moral (e.g., Moll et al. 2001, 2002a; Heekeren et al. 2003) or aesthetical

judgments (Tsukiura and Cabeza 2010). Others investigated the neural correlates of specific moral emotions (guilt, shame, regret and moral disgust or indignation; e.g. Coricelli et al. 2005; Moll et al. 2005a; Wagner et al. 2011). Moreover, studies have focused on the evaluation of one's own or other agents' actions and whether it matters if harm was caused intentionally or accidentally (e.g., Berthoz et al. 2002, 2006; Schaich Borg et al. 2006; Young et al. 2007; Young and Saxe 2008), on the influence of bodily harm on neural correlates of moral decision making (Heekeren et al. 2005), on the regulation of emotional responses (Harenski and Hamann 2006), and the impact of audience on moral judgments (Finger et al. 2006). Recent work was also dedicated to a differentiation of moral intuition and moral reasoning (Harenski et al. 2010a).

Neuropsychiatrists dealing with antisocial individuals and the biological foundations of criminal behavior have also contributed to the study of morality and have applied structural and functional MRI to investigate the "immoral brain" in clinical populations with difficulties in moral judgment and behavior, such as psychopaths (de Oliveira-Souza et al. 2008; Harenski et al. 2009, 2010b; Harenski and Kiehl 2010; Prehn et al. 2013). Kent Kiehl, in particular, has contributed enormously to the field by traveling with a mobile MRI scanner mounted on a truck and investigating more than 1,000 prison inmates. He linked emotional hypo-reactivity found in psychopaths to a dysfunctional paralimbic system including anterior and posterior cingulate, insula, OFC, amygdala, parahippocampal gyrus and superior temporal gyrus (Kiehl 2006; for reviews, see also Raine and Yang 2006; Blair 2008; Glenn and Raine 2008).

Below, we will look at three research foci, dedicated to the question of how and where moral judgment is processed in the human brain: (1) the relationship of emotional and cognitive subsystems contributing to moral judgment and behavior, (2) the role of social cognitive processes and mental state reasoning, and (3) the influence of individual differences (see also the review by Young and Dungan 2011).

8.4.1 Competing Emotional and Cognitive Subsystems

As presented in the Theories section, recent psychological theories on morality as well as neuropsychological models (e.g., the Somatic Marker Theory) claim that emotions are central for moral judgment and behavior. Following this "affective revolution" in psychology and cognitive sciences, many (early) neuroscientific studies were dedicated to the question of whether moral judgment and behavior is guided by reason or emotion.

One of the first studies investigating which brain regions are involved in moral judgment was the study by Greene and colleagues published in 2001. In this study, participants were presented with two types of dilemmas. One type is represented by the "trolley dilemma" and the other by the "footbridge dilemma". In the trolley dilemma, the participant is asked to consider the following situation: A runaway

trolley is quickly approaching a fork in the tracks. On the tracks extending to the left is a group of five railway workmen. On the tracks extending to the right is a single railway workman. If one does nothing the trolley will proceed to the left, causing the deaths of the five workmen. The only way to avoid the deaths of these workmen is to hit a switch on your dashboard that will cause the trolley to proceed to the right, causing the death of the single workman. After presenting this story, the participant in the experiment is asked to respond whether it is appropriate to hit the switch to avoid the deaths of the five workmen. In the footbridge dilemma the situation is slightly different. Again, a runaway trolley is heading down the tracks toward five workmen who will be killed if the trolley proceeds on its present course. The participant now has to imagine being on a footbridge over the tracks with a stranger and in between the approaching trolley and the five workmen. The only way to save the lives of the five workmen is to push the stranger off the bridge and onto the tracks below where his large body will stop the trolley. The stranger will die as a result, but the five workmen will be saved. After presenting this situation, the participants again are asked to respond whether it is appropriate to push the stranger onto the tracks to save the five workmen.

By comparing neural activity during reasoning about these two types of dilemmas, Greene et al. (2001) found that reasoning about dilemmas that are emotionally engaging such as the footbridge dilemma (i.e., personal dilemmas or dilemmas in which physical harm is caused to another person directly by the agent) as compared to dilemmas that are less emotionally engaging such as the trolley dilemma (i.e., impersonal dilemmas or dilemmas in which physical harm is caused to another person only indirectly) activate the medial prefrontal cortex, the PCC, and the PSTS.

In a later study, Greene et al. (2004) further investigated how people solve particularly difficult personal moral dilemmas. An example for a very difficult personal dilemma is the "crying baby dilemma" that is used to bring cognitive and emotional processes into tension. In this dilemma, the participant has to decide whether it is appropriate to smother his or her own crying baby to save his or her life and the lives of other refugees hiding from enemy soldiers in a basement. Participants usually answer very slowly when presented with such a dilemma and do not reach a consensus on this issue.

The comparison of neural activity during utilitarian and non-utilitarian decisions (i.e., neural activity when participants decided that smothering the crying baby to save more lives is appropriate vs. neural activity when participants decided that smothering the crying baby is not appropriate) revealed increased activity in brain regions associated with abstract reasoning, conflict processing, and cognitive control such as the DLPFC and the anterior cingulate cortex (Greene et al. 2004). One interpretation of these results is that a conflict associated with such a difficult moral question is detected by the anterior cingulate cortex which then recruits control mechanisms and rational reasoning processes associated with neuronal activity in the DLPFC. These control processes help to resolve the conflict and to override prepotent emotional responses to make a utilitarian decision (see also Greene 2007; Greene et al. 2008).

To further investigate the role of emotion in moral judgment, Koenigs et al. (2007) tested a group of patients with VMPFC lesions. One of the most robust clinical finding in VMPFC patients is that they have blunt or flattened affects. For instance, in laboratory investigations VMPFC patients exhibit diminished autonomic arousal and subjective feelings in response to emotionally charged pictures and, according to their spouses, reduced feelings of empathy and guilt (Eslinger and Damasio 1985; Barrash et al. 2000). When confronted with moral dilemmas, these patients were more likely to choose a "rational" and utilitarian option (e.g., smothering a crying baby to save a group of refugees hiding from soldiers that normally would elicit a strong emotional response, see above) than healthy controls (Koenigs et al. 2007; Thomas et al. 2011; for similar results in a different sample of patients with VMPFC lesions, see Ciaramelli et al. 2007; but see also Kahane and Shackel 2008 for methodological problems in the study of utilitarian and non-utilitarian moral judgments).

However, it cannot be concluded that VMPFC patients in general decide more rationally. When engaged in real social situations involving frustration or provocation, the same participants exhibit exaggerated anger and emotional outbursts (Barrash et al. 2000). Using another experimental paradigm, namely the Ultimatum Game, Koenigs and Tranel (2007) found that patients with VMPFC lesions were influenced even more strongly by emotional reactions in their decisions. In the Ultimatum Game, two players are given a sum of money ($100) and one opportunity to split it. The first player proposes how to divide the sum between each other, and the second player can either accept or reject this proposal. If the second player accepts, the money is split according to the offer. If the second player rejects, neither player receives anything. Therefore, a "rational" second player would accept any offer, no matter how low, because getting a low amount of money should be more rewarding and better than getting nothing at all. The "irrational" rejection of a low and unfair offer (e.g., when the offer is below $20), in contrast, has been attributed to an impulsive reaction related to negative feelings such as anger and poor regulation thereof (Sanfey et al. 2003). Using this economic decision-making task, the authors demonstrated that patients with VMPFC lesions were more likely to make irrational choices and rejected low and unfair offers more often than healthy controls.

Together, the two different lesion studies show that VMPFC patients respond rationally in the face of abstract hypothetical scenarios related to the welfare of others, but irrationally in a real social setting involving their own self-interest. Although the precise role of VMPFC in moral judgment needs to be further investigated, it can be concluded that damage to this region disrupts the integration of emotion and reason in decision making. Further research is needed to investigate why a disruption of VMPFC function takes different forms in different circumstances (moral dilemmas vs. economic decision making) and whether emotion in general makes moral judgment better or worse (for further discussion and different explanations, such as a selective impairment of "prosocial sentiments" with a preserved capacity for anger or indignation, see reviews by Greene 2007; Young and Koenigs 2007; Moll and de Oliveira-Souza 2007; Young and Dungan 2011).

8.4.2 Social Cognitive Processes and Mental State Reasoning During Moral Judgment

As mentioned in the Theories section, the cognitive-developmental theory proposes that morality significantly relies on the way individuals understand and think about social situations. How people think about social situations is assumed to mature as a result of an active interaction of the individual with his or her social environment. Although this theory has been criticized and replaced by more recent models that place more emphasis on the role of emotion, neuroscientific work provides evidence that morality critically depends on a set of social cognitive abilities that allow people to take others' intentions, beliefs, and desires (or any kind of mental state) into account when making moral judgments.

Following the results of Rebecca Saxe and colleagues, the TPJ (mostly on the right hemisphere) appears to support important cognitive functions of mental state reasoning ("theory of mind") in moral judgment. These functions include the initial encoding of the agent's mental state (Young and Saxe 2008), the integration of that information (Young et al. 2007), spontaneous mental state inference (Young and Saxe 2009), and even post-hoc mental state reasoning to justify moral judgments (Kliemann et al. 2008; Young et al. 2011).

To investigate the impact of mental state reasoning in moral judgment, Young et al. (2007) used highly hypothetical scenarios in their studies. An example for such a hypothetical scenario is the following story: "Grace and her friend are taking a tour of a chemical plant. When Grace goes over to the coffee machine to pour some coffee, Grace' friend asks for some sugar in hers. The white powder by the coffee is not sugar but a toxic substance left behind by a scientist. Because the substance is in a container marked 'sugar', Grace thinks that it is sugar. Grace puts the substance in her friend's coffee. Her friend drinks the coffee and dies." (example from Young et al. 2007). By a systematic variation of outcomes (dying or not) and beliefs (she thinks it is sugar or she thinks it is toxic) the authors showed that neural activity in right TPJ was greatest for attempted harm; that is, in cases where protagonists were condemned for actions that they believed would cause harm, even though the harm did not occur.

Disrupting activity of the right TPJ by rTMS also disrupts the impact of mental state reasoning for moral judgment (Young et al. 2010a). Specifically, it reduces the role of intentions in moral judgment and increases, in contrast, the role of outcomes. For example, participants judged cases when actors intended but failed to do harm (including murder) as less morally blameworthy than cases in which harm was caused accidentally. However, TMS significantly reduced but did not completely eliminate the impact of mental state reasoning. In fact, there is evidence from many neuroimaging studies that the VMPFC, OFC, the temporal poles, and the PSTS also contribute to social cognition and mentalizing (Saxe et al. 2004; Amodio and Frith 2006; Mitchell 2009). Further evidence of a particular role of the VMPFC is given by a lesion study conducted by Young et al. (2010b) showing that patients with lesions in the VMPFC also judged cases with intended but failed harm as less blameworthy than controls.

In some research performed in our working group (Bahnemann et al. 2010), we investigated whether activity in the PSTS/TPJ region evoked by three tasks with increasing complexity (namely, detecting movements of bodies, making inferences concerning intentions, judging whether a behavior is morally good or bad) represents a common or distinct processes. We found an overlap of neural activity between all three tasks in right PSTS, but also a hierarchically increasing recruitment of the left PSTS and bilateral TPJ representing increasingly more complex processing of the social situation in the intention reading and moral judgment task.

In sum, these findings suggest that mental state reasoning represents a key cognitive component of moral judgment (i.e., moral judgments depend on information about agents' beliefs and intentions). The neural substrates that support this function, therefore, constitute an important part of the "moral brain network," in which the PSTS/TPJ and VMPFC are critical nodes.

8.4.3 The Influence of Individual Differences in Moral Judgment Competence

As already mentioned, to investigate the moral brain, some studies also investigated deviations and limitations in mental capacities thought to be relevant for moral judgment and behavior. These studies highlight the role of individual differences and provide direct evidence that particular abilities, such as empathy, perspective taking, and mental state inferences (in which patients differ from healthy controls) have a great impact on moral judgment and behavior.

The two studies by Koenigs and colleagues in patients with lesions of the VMPFC (Koenigs and Tranel 2007; Koenigs et al. 2007) discussed earlier, however, showed that a disruption of the affective/intuitive decision making component, on the one hand, improves utilitarian moral judgment, whereas on the other hand, economic decision making in the Ultimatum Game was impaired. Referring to the question of whether emotion makes moral cognition better or worse, Talmi and Frith (2007) stated that "The challenge, then, is for decision-makers to cultivate an intelligent use of their emotional responses by integrating them with a reflective reasoning process, sensitive to the context and goals of the moral dilemmas they face. If decision-makers meet this challenge, they may be better able to decide when to rely upon their emotions, and when to regulate them." (Talmi and Frith 2007: 866). Therefore, the question is not only which processes are involved in moral judgment but also how competently a decision maker can integrate the different (emotional and cognitive) processes sensitive to the context of the particular social situation he or she faces.

As mentioned in the Theories section, a current theoretical approach addressing this particular "intelligent use" of emotional and reasoning processes sensitive to the context of the specific social situation is the Dual Aspect Theory by Georg Lind. Referring to Kohlberg's notion of morality as an ability, Lind defines "moral judgment competence" as the ability to apply certain moral orientations in a consistent

and differentiated manner in varying social situations. Thus, social norms and values held to be virtuous in a culture or subculture are linked by means of moral judgment competence with everyday behavior and decision making (Lind 2008).

To investigate how individual differences in moral judgment competence are reflected in changes in brain activity during a moral judgment task, we conducted an fMRI study and measured neural activity while 23 participants made either moral or grammatical judgments. Participants were required to decide whether sentences were morally or grammatically correct or not. We correlated neural activity during these tasks with individual scores in moral judgment competence (Prehn et al. 2008). Individual moral judgment competence was measured using the Moral Judgment Test (MJT; Lind and Wakenhut 1980; Lind 2006, 2008; www.uni-konstanz.de/ag-moral/mut/mjt-intro.htm).

The MJT confronts a participant with two complex moral dilemmas. In one dilemma (the doctor dilemma), for instance, a woman had cancer with no hope of being cured. She suffered terrible pain and begged the doctor to aid her in committing medically assisted suicide, and the doctor complied with her wish. After presentation of this short story, the participant has to indicate to which degree he or she agrees or disagrees with the solution chosen by the protagonist. After that, the participant is presented with six arguments supporting (pro-arguments) and six arguments rejecting (counter-arguments) the protagonist's solution which the participant has to rate with regard to its acceptability on a nine point rating scale ranging from –4 (highly unacceptable) to +4 (highly acceptable). Each pro- and counter- argument represents a certain moral orientation according to the six Kohlbergian stages.

In general, adult participants—in contrast to children or adolescents—prefer more elaborate arguments (i.e., adults rate more elaborate arguments as more acceptable than low level arguments) in line with having achieved a higher developmental stage of moral judgment. However, adult participants differ greatly in their ability to apply these orientations consistently especially when confronted with counter-arguments (i.e., arguments which are against their own opinion). This means that they rate more elaborate arguments as acceptable only when they represent their own opinion and reject all counter-arguments regardless of whether they are elaborate or not. The moral judgment competence score (C-score, the MJT's main score) reflects the ability to consistently or, in Lind's terms, competently apply a certain moral orientation and is calculated as an individual's total response variation.

By providing a measure of consistency, Lind's approach clearly goes beyond what we may ordinarily call "moral competence" as well as the Kohlbergian approach which focuses merely on moral orientations and the level of reasoning.

To our knowledge, the MJT is the only available test that provides a measure of moral judgment competence independent from a person's moral attitudes and values, in contrast to other instruments such as Kohlberg's Moral Judgment Interview (Colby et al. 1987), the Defining Issue Test (Rest 1974), the Sociomoral Reflection Measure (Gibbs et al. 1992), and the newly developed Moral Foundations Questionnaire (Graham et al. 2011), which all mostly assess individual moral orientations. The MJT has been proven to be a valid and reliable psychometric test.

For instance, moral judgment competence has been associated with responsible and democratic behavior. Translated in many languages, it also has been successfully used in scientific research (i.e., probing theoretical assumptions on moral development) and in evaluating educational programs (Lind 2008).

Contrasting neural activity during moral judgments with grammatical judgments, we found in line with the literature increased activation in the left VMPFC, the left OFC, the temporal poles, and the left PSTS for moral judgment. Regarding moral judgment competence, our sample of 23 participants showed a wide range of C-scores (maximum score = 62.74, minimum score = 5.55, mean = 36.93, standard deviation = 16.67). We correlated individual scores of moral judgment competence with neural activity during moral judgment and found that C-scores were negatively correlated with changes in BOLD activity in the right DLPFC during moral judgments contrasted with grammatical judgments. That is, participants with lower C-scores recruited the right DLPFC more than those with higher competence during moral judgment. Additionally, we investigated whether individual differences in moral judgment competence also modulate BOLD activity in the cerebral network engaged in moral judgment. An additional median split analysis revealed greater activity in the left VMPFC and the left PSTS in participants with comparably low moral judgment competence, specifically during identification of moral transgressions.

Finding a specific neural activation that reflects differences in moral judgment competence provides neuroscientific support for the Dual Aspect Theory by Lind. In the literature, greater neural activity in participants with lower ability in a certain cognitive task has been associated with compensation and an increased recruitment of mental resources (Rypma et al. 2006). As described earlier, moral judgment competence assessed with the MJT represents the ability to apply individual moral orientations in a consistent and differentiated manner in varying social situations. The increased activity in right DLPFC and left VMPFC/PSTS in participants with lower competence can thus be interpreted as reflecting higher processing demand due to a controlled application of moral orientations and an increased involvement of social cognitive and affective processes (such as mentalizing, estimating the value of possible outcomes of a behavior, and the experience of moral emotions) during the decision-making process (for extended discussion of the results regarding the brain regions involved, see Prehn et al. 2008; Prehn and Heekeren 2009).

Further neuroscientific evidence for a role of the right DLPFC in moral judgment and the implementation of morally appropriate behavior comes from a study using rTMS. Here also, a disruption of the right (but not the left) DLPFC reduces the subject's willingness to reject their partner's intentionally unfair monetary offers in the Ultimatum Game. Importantly, subjects were still able to judge the unfair offers as unfair. This indicates that the right DLPFC plays a key role especially in the implementation of fairness-related behaviors (Knoch et al. 2006).

Thus, both our own study and the rTMS study provide complementary evidence that there are specific brain regions crucial to the execution of morally appropriate behavior (see also Fecteau et al. 2007a, b; Knoch and Fehr 2007; Knoch et al. 2008).

8.5 Conclusion

In this chapter, we took a look at current psychological models on moral judgment from a neuroscientific point of view, specifically introducing neuroscientific methods (clinical studies, fMRI, and neurostimulation) as powerful tools to investigate the processes underlying moral judgment and behavior in the human brain.

Since the availability of these tools, a multitude of studies have been conducted to investigate how and where morality is represented in the human brain. These studies used a variety of tasks and experimental paradigms, which are more or less ecologically valid, to investigate numerous different aspects, such as the role of emotion in moral judgment (see Sect. 8.4.1), social cognitive processes (Sect. 8.4.2), or the impact of inter-individual differences (Sect. 8.4.3).

What can we learn from these studies conducted so far? First of all, complex tasks, such as judging whether a certain behavior is "wrong" with regard to moral conventions, recruit a network of distributed brain regions supporting various subsystems and a number of different processes. For example, when confronted with a difficult moral dilemma, the situation presented to participants of a study needs to be represented in working memory. Simultaneously, accepted standards for social and moral behavior (i.e., norms and values considered to be virtuous in a culture or subculture) have to be retrieved from long-term memory. The presented behavior will then, subsequently, be evaluated with regard to these (re-)presentations. This evaluation process involves cognitive and affective sequences alike. Processes and mechanisms at the "cognitive" end of the spectrum, for example, might include social cognition, providing an understanding of the social situation and taking into account the cultural context and the intentions of the people involved. The individual would have to infer whether an actor did what he or she did on purpose or not. Processes at the "affective" end of the spectrum include the feeling of emotions such as guilt, sympathy, shame, anger, or disgust, when people are harmed and social norms are violated. The processes involved can be rational and accessible for conscious reflection (e.g., the controlled application of legal rules) or can be more subconscious, intuitive, and automatic. Depending on the circumstances (the emotional content of the situation, a person's involvement, the necessity of reaching a utilitarian decision, etc.), either the cognitive or the affective aspect dominates in the decision-making process. For situations containing issues of life and death, which immediately affect the survival of an individual or his or her successful reproduction (e.g., in the case of consensual sex between siblings), judgment mechanisms might have evolved as a result of biological adaptation over the course of evolution and go on to proceed automatically and effortlessly without conscious reflection. Such mechanisms are suggested in the social intuitionist model by Jonathan Haidt. In contrast, coming to the conclusion that downloading music files illegally from the internet is harmful to society, for example, requires more abstract reasoning processes as proposed by the Kohlbergian model. As psychological theories and our own studies suggest, individual differences in information processing can also influence judgment processes and moral behavior. Individuals, for instance, considerably

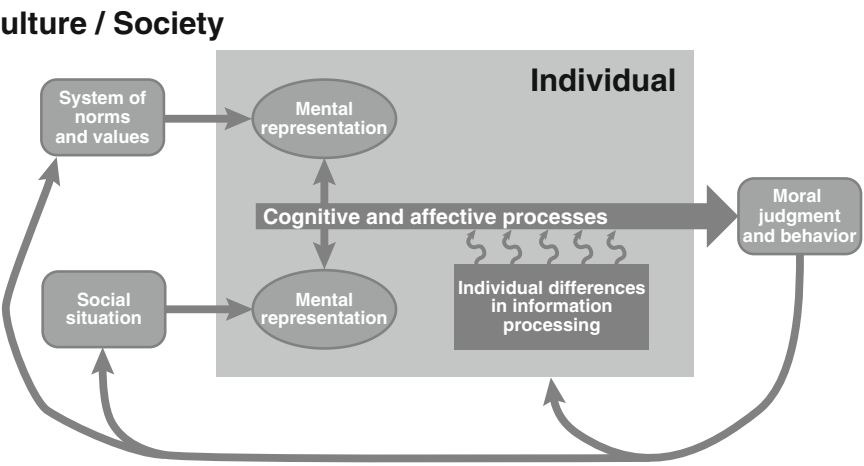

Fig. 8.1 A working model of moral judgment and behavior comprising different information processing modules described in this chapter. The judgment process includes cognitive and affective processes, which might be either more rational/conscious or intuitive/automated. In addition, information processing is modulated by individual differences such as the developmental stage of reasoning, moral judgment competence, or empathy. Finally, moral judgment and behavior have, in turn, an impact on the individual, his or her mental representations, as well as on future social interactions and the system of norms and values

differ in moral development, moral judgment competence, and empathy. Finally, moral judgments taken and the (moral or immoral) behavior shown could have, in turn, an impact on the individual (specifically on his or her experiences and mental representations), as well as on future social interactions and the society (e.g., when new laws are adopted to protect people which are treated immorally).

Taking all these different processes and aspects into consideration, we cannot expect to find morality located in a specific and distinct brain area ("a moral center"). The current model of the brain is an interconnecting networking model of information processing that integrates different kinds of information. That is, the "(moral) brain" can be broken up into several modules whose functions originally have nothing to do with morality (e.g., emotion, social cognition, cognitive control, etc.). By employing neuroscientific methods, the different "brain modules" or "process units" contributing to moral judgment and behavior can be (re-)presented, "visualized," relatively analyzed, manipulated, selectively impaired, and (potentially) also modulated and trained. Neuroscience, thus, helps to imagine how processes are organized and grants a potential opportunity to alter human brain processing if something "is going wrong." The goal of an empirically informed neuroscience of ethics is to integrate the various sets of subsystems contributing to moral judgment and behavior based on solid hypotheses stemming from all kinds of research fields. For a simple graphic idea of how the brain organizes moral processes, see Fig. 8.1.

A neuroscience of ethics, with its existing theoretical models, in our view, is highly beneficial to psychological studies and the field of psychology as such, both on a theoretical and practical level. We have taken the first steps toward "neuromorality tests" and "neuropsychotherapy," although we are still far from individual-based testing and neuropsychological assessment in forensic and pedagogical settings. One example of this endeavor is our study (Prehn et al. 2008) on the neural correlates of moral judgment competence (as defined following the Dual Aspect Theory by Lind). The data presented in this study strongly support the notion that morality can be considered as both a capacity and in terms of individual differences in the ability to apply frequently conflictual moral standards in a consistent and differentiated manner in varying social situations. It should be noted that it is presently unclear how exactly moral judgment competence, as measured by the MJT, maps on other cognitive abilities such as general intelligence. Future studies will have to address this question together with the question of how individual differences in moral judgment competence modulate processing during other kinds of tasks or day-to-day behavior. In addition, research is needed to investigate the possibility of changing one's "ethics" through methodological training.

We hope that the review of neuroscientific studies on morality demonstrated that neuroscientific empirical research, to say the least, helps to disentangle the different processes involved in moral judgment and behavior. Thus, neuroscience helps to test and verify numerous assumptions made in psychological theories and can give an empirically informed opinion on long-time philosophical discussions aiming to differentiate between hotly debated concepts such as "intentions vs. consequences," "intuition vs. reason," or "morality of justice vs. care ethics." However, and as we also emphasized in this chapter, one should not overlook the limitations or preconditions of neuroscientific methods and should always keep in mind the specific part of the broad philosophical notion of morality or ethics we are talking about at a given point in time.

Chapter 9
Using Experiments in Ethics – Ethical Conservatism and the Psychology of Moral Luck

Shaun Nichols, Mark Timmons, and Theresa Lopez

We should ask why blame should be thought to be such a fearsome thing that, like weapons of destruction, it can be loosed only in circumstances that ultimately justify it. (Williams 1995: 243–244)

9.1 Ethical Conservatism

Psychological evidence detailing why we believe what we believe can provide a powerful basis for challenging the warrant of those beliefs. If psychology shows that a certain common belief derives from epistemically defective processes, then this threatens to undercut the epistemic authority of the belief. The canonical case of this sort of debunking argument is Freud's critique of religion (1927). Freud argues that religious belief is a product of wish-fulfillment, not reason; as a result, he says, we should regard our religious beliefs as unwarranted. On the heels of psychological research on philosophically charged domains, these kinds of debunking arguments have been making a comeback. Debunking arguments have been advanced in the domains of metaphysics (e.g., Scholl 2007), metaethics (e.g., Nichols 2008), consciousness (e.g., Fiala et al. 2011), and normative ethics (e.g., Greene 2008; Singer 2005).

Underlying many of these debunking arguments is a commitment to a plausible principle: if a belief results from a process that is neither rational nor reliable, this calls into question the justificatory status of the belief. The advocates of debunking arguments have typically given very few details on what exactly makes a process epistemically defective, but the basic idea is obvious enough. Some beliefs are clearly based

S. Nichols (✉) • M. Timmons
Department of Philosophy, University of Arizona, Tucson, NY, USA
e-mail: sbn@email.arizona.edu

T. Lopez
Department of Philosophy, Hamilton College, Clinton, NY, USA

M. Christen et al. (eds.), *Empirically Informed Ethics: Morality between Facts and Norms*, Library of Ethics and Applied Philosophy 32, DOI 10.1007/978-3-319-01369-5_9,

on rational inference. Consider, for instance, the belief that dachshunds are mammals. I believe this because I believe that dachshunds are dogs and dogs are mammals. Thus, that belief will count as "rational" by present lights. Other beliefs do not derive from rational inference but come from reliable processes. My current perceptual belief that the paper before me is rectangular is based on a visual process of shape-recovery that is (apparently) highly reliable and tends to produce true beliefs about the world. Beliefs based on rational or reliable processes are typically taken to be in decent justificatory standing. The principle underlying debunking arguments is that beliefs that are formed by processes that are *neither* rational nor reliable are in bad justificatory repair. Although we find this principle to be plausible when it comes to metaphysics, metaethics, and consciousness, we think the situation is different in the domain of normative ethics, where there is room for psychology to play a positive rather than destructive role. The contrast with metaphysics is instructive. If we find out that our metaphysical commitment to causal powers is just the residue of an a-rational and a-reliable bag of tricks, this would be reason to suspend the commitment to causal powers in our reflective moments. We maintain (*pace* Singer 2005 and Greene 2008) that the situation is different with normative ethics. It is quite likely that much of commonsense ethics derives in part from fundamentally a-rational and a-reliable processes. In particular, it is likely that much of what we care about in ethics depends in part on a-rational and a-reliable emotional processes (see, e.g., Blair 1995; Nichols 2004; Prinz 2007b). For instance, if we did not have a natural revulsion to suffering in others, we would likely not have deep-seated norms that prohibit killing and maiming, nor norms that promote helping suffering strangers.[1] Even if these norms are rooted in processes that are neither rational nor reliable, we would not conclude from this that we should suspend our rejection of killing and maiming. If we give up all of the ethical judgments that critically depend on our a-rational and a-reliable processes, then we might well be left with an ethical world view more barren than almost anyone is willing to accept.

Rather than adopt such an emaciated ethics, one might take on a conservative position regarding certain ethical commitments. This view, which we'll call *ethical conservatism,* holds that for a certain class of ethical commitments, the (presumed) fact that they lack a rational or reliable grounding does not undermine their normative authority. This is not the same as saying that the commitments are unassailable. Rather, the claim is simply that we are not compelled to suspend the commitments just because the commitments are not based on reason.

According to ethical conservatism, certain normative commitments need not flow from rationality to retain their authority.[2] Some of our normative commitments can

[1] This is not to say that the norms we have are equivalent to emotional responses. Rather, the claim is that emotions played a critical role in our coming to have, embrace, and retain those particular norms (Nichols 2004).

[2] What we are calling "ethical conservatism" is to be distinguished from what is called "epistemic conservatism"—a cluster of related positions in epistemology according to which (roughly) one is justified in holding a belief so long as one does not possess good reasons for believing that the proposition believed is false. (Versions of this kind of view have been defended by Chisholm 1981, Harman 1986, Lycan 1988, Kvanvig 1989, and Adler 1996. The main varieties of this form of

remain legitimate even if it turns out that they are the product of patently a-rational and a-reliable processes. It remains to be said *which* normative commitments have this special status. We propose that this special status be accorded to normative commitments that are *entrenched* in our psychology. There are two elements to a commitment's being psychologically entrenched. First, entrenched commitments are not the product of consciously available inferences from other norms or facts. So, for many people, norms about the propriety of abortion, for instance, are not entrenched. For norms about the propriety of abortion depend, for many people, on views about factual matters, like whether the fetus has a soul. By contrast, an entrenched norm does not depend on inferences from factual beliefs.[3] The second feature is that entrenched commitments are rooted in human emotions. We are naturally inclined to be emotionally committed to them. This excludes certain norms of etiquette, such as *the napkin should be placed to the left of the plate*. That norm is arbitrary with respect to our natural emotional endowment. Not so for a norm like "it's wrong to kill people." That's a norm that resonates with our emotional repertoire.

In sum, then, our ethical conservatism reserves a special status for normative commitments that are rooted in human emotion and are not inferentially dependent on other norms or facts. The legitimacy of such entrenched commitments is not undermined merely by discovering that those commitments derive from a-rational and a-reliable factors.

With this background in place, we maintain that psychology can play a positive, rather than a debunking, role for normative ethics. From the armchair it is not easy to determine whether an ethical conviction is entrenched. Some of our ethical convictions depend on factual beliefs. Others turn out to be the product of general biases. Psychology provides a critical resource for assessing whether a normative judgment is the result of prior factual beliefs, general biases, or some other intervening factor. As a result, psychology can help us to assess whether a given commitment is entrenched.

Moral luck is a promising test case here. A drunk driver who runs over a pedestrian seems more blameworthy than an equally negligent driver who, through sheer luck, doesn't encounter any pedestrians. Let's call these asymmetric judgments "luck-based," since the judgments differ even though the only difference between the agents is that one is unlucky and the other is not. Some maintain that in luck-based judgments the influence of outcome on blame is mediated by a rationally proper inference: a bad outcome provides evidence about the agent's prior epistemic state.

---

conservatism are usefully discussed in Vahid 2004.) Our brand of ethical conservatism is more robust than epistemic conservatism. Finding out that some empirical belief one holds is the product of an a-rational and a-reliable mechanism would presumably be taken by the epistemic conservative to constitute a good reason to stop holding the belief in question. However, according to our ethical conservatism, even if it turns out that some of one's ethical commitments are products of a-rational and a-reliable mechanisms, this fact about them does not automatically count as a good reason to reject their normative authority. In this way, ethical commitments are similar to certain aesthetic judgments. Finding out that one's aesthetic tastes (and related judgments) in music are grounded in a-rational and a-reliable mechanisms is not itself a good reason for rejecting those tastes and related judgments.

[3] We leave open the possibility that an entrenched norm might depend on some kind of subdoxastic inference (Stich 1978).

Others maintain that outcome has an *improper* influence on blame, and that its influence is the result of a general epistemic bias. If either of these accounts is right, then, while outcome influences blame, it does not do so because of an entrenched commitment to moral luck. Rather it is simply the product of differential epistemic evaluation in light of bad outcomes. If luck-based judgments are mediated just by epistemic considerations, then the commitment to moral luck is not entrenched. In order to see whether we have an entrenched commitment to moral luck, we need to explore in detail how outcome affects judgments of blame.[4]

Drawing on recent experimental results, we will suggest that outcome has an effect on blame that is *not* mediated by epistemic inference. This, we argue, provides prima facie evidence that outcome-based blame is an entrenched commitment, and hence that a further defense of the authority of this commitment is not required. In addition, we will present some evidence of our own that suggests a rather natural way to relieve the tension associated with judgments of moral luck.

9.2 The Problem of Moral Luck

In 1989, the Exxon Valdez ran aground in Prince William Sound, dumping over ten million gallons of oil into the Gulf of Alaska. Ever since, the captain of the vessel, Joseph Hazelwood, has been the focus of intense blame. Lots of other ship captains have, no doubt, been equally negligent. However, if, say, Captain Jones was just as negligent as Hazelwood, but Jones got lucky with the currents and avoided a disaster, we would not heap an equal amount of blame on him. The fact that, through sheer (bad) luck, Hazelwood is responsible for an environmental catastrophe seems to earn him an extra dose of blame.

The impact of outcome on blame is enshrined in the law. In February 2009, Briana Bonds was shot in the head by her ex-boyfriend, Juan Kemp. Bonds' hair was very tightly weaved, however, and apparently this stopped the bullet from entering her skull. Kemp was charged with domestic assault and armed criminal action. These charges will carry much lighter sentences than murder. But it was through pure luck that Kemp's bullet didn't kill Bonds. Who could have anticipated that *hair* would stop a bullet?

9.2.1 Setting the Problem

The reason blaming and punishing someone more for bad luck in outcomes seems puzzling is because it conflicts with the idea that a person can only be blamed for

[4] Most of the literature on moral luck focuses on cases in which agents have bad luck and are consequently blamed more harshly. It is an interesting question whether when agents have *good* luck, this will have a parallel affect on judgments of praise. But since there is very little work on moral good luck, we set it aside for the purposes of this paper.

what is within their control.[5] As Nagel puts it, "[p]rior to reflection it is intuitively plausible that people cannot be morally assessed for … what is due to factors beyond their control" (Nagel 1979: 25). Some version of this idea, the *control principle,* is widely accepted in the philosophical literature. Ed Royzman and Rahul Kumar characterize the principle as follows: "in assessing how morally faulty a person's conduct has been, only those considerations that it is reasonable to treat as having been within the sphere of the agent's rational control ought to be taken into account" (2004: 330; see also Nelkin 2008).

Moral luck runs in direct opposition to this principle. The idea behind moral luck is that it is sometimes appropriate to assign more blame just because of a bad outcome. Dana Nelkin puts it as follows: "Moral luck occurs when an agent can be *correctly* treated as an object of moral judgment, despite the fact that a significant aspect of what he is assessed for depends on factors beyond his control" (Nelkin 2008).[6] Typically, cases of moral luck involve agents who have been negligent in one way or another. Accordingly, Darren Domsky characterizes moral luck as the idea that "Negligent agents who by luck bring about bad outcomes are more blameworthy than equally negligent agents who by luck do not" (Domsky 2004: 445).

The problem, then, is that we hold people more blameworthy when their actions result in bad outcomes, even when the outcome was out of their control. At the same time, it seems like people should only be blamed for what is under their control. On the one hand it seems like luck should not matter to blameworthiness. On the other hand, our judgments of blameworthiness often seem to vary depending on the luck of the agent. It should be noted that the problem of moral luck is not a problem restricted to the judgments of trained philosophers. On the contrary, Nagel says that the problem emerges from our "ordinary idea of moral assessment" and the "application of ordinary standards" (1979: 27). But these ordinary ideas are also supposed to reflect *prima facie* ethically correct positions. That is why moral luck presents a philosophical puzzle rather than just a folk confusion.

9.2.2 Epistemic Explanations of Outcome-Based Blame

One prominent approach to moral luck holds that luck mediates our judgments of blame by affecting our evaluations of agents' *epistemic* states at the time of their decision. Epistemic accounts of luck-based judgments start from the idea that it is perfectly appropriate to judge a person more harshly when he *should have known better*.

[5] Recognition of this tension goes back at least to Plato. In *The Laws*, Plato registers that attempted murder really should be treated the same as murder, but he can't bring himself to follow through on this. He says that while murder should carry capital punishment, attempted murder should not. His justification for differential treatment is that we should refrain from executing attempted murderers "as a thank-offering to the deity, and in order not to oppose his will." (Plato 1873: 390).

[6] Nagel (1979) identifies several different kinds of moral luck. Our interest in this paper is restricted to what he calls "resultant luck."

If Ken acted in an *epistemically irresponsible* way, then Ken deserves more blame than if he had acted under better epistemic conditions. Consider if Ken and Jan each serve their guests mushrooms from their garden. Jan has carefully researched the mushrooms and as a result, she comes to think that the mushrooms are safe. Ken, on the other hand, hasn't researched the mushrooms at all but has seen similarly colored mushrooms in the grocery store, so he concludes that these mushrooms are safe. Regardless of whether the mushrooms are safe, Ken's action of serving the mushrooms is obviously more blameworthy than Jan's. Ken should have done more to be sure that the mushrooms were safe. Ken had a terrible reason for thinking the mushrooms were safe.

It is taken to be normatively appropriate to blame someone more when she or he acted under epistemically bad reasons. This is the basis for two quite different epistemic accounts of luck-based blame judgments.

The rational inference account: One venerable epistemic explanation of outcome-based blame attributions is that outcome provides important information for drawing inferences about the epistemic status of the agent (e.g., Richards 1986; Rosebury 1995; Thomson 1993; for a classic statement of this general position, see Heider 1958). The world is a noisy place, and when we are trying to allocate blame, we need to draw on a wide variety of information to determine the culpability of an agent. One source of evidence is outcome. If Fred gets salmonellosis after eating Tony's chicken, then that outcome, Fred's food poisoning, is perceived as *evidence* that Tony was insufficiently attentive to food preparation. So, we feel justified in blaming Tony more than we would have if Fred did not contract salmonellosis. This explanation of outcome-based blame will be dubbed the *rational inference account,* since it takes our harsher judgments of blame to be based on rational inferences about the epistemic status of the agent. This model is represented in Fig. 9.1.

The epistemic bias account: The *epistemic bias* account of outcome-based blame also starts with the observation that it is perfectly legitimate to assign more blame to someone who acts on the basis of bad reasons. If a person performs an action he *should have known* was very risky, then we are right to assign more blame. In addition, the epistemic bias account holds, along with the rational inference account, that outcomes affect our epistemic assessments of agents. However, unlike the rational inference account, the epistemic bias account proposes that when we learn of a bad outcome, we fall prey to a kind of egocentric bias. Knowing that the outcome was bad, we view the agent's prior epistemic state through the lens of hindsight. As Ed Royzman and Rahul Kumar put it, "we are commonly mistaken in our judgment of what was reasonably foreseeable" (Royzman and Kumar 2004: 338). This is often because we overestimate the extent to which others know what we know. Royzman & Kumar call this egocentric tendency the *I know, you know* bias (following Royzman et al. 2003). There is a wealth of independent evidence that we are indeed subject to an egocentric bias in judging others' epistemic situations.[7]

[7] It is not obvious that egocentric attributions are all subserved by a single mechanism. Indeed, there might be several different pathways that generate egocentric biases. But this doesn't affect our current concerns, and so we will set aside this complication.

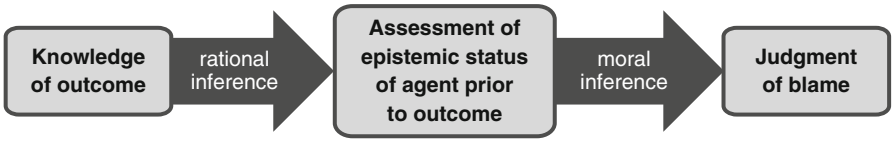

Fig. 9.1 The rational inference account of outcome-based blame judgments

Hindsight bias provides the most familiar illustrations. Hindsight bias is richly documented (see, e.g. Hawkins and Hastie 1990), but one study will suffice for present purposes. Shortly before Nixon's 1972 trip to China, participants were asked to estimate the probabilities of various events occurring (e.g., that Nixon meets Chairman Mao). Then after Nixon's trip, the same participants were asked to recall their predictions. Participants' memories erred on the high side when the event actually happened, and their memories erred on the low side when the event didn't occur, indicating that their belief about what actually happened affects their memory about what they *thought* would happen (Fischhoff and Beyth 1975).

This kind of epistemic egocentrism is a *general* cognitive bias (Royzman and Kumar 342; see also review by Hawkins and Hastie 1990). It is not specific to the domain of morality, let alone the phenomenon of outcome-based blame. However, it provides a natural explanation of outcome-based blame. According to the epistemic bias account, the process goes roughly as follows. Knowledge of the outcome triggers an epistemically egocentric evaluation of what the agent was in a position to foresee prior to acting. When I learn that Fred got salmonellosis, my knowledge of that outcome biases me to think that the outcome was largely foreseeable. Now that I know what happened, I conclude, via an egocentric bias, that Tony really *should have* known the chicken required more careful preparation. From there the process is perfectly legitimate. If my perception is that Tony was epistemically irresponsible, then I feel justified to draw a presumably valid moral inference that he is more blameworthy. This model is represented in Fig. 9.2.

As is clear from the figures, the rational inference model and the epistemic bias model are very similar. The only difference is in the nature of the process that delivers the epistemic assessment. This difference is far from trivial. The rational inference theory proposes that outcome-based blame judgments are, in fact, produced by a rationally respectable process. By contrast, the epistemic bias account proposes that such judgments are the result of a rationally defective process. Critically, however, both accounts maintain that the effects of outcome on blame judgments are *mediated by epistemic assessment.* There is not a more direct link between outcome and blame on these models. As a result, both accounts purport to explain outcome-based blame without adverting to an entrenched commitment to outcome-based blaming. For our purposes, this is a crucial commonality. For our aim in this paper is to determine whether blaming a person on the basis of outcome is an entrenched moral commitment. On both of epistemic accounts of luck-based judgments, moral luck is *not* entrenched.

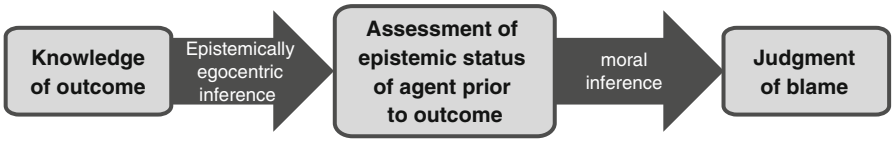

Fig. 9.2 The epistemic bias account of outcome-based blame judgments

We now turn to examine whether these epistemic models adequately account for the phenomena. The psychological evidence on moral luck currently presents a rather mixed picture. But we will argue that there is reason to favor the idea that outcome has an effect on blame judgments that is not mediated by epistemic inference.

9.3 Psychological Research on Moral Luck

Although moral luck has been hotly debated in philosophy for decades, the first extended exploration of outcome-based blame in psychology only appeared in 2008 (Cushman 2008; but see Walster 1966; Baron and Hershey 1988). In an important series of studies, Fiery Cushman finds that outcomes do indeed impact judgments of blame. It will be important to explain the studies in some detail.

9.3.1 Cushman's Experiments

In Cushman's studies, participants are presented with several scenarios, with the intentions and outcomes of actions systematically varied.[8] All participants are given the same initial scenario. For instance: "Jenny is taking a class in sculpture. She is assigned to work with a partner to weld together pieces of metal." The scenario then develops in different ways for different conditions. For example, for some subjects, the story is elaborated in a way that makes clear that Jenny has a *bad intention*:

> Jenny wants to burn her partner's hand. Jenny thinks that if she welds a piece of metal that her partner is holding the heat will travel down the metal and burn her partner's hand.

For other subjects, the story is elaborated with Jenny having a *neutral intention*:

> Jenny does not want to burn her partner's hand. Jenny only wants to weld together the metal. Jenny does not think that if she welds a piece of metal that her partner is holding the heat will travel down the metal and burn her partner's hand. Jenny thinks that the metal will weld without causing her partner any injury at all.

[8] The category of "intention" here actually collapses two different factors in Cushman's studies—belief and desire. This additional complication is not relevant to our current interests, so we opt for a more streamlined presentation.

In addition to varying the quality of the agent's intentions, Cushman also varies the quality of the outcome. For some subjects, the story ends with a *bad outcome*:

Jenny welds the metal, and her partner's hand is burned.

For other subjects, the story ends with a *neutral outcome:*

Jenny welds the metal, but her partner happens to let go and is not burned at all.

All of these factors are crossed. So some subjects get bad intention + bad outcome; others get bad intention + neutral outcome; others get neutral intention + bad outcome; and others get neutral intention + neutral outcome.

After reading the scenario, participants are asked to indicate on a 7 point scale, "How much blame does Jenny deserve?" Not surprisingly, Cushman found that intention played a major role in the assignment of blame. When Jenny had a bad intention, she was blamed more than when her intention was neutral. More importantly for us, however, is that *outcome* also played a major role in the assignment of blame. The results showed a stepwise progression. Neutral intention + neutral outcome was rated very low; neutral intention + bad outcome was rated significantly higher; bad intention + neutral outcome was rated higher still, and bad intention + bad outcome approached the top of the scale. In a second study, Cushman asked about punishment rather than blame. Once again, outcome had a major impact on subjects' judgments about how much a person should be punished.

In addition to questions about blame and punishment, Cushman also had subjects answer a question about *wrongness*: "How wrong was Jenny's behavior?" Once again, intention was a major factor in determining subjects' judgments of wrongness. However, outcome had almost no impact on these judgments. While outcome does affect judgments of blame, it seems to have little effect on judgments of wrongness.

9.3.2 A Psychological Model

Based on his results, Cushman offers a "two-process" model of moral judgment. Judgments of wrongness, on this model, are driven by an evaluation of the agent's intention. But they are not driven by the outcome. Judgments of blame, on the other hand, are driven by both the outcome and an evaluation of intention (see Fig. 9.3).

In one sense, Cushman's model is too narrow for our purposes. For his 'analysis' stage only countenances the quality of the agent's intention and not the quality of her *reasons*. Even if an agent didn't *intend* to harm, our judgments of blame are likely sensitive to whether the agent acted on the basis of good or bad reasons. For instance, when we judge someone as blameworthy because they are negligent, presumably we are often trading on the poor quality of the agent's reasons. For present purposes it will be essential to include *quality of reason* as a factor in the model.

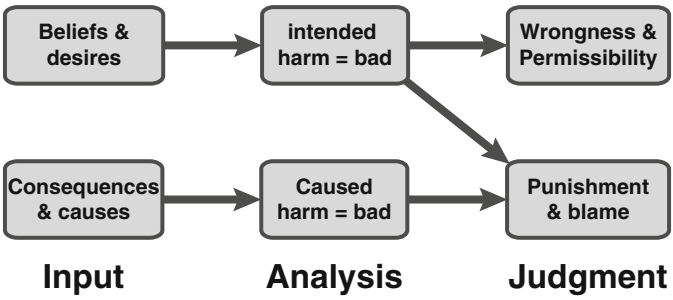

Fig. 9.3 Cushman's two-process model of moral judgment (2008: 364)

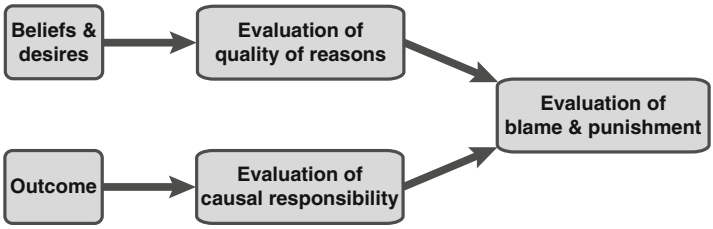

Fig. 9.4 Core blame model of outcome-based blame judgments

To accommodate this, and to focus just on blame judgments, we can recast Cushman's model slightly, as depicted in Fig. 9.4. We'll call this the "core blame" model.[9]

The critical feature of the core blame model is that outcome has an effect on blame judgments that is *not* mediated through epistemic judgments. Epistemic assessments do, on this model, affect judgments of blame. But outcome is an independent pathway that also affects blame judgments. By contrast, the epistemic models (Sect. 9.2.2) make no appeal to such a pathway. To emphasize the difference between the core blame model and the epistemic models, it is useful to frame the epistemic models in a general fashion, while remaining neutral on whether the epistemic evaluation is itself rational. Broadly speaking, the epistemic models hold that outcome-based blaming is always mediated by epistemic evaluation, as reflected in Fig. 9.5.

The core blame model has some apparent advantages over the epistemic model. First, the fact that Cushman found that outcome strongly affects judgments of blame but not wrongness, is not easily explained by the epistemic accounts. For if a bad outcome triggers a negative evaluation of the agent's epistemic state, then one might expect this to impact both blame and wrongness judgments. That is, if I come to think that Jenny *should have known better*, then it seems like this should lead to

[9] We take this to be a friendly amendment to Cushman's model. Thus, the "core blame" model that we discuss in the text is, we take it, effectively Cushman's model.

Fig. 9.5 Epistemic model of outcome-based blame

inflated judgments of both wrongness and blameworthiness. But while Cushman finds outcome to have a major effect on blame judgments, outcome does not have much of an effect on wrongness judgments.[10]

The second advantage of the core blame model is that it fits with broader developmental findings on the role of outcome in early judgments of punishment. In his classic work on children's views of punishment, Piaget presented children with the following two stories:

> A little boy who is called John is in his room. He is called to dinner. He goes into the dining room. But behind the door there was a chair, and on the chair there was a tray with fifteen cups on it. John couldn't have known that there was all this behind the door. He goes in, the door knocks against the tray, bang go the fifteen cups and they all get broken!
>
> Once there was a little boy whose name was Henry. One day when his mother was out he tried to get some jam out of the cupboard. He climbed up on to a chair and stretched out his arm. But the jam was too high up and he couldn't reach it and have any. But while he was trying to get it he knocked over a cup. The cup fell down and broke (Piaget 1932/1997: 122).

Following the cases, Piaget asked the children how much punishment each child should get. He found that children allocated punishment according to outcome rather than intention. For instance, on being asked how much punishment each should get, one child said: "…the one who broke the fifteen cups: two slaps. The other one, one slap" (Piaget 1932/1997: 125).

According to Cushman, the outcome-based system that we see in the young child persists into adulthood. As the child matures, she gains facility with an additional pathway, the intention pathway. But the early-emerging outcome-system does not get replaced. Rather, on the core blame model, what we see in adults' outcome-based blame judgments is just the response of this early-emerging mechanism. In this sense, the core blame hypothesis is naturally allied with the core knowledge program in developmental psychology (Spelke 2000): certain systems emerge early in development and are preserved through adulthood.

[10] While this poses something of a problem for the epistemic accounts, it is not yet a crushing objection. For one might maintain that judgments about the quality of a person's reasons are in fact more important to blame judgments than wrongness judgments. This is, of course, an empirical question.

9.3.3 Evidence Consonant with Epistemic Accounts

Although the foregoing observations provide some prima facie reasons in favor of the core blame hypothesis, they do not provide a very direct comparison of the core blame hypothesis with the epistemic account. Moreover, Cushman's results on blame judgments can be easily accommodated by the epistemic accounts. The psychological process might unfold as follows: The bad outcome triggers (whether by rational inference or egocentric bias) a negative evaluation of the agent's *reason,* and this then leads to an increased allocation of blame. When Jenny's partner's hand is burned, that might lead people to think that Jenny *should have known* the risk*,* and that evaluation might then lead to the increased blame judgment. Since Cushman's studies don't include *quality of reason* as a variable, it is impossible to directly rule out this explanation of the result.

Liane Young and colleagues recently set out to address this limitation in Cushman's studies (Young et al. 2010c). In addition to varying whether the outcome of their scenarios was good or bad, they also varied the quality of reasons that the agent had. Participants were given several different scenarios, with the *reason* and the *outcome* systematically varied between subjects. A representative scenario went as follows: "Mitch is at home on his day off, giving his 2-year-old son a bath. He fills the bath while his son stands by the tub. The phone rings in the next room. Mitch tells his son to hang on while he gets the phone." As in Cushman's study, the scenario then develops in different ways for different conditions. For some subjects, the story is elaborated in a way that makes clear that Mitch has a *bad reason*: "Mitch's son never does what he's told. But Mitch believes his son will wait for him for just a moment. Mitch leaves the room for 2 min." For other subjects, Mitch had a *good reason*: "Mitch's son always does what he's told. So, Mitch believes his son will wait for him for just a moment. Mitch leaves the room for 2 min." In addition, Young et al. varied the ending of the story. Some subjects received the *bad outcome*: **"**When Mitch returns, his son is in the tub, face down in the water." Others were given the *neutral outcome*: "When Mitch returns, his son is where he left him, outside the tub. He enjoys his bath."[11]

After reading the scenario, some participants were asked a question about the quality of reason: "Did Mitch have good reason to believe that his son would wait by the tub?" Other participants were asked the blame question: "How morally blameworthy is Mitch for leaving his son alone by the tub?" As expected, people attributed more blame when the agent had a bad reason. That is, when Mitch in fact had a bad reason ("Mitch's son never does what he's told…"), he is judged more blameworthy than when he had a good reason ("Mitch's son always does what he's

[11] Young et al. also included an "extra lucky" condition in which the outcome was neutral even though the agent had a false belief about what would happen. For example, in the Mitch case, when he returned to the bathroom, his son was in the tub but he was fine. This extra wrinkle is not essential to our interests here, so we will focus on the simpler contrast: true belief + neutral outcome vs. false belief + bad outcome. (We note that this is the familiar contrast from moral luck cases in the philosophical literature.)

told…"). Like Cushman, Young et al. found people attributed more blame when the outcome was bad. When Mitch's son is face down in the tub, people judge Mitch to be more blameworthy. This all looks fine for the core blame proposal.

When we turn to people's judgments about agents' *reasons*, the data begin to look less friendly to the core blame model. Although people attributed more blame when the outcome is bad, it is also the case that people judged the agent's *reason* to be worse when the outcome was bad. So, if Mitch's son is face down in the tub, people tended to give worse assessments on the question "Did Mitch have good reason to believe that his son would wait by the tub?" The worry for the core blame hypothesis should be clear. One plausible explanation of people's responses is that the bad outcome leads people to think Mitch was being epistemically irresponsible, and that is why they judge him more blameworthy when there is a bad outcome.

In addition to the verbal responses, Young and colleagues also measured activity in the right temporal parietal junction, a brain region associated with theory of mind (Saxe and Kanwisher 2003). The results indicated significantly greater activation when the outcome was bad as compared to when the outcome was neutral. The natural interpretation of these results is, again, friendly to the epistemic accounts. When the outcome is bad (as when Mitch's son is face down in the tub), this leads participants to reevaluate the mental states of the agent.

The results from Young and colleagues suggest that outcomes do affect epistemic evaluations. And this provides some support for the epistemic accounts of moral luck. That is, the results suggest that the epistemic account (as represented in Fig. 9.5) does likely explain a significant chunk of the phenomenon of outcome-based blaming.

9.3.4 Evidence for the Core Blame Model

Although Young and colleagues' results appear to favor the epistemic models, more detailed statistical analyses do provide support for the core blame model. Using mediation analyses,[12] they found, not surprisingly, that when the agent had a bad reason, this affected judgments of blame via judgments of justification. When Mitch had a bad reason, people judged him more blameworthy *because* they registered that he had a bad reason. More important for our purposes is the effect of outcome. Bad outcomes led both to harsher epistemic assessments and harsher blame judgments. Strikingly, mediation analysis did not support the hypothesis that outcome leads to harsher blame by way of harsher epistemic assessments.[13] Rather, the opposite was found: outcome leads to harsher blame, which then leads to harsher epistemic

[12] Mediation analysis is a statistical technique for inferring causal relations among multiple variables. For details of the mediation analysis discussed here, see Young et al. (2010c, 341–343).

[13] That is, there was no significant effect indicating that outcome influences blame via altering judgments of justification. This does not, of course, mean that outcome does *not* affect judgments of justification. Rather, it just means that in this study, no significant difference emerged.

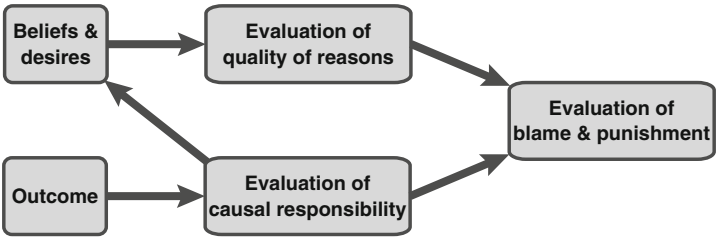

Fig. 9.6 A fuller account of outcome-based blaming

assessment. So, when Mitch's son was face down in the water this leads subjects to blame Mitch more, and their blaming him more leads them to evaluate his reasons more harshly. The important conclusion for our purposes is Young and colleagues' suggestion that there is an effect of outcome on moral judgments that is not mediated by epistemic evaluation. The results of the mediation analysis thus provide support for the core blame model.

Thus, it seems that the core blame model, as represented in Fig. 9.4, captures an important part of the phenomenon of outcome-based blaming. Of course, Fig. 9.4 is still quite incomplete as an account of the total functional profile of blame judgments. It remains plausible that epistemic considerations (whether through rational inference or epistemic bias) also play a role. A more complete account of the effect of outcome might incorporate both an epistemically mediated process and a process that is more direct (see Fig. 9.6). However, since our interests are limited to core blame, we will focus on that more restricted model.

9.4 Outcome-Based Blame: An Emotional Bias?

The evidence currently available suggests that our commitment to outcome-based blame is an entrenched commitment, which is in alignment with extant attribution theory work. We seem to have a commitment to moral luck that cannot be explained away as mediated by epistemic inference. This is an important step in showing that the commitment to outcome-based blame is entrenched, since part of what it is for a moral commitment to be entrenched is that the commitment is not inferred from other beliefs. The other element of entrenchment is that the commitment be grounded in human emotion. Although there is currently little evidence on the matter, it is plausible that anger plays a critical role in outcome-based blame judgments. Let us take for granted, then, that people do have an entrenched commitment to outcome-based blame. The ethical conservatism with which we began entails that people's entrenched commitment to moral luck is not undermined even if moral luck lacks positive justification. But of course entrenched commitments are not unassailable. There might be powerful considerations that should lead us to give up our commitment to moral luck. One important challenge is that the commitment to moral luck might be dismissed as the product of a distorting bias. Emotions sometimes exert what is obviously a distorting bias on judgments of blame and of the basic perception

of events upon which the judgments are based. For instance, showing people an emotional movie leads them to attribute more blame when evaluating an agent in an unrelated scenario (e.g. Lerner et al. 1998). Of course, the mere fact that emotion is implicated in judgment does not provide grounds for dismissing the judgment as biased. In many contexts, (e.g., judgments of taste and beauty) emotions contribute to judgments without thereby distorting judgment.

It is a difficult matter to assess whether a particular class of judgments is the product of a distorting bias. One strategy for exposing bias is to see whether people withdraw their judgments under full information. Our idea, borne of philosophical temperament, is to see whether people will quickly renounce moral luck under conditions of full information. If people do quickly renounce moral luck, then this is some evidence that their luck-based judgments result from a bias. If, on the other hand, people persist in making luck-based judgments under full information, this is some reason to doubt that moral luck can be readily dismissed as a bias. In designing our studies, we took a cue from the findings of Young and colleagues. In their study, they found that the impact of outcome on blame judgments was strongest when the specified reason was weak. That is, outcome had a greater impact on blame when agent was clearly negligent (Young et al. 2010c: 339). As a result, we thought that under full information, people might be more willing to embrace the outcome-based blaming when the agent is clearly negligent. For our experiment, we contrasted two cases. In one case, the agent had manifestly good reason for her action, while in the other case, everything was the same except that the agent was clearly negligent. All participants received the same initial set up: "Susan is taking a class in sculpture. She is using a special saw to cut a large sheet of metal in half. Susan is about to cut the sheet of metal on the table in half." Then, for one group of subjects, Susan is ascribed what is clearly a *good reason* for her action: "Susan follows the safety instructions exactly and carefully secures the metal to the table. On this basis, Susan believes that when she cuts the sheet of metal, the sawed off half will remain on the table, and not fall off the table or hurt anyone." For the other group of participants, Susan is clearly *negligent*: "Susan does not consider if when she cuts the metal, the sawed off half will remain on the table, or whether it might fall off the table and hurt someone." All participants are then presented with an explicit question about outcome-based blame. They are told that Susan cuts the metal in half, and are then told to consider two different endings to the scenario:

> In case A, after Susan cuts the metal, the metal stays in place and no one is hurt.
>
> In case B, after Susan cuts the metal, half of the sheet falls off the table and onto another student's foot, breaking several bones.
>
> Susan acts the same way in both cases, but in case A it turns out no one is hurt, and in case B it turns out a student's foot is badly injured. What Susan does and thinks in the two cases is exactly the same.

After this explicit presentation of luck-based outcomes, all participants are asked two questions:

> Given that Susan acted in the same way in both cases, does Susan deserve more blame for cutting the metal in case B than in case A?

> Given that Susan acted in the same way in both cases, does Susan deserve more punishment for cutting the metal in case B than in case A?

In the *good reason* condition, participants tended to give low rankings to both claims, indicating that when a person acts based on a good reason, they do not deserve more blame or punishment when the outcome is bad. By contrast, participants gave significantly higher scores in the *negligence* condition, indicating when a person is negligent, they *do* deserve more blame and punishment when the outcome is bad.[14]

Thus, under full information, people embrace their outcome-based judgments much more when the agent is obviously negligent. Indeed, they largely reject outcome-based blame when the agent had a good reason. This, of course, fits nicely with Young & colleagues' finding that outcome had the strongest effect on blame when the agent was negligent. The fact that people embraced outcome-based blaming under the negligence condition, but not under the good reason condition, suggests that we cannot simply dismiss the responses as the product of a distorting general emotional bias. For people's judgments of outcome-based blame is *sensitive* to whether or not the agent had good reasons.

9.5 Luck, Control, and Negligence

In the previous section, we considered one challenge to the normative authority of our entrenched commitment to moral luck—that luck-based judgments are the product of a distorting bias. Strikingly, we found that under full information people will explicitly endorse blaming people more when they unluckily produce a bad outcome, *but only* when agents have been negligent. This explicit endorsement of luck-based blame for negligent agents suggests that the judgments cannot be lightly dismissed as a distorting bias. There is a much more familiar philosophical objection to luck-based judgment, however—the control principle. For the bulk of the paper, we set the control principle aside for purposes of getting clear about the commitment to moral luck itself. But we now need to return to the control principle. We think that an important consequence of the empirical work is that it naturally leads to a proposal that can reconcile the control

[14] Judgments that Susan deserved more blame in the bad outcome case were significantly higher in the *negligent* condition ($M = 4.18$ out of 7) than in the *good reason* condition ($M = 3.06$) ($t(84) = 2.2$, $p < .05$). Similarly, judgments that she deserved more punishment in the bad outcome case were significantly higher in the *negligent* condition ($M = 4.42$) than in the *good reason* condition ($M = 3.00$), ($t(83) = 2.91$, $p < .01$). Judith Thomson ventures a different opinion. She considers the case of two negligent drivers, one of whom, Bert, causes a death. The other, Carol, is equally negligent but causes no harm. Thomson writes, "Well, *do* we regard Bert with an indignation that would be out of place in respect to Carol? Even after we have been told about how bad luck figured in his history and good luck in hers? I do not find it in myself to do so" (1993: 205). Our experiments suggest that many people would demur from Thomson's verdict.

principle with moral luck. To begin, we must look more carefully at how the principle ought to be formulated.

The control principle is typically framed in a global fashion. Recall Nagel's statement of the principle: "people cannot be morally assessed for … what is due to factors beyond their control" (Nagel 1979: 25).[15] This statement of the control principle does not restrict application to *non-negligent* agents. Thus, it seems to apply to all agents, regardless of their degree of conscientiousness or negligence. And the basis for formulating the control principle in this way is, at least often, taken to be a generalization from intuitions about cases. Here's Nagel again: "the condition of control does not suggest itself merely as a generalization from certain clear cases. It seems *correct* in the further cases to which it is extended beyond the original set" (26). So, our reactions in a range of cases are supposed to lead to a formulation of the control principle.

We have no objection to the idea that we can formulate ethical principles by generalizing from cases. Indeed, we are inclined to embrace some form of the control principle precisely because it handles clear cases well. However, if the principle is built up from cases, then we need to take seriously the finding that people have different views about cases in which agents are negligent as compared to cases in which agents are not negligent. When agents are fully conscientious, then people do indeed seem to abide by the control principle. They reject the idea that a fully conscientious but unlucky agent deserves more blame than a correspondingly conscientious lucky agent. But the situation is different for negligent agents. Many people accept the idea that an unlucky *negligent* agent deserves more blame than a lucky negligent agent. This suggests that, while the cases support some version of the control principle, they don't support a global version of the principle. Rather, if we are to construct a control principle from observed lay intuitions about cases, a more accurate rendering might be:

> When agents act *on the basis of good reasons,* reasons that rule out negligence, then it is inappropriate to blame or punish them for anything that is beyond their control.

This version of the control principle does indeed seem plausible. It also allows for the possibility of moral luck when agents are negligent.

Thus, we think that a control principle that really answers to the cases might well be consistent with moral luck. For such a control principle will be narrower than the global renditions that are typical in the literature. This more nuanced version of the control principle is not in tension with conserving luck-based judgments in cases of negligence. Collectively, the folk view would be that, while lucky outcomes ought not affect our blame judgments of agents who act on the basis of good reasons, those who act on the basis of bad reasons should be held accountable for the results of their actions, even if the results are partly a matter of luck. Not only does this dampen the apparent tension in folk morality, it also appears to be a defensible view, normatively speaking.

[15] See also Royzman and Kumar's characterization, quoted above (Sect. 9.2.1).

9.6 Conclusion

This paper has offered an exploration of the intersection of moral psychology and normative ethics. The psychological evidence indicates that people give harsher blame judgments to unlucky agents than to equivalently situated lucky agents. The process that gives rise to these harsher judgments seems not simply to be mediated by inferences about the epistemic status of the unlucky agent. Rather, outcome seems to have a more direct effect on blame judgments. This suggests that our commitment to allotting greater blame to unlucky agents is an entrenched commitment that runs fairly deep in human psychology. Such commitments, we have maintained, carry a normative authority that is not undercut by the mere fact that the commitments are based on a-rational and a-reliable processes. Thus, we take luck-based judgments to carry some initial normative authority. This initial authority is not beyond critique, but as it happens, people's commitment to outcome-based blame is more sensitive than has been recognized. People are much more likely to embrace outcome-based blame when agents are negligent than when agents are conscientious. This, we have argued, provides the basis for a more plausible rendering of the control principle. Thus, the psychological work not only helps us assess the nature of our normative commitments, it also helps to articulate normatively plausible principles.

Acknowledgements Thanks to Jan Gertken, Michael Gill, Jesse Prinz, and Liane Young for comments and discussion on an earlier version of this paper.

Part IV
Justifications Between Rational Reflections and Intuitions

Chapter 10
Intuitions in Moral Reasoning – Normative Empirical Reflective Equilibrium as a Model for Substantial Justification of Moral Claims

Ghislaine J.M.W. van Thiel and Johannes J.M. van Delden

10.1 Introduction

Moral questions and dilemmas in everyday life prompt us to take a normative stance. Sometimes we rely on our moral intuitions and make judgments accordingly. In other cases, we feel the need for more extensive deliberation of a moral case. After a judgment is made, we have to ask: Can we justify our moral view to others who may have come to a different conclusion? Ethicists have long tried to describe fundamental moral principles from which justified judgments can be derived. However, until now, a set of foundations that received general assent has not been found. Moreover, the development of action guiding principles and rules always requires some kind of interpretation or specification of general principles (Richardson 2000). Currently, most ethicists hold the view that theory and practice should mutually influence each other in the process of searching for reliable moral judgments and theories.

A main theory that is put forward to seek justifiable resolutions is Reflective Equilibrium (RE) (Rawls 1971/1999: xiii). In a nutshell, RE is a coherentist model for moral justification in which the key idea is that we "test" various parts of our system of beliefs (including considered moral judgments, principles, relevant facts and background theories) against the other beliefs we hold. In this chapter, we aim to modify the model of RE in such a way that the moral experience of agents other than the thinker can play a role. We present our version of RE, called Normative Empirical Reflective Equilibrium, or NE-RE. NE-RE differs from RE in two respects: (i) moral intuitions of agents other than the thinker are included and (ii) empirical research is used to obtain information about these intuitions. Second, we acknowledge that NE-RE is susceptible to a major criticism, the so-called no-credibility objection. With reference to DePaul's (1993) work, we propose to enforce the quality of moral

G.J.M.W. van Thiel (✉) • J.J.M. van Delden
Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, The Netherlands
e-mail: g.j.m.w.vanthiel@umcutrecht.nl

M. Christen et al. (eds.), *Empirically Informed Ethics: Morality between Facts and Norms*, Library of Ethics and Applied Philosophy 32, DOI 10.1007/978-3-319-01369-5_10,

reasoning in NE-RE. We move away from the thought that only elements with high epistemic status should be allowed in the reasoning process of NE-RE. Instead, we arg*ue that in the good reasoning*-justified outcome strategy, beliefs can gain or lose justificatory power. Finally, the justificatory power of moral intuitions in a coherentist reasoning model such as NE-RE, is assessed.

10.2 Expanding the Scope: The Value of Moral Wisdom

In the current understanding of Reflective Equilibrium theory, there is no place for considered moral judgments of agents other than the person who performs the ethical reasoning (we call this person the thinker). However, a single thinker may come up with only a small part of the moral intuitions that may be relevant in a given case. Ethical reflection can benefit from the intuitions of agents other than the thinker because they add to the complexity of reasoning (Woods 1999). To achieve insight concerning the moral intuitions of others, a theorist can seek encounters with experience (Ives 2008). This means, for example, that the thinker collects data on the intuitions of several stakeholders in a given case. Some of these stakeholders may have intuitions that reflect a specific moral wisdom. This moral wisdom is present among experienced agents in a practice or situation, and is acquired through a learning process. In this learning process, formative experiences cause a person to adopt a different moral perspective. It can be provoked by a single confrontation with a work of art or literature. However, activities that lead to a formative experience usually influence a person's moral outlook over time. This implies that moral wisdom is in part dependent on, and can vary with, experience (DePaul 1993). It is moral wisdom in the Aristotelian sense, which refers to the ability to deliberate about human actions in terms of what contributes best to the good life (Edmondson and Pearce 2007).

Moral wisdom is thus acquired through moral experiences and it is reflected in moral intuitions. Wisdom is a key in dealing with problems that lack prescribed solutions and for which uncertainty and fluidity must be tolerated in seeking to resolve them. Wise responses to circumstances are often needed when moral judgments in health care have to be made, because of the profound effects health care decisions can have on people's lives (Edmondson and Pearce 2007).

Our perspective on the role of the moral views of moral agents in a practice is the background of our proposal to adjust the model of RE. We aim to modify the model in such a way that (i) moral intuitions of agents other than the thinker are included, and (ii) empirical research is used to obtain information about these intuitions. We call our version of RE the Normative Empirical Reflective Equilibrium (hereafter NE-RE) (van Delden and van Thiel 1998). The use of moral intuitions in normative reasoning is considered inescapable by some, but it is certainly not unproblematic (Daniels 1979; Beauchamp and Childress 2008). Moreover, mixing empirical with normative elements is at best regarded as risky business. We will address these issues from the perspective of NE-RE. But first, we outline the basics of reflective equilibrium theory.

10.3 Reflective Equilibrium Theory

Reflective Equilibrium was developed by John Rawls for the theoretical purpose of formulating the most appropriate conception of justice (Rawls 1971/1999: xiii). It is a model for justification that revolves around the idea that seeking justified moral principles requires an argumentative process in which general principles and background theories are considered together with a person's considered moral judgments. The term RE refers to both the process and the result of moral reasoning. A RE as the result of moral reasoning is a coherent and interconnected set of moral and non-moral beliefs at various reflective levels. The number of beliefs that can be included in the reasoning process of a researcher is necessarily limited, but ideally, in working toward a RE all relevant considerations are addressed.

In broad outline, moral reasoning according to RE proceeds in three stages. A person—the thinker—starts with identifying his or her considered moral judgments or moral intuitions. In theory, the starting point of reasoning can be any of the elements mentioned. But often moral reasoning is triggered by spontaneous and personal interpretation of the facts of a case. The next step is to formulate moral principles that are relevant for the situation under consideration. In theory, these can be new principles, but it is likely that a thinker will come up with at least some of our commonsensical moral principles (i.e., keep promises, respect autonomous choices; Arras 2007). These elements may be conflicting, inconsistent, or both. In that case, the thinker will have to respond to the divergence between the principles and his initial beliefs. He has to work back and forth between principles and judgments and make adjustments in both his considered moral judgments and moral principles. This process ends when the thinker accepts a set of principles that coheres with his considered moral judgments. The person's beliefs are now said to be in reflective equilibrium.

Rawls' idea was both welcomed and criticized. One major criticism was—and still is—the alleged subjectivism inherent to the method. Since all the beliefs taken up in RE reasoning are the beliefs of one single thinker, RE would amount to no more than a neat systematization of the preliminary ideas a thinker has about the case. In his influential article, Daniels (1979) acknowledges the problem of circularity and proposes to speak of narrow reflective equilibrium when considered moral judgments and principles are made coherent. To diminish the risk of subjectivism and circularity, Daniels proposed to add an extra round of reasoning to the method, in which the thinker attempts to disrupt the state of narrow reflective equilibrium by considering background theories and alternatives to his moral theory (Daniels 1979). Examples of background theories are a theory of personhood or a general social theory. These background theories have to be chosen for their potential to provide critical input. Again, mutual adjustment of the elements is required until the final result, a wide reflective equilibrium, is achieved.

Contrary to other approaches in ethics, i.e. principlism and casuistry, RE claims no locus of authority in one of the elements. In the process of consideration and amendment by the individual thinker, none of the elements has a privileged status

and all are open to revision. It is the thinkers' task to fit the most comprehensive and interconnected set of beliefs into a balanced view of the moral case at hand. When the thinker decides the adjusted moral beliefs form a coherent whole, the reasoning process ends. Rawls developed RE for the purpose of formulating a general theory of justice. However, the result of RE can also be a moral judgment about a case, or a so-called modest theory, which is more limited in scope (Van Willigenburg 1991: 191).

There are several possible interpretations of RE (Rawls 1971/1999: 43). Many authors besides Daniels (1979) have suggested changes to the type and number of considerations that can be included. By means of illustration, we give some examples. Nielsen suggested only letting judgments that are actually agreed upon within a community take the place of considered moral judgments (Nielsen 1982a). Beauchamp and Childress proposed to extract considered moral judgments from the common morality and thus include those judgments that all serious moral persons share (2008: 387). Heeger and Van Willigenburg added morally relevant facts as a separate element to be included in the reasoning towards RE (Van der Burg and Van Willigenburg 1998: 14). Van der Burg (1997) added ideals.

RE is the reference theory from which we developed Normative Empirical Reflective Equilibrium (NE-RE). In general, the strength of a moral view, achieved through (NE-)RE depends on three aspects:

(i) the comprehensiveness of the set of beliefs;
(ii) the strength of argumentation in the reasoning process;
(iii) the level of coherence among beliefs in the end.

In the following sections of this article, we describe how the use of empirical research in moral intuitions adds to the comprehensiveness of NE-RE. Subsequently, we address the strength of the reasoning process and the role of coherence in the light of an important objection against RE, associated with the use of moral intuitions in ethics.

10.4 Enriching the Thinker's Perspective: Moral Intuitions

In RE, the initial beliefs that are allowed in the reasoning process are called Considered Moral Judgments (hereafter CMJ). When Rawls talked about considered moral judgments, he meant the moral judgments that seem clearly to be correct under conditions that were suitable for making good judgments of the relevant kind (Scanlon 2003). The concept of CMJ has been criticized for the vagueness of the requirements that need to be met. It seems that not only should the thinker be unaffected by conditions that threaten his or her ability to exercise his or her sense of justice, but he or she must also have the will or the intention to reach the correct decision (Rawls 1971/1999: 42). For the practical use of RE, identifying a thinker or other persons CMJ would require that we rule out the judgments made under

non-ideal circumstances. Moreover, the thinker's desire to reach a correct decision has to be verified. These requirements are formulated with the aim to warrant a minimal level of credibility of CMJ, that is needed to justify their role in a coherentist model of justification like RE.

The concept of moral intuition is suggested as an alternative to CMJ (Van der Burg and Van Willigenburg 1998: 14). We prefer this concept over CMJ for several reasons. First, it is appropriate to characterize the beliefs that a person comes to hold without extensive deliberation. The appearance of an intuition is rather sudden and not well thought-out. Second, the concept of moral intuition offers an account of the type of belief we think is relevant to NE-RE. Moral intuitions are the preliminary interpretations of people that give the holder a sense of the direction in which a judgment about the case should go. These intuitions can be both pre-reflective and post-reflective. Pre-reflective interpretation occurs when a person is confronted with a moral situation with which he is unfamiliar. In other cases the interpretation of a person is based on structuring of facts in previous cases, and in this way is influenced by experienced perception. This is called post-reflective interpretation (Haidt 2001).

There is evidence for the claim that most of our moral judgments are intuitive and automatic responses to challenges, elicited without awareness of underlying mental processes (Musschenga 2008). It is thus highly likely that the moral judgments of people, who work and live in a certain practice, are usually at an intuitive level. Thus, it is through these intuitions that moral theorizing can gain access to detailed information on the moral experience of relevant agents. If we incorporate the intuitions of, for example doctors and nurses, we can work towards a NE-RE that grasps a moral experience that generally cannot be found among people outside the health care practice. An example of the value of collecting intuitions from experienced agents is provided by Ives and Draper (2009). They describe their so-called 'encounter with experience' in a project on paternal rights. They initially had come to the conclusion—based on their own intuitions—that there were *no good* grounds for basing paternal responsibilities or rights on genetic relatedness. However, data from men who were separated from their children and had no other connection to their child than a genetic one, made the researchers aware of the limitations of a non-genetic account of paternity (Ives and Draper 2009).

Moral intuitions are relevant to ethical judgments first, because they are usually the starting point of deliberation. Second, when a person comes to hold a moral intuition, he will generally feel the urge to look closer at the case and seek alternative interpretations of the circumstances (Van Willigenburg 1991: 111). Finally, moral intuitions connect ethical reflection through NE-RE to our everyday moral experiences. An important aim of the practical ethicist is to justify claims to other moral agents. This requires sharing reasons in a manner that makes considerations relevant to those claims as 'vivid and motivationally compelling' as possible (London 2000). Moral reasoning that avoids moral intuitions amounts to requiring that a person adopts moral principles without referring to the intuitions that guide her moral action in daily life (DePaul 1993: 2–3).

10.5 Moral Intuitions and Moral Justification

There seems to be one major reason to stay away from moral intuitions: it raises the question of how it can be done without endangering the normative force of ethical reasoning. This is a matter of moral justification. Suppose a person has considered a set of beliefs at different levels of reflection, tested and adjusted these beliefs and came up with a coherent moral view. We may conclude that this thinker rightly claims that he was successful in achieving an RE. But is his moral view convincing? On what grounds can this thinker defend his moral judgments toward others who may have come to another conclusion? To defend moral claims, we have to elaborate on the moral justification of judgments and theories. Critics have put forward the no-credibility objection, which holds that moral justification through RE is impossible precisely because moral intuitions are allowed in the reasoning process.

The no-credibility objection (DePaul 1993: 25) rests on the combination of two features of RE: (i) the role of considered moral judgments (or moral intuitions) and (ii) the coherentist nature of the method. The no-credibility argument entails the following: if justification of the result of an RE process is based on the coherence of a set of beliefs, each of the individual beliefs has to be reliable enough to guide the process of reasoning. Moral intuitions, it is argued, are subjective and can be erroneous. Therefore, they lack the credibility that is necessary to add to the justification of judgments in a coherentist model of moral reasoning. It is generally recognized that the no-credibility objection poses a serious problem for RE. However, proponents of RE have made things more complicated, we believe, by seeking moral judgments and moral intuitions that are considered more credible. In the next section, we explain why this is the wrong track, and we use DePaul's (1993) work to develop a different strategy.

10.6 Enhance Credibility or Embrace Openness to Revision?

Moral intuitions are personal interpretations of the facts of a case by individuals and thus marked subjective. This makes them—and (NE-)RE as a whole—vulnerable to the no-credibility-objection. To defeat the no-credibility objection, proponents of RE developed what we call the credible input-justified outcome strategy:

10.6.1 Enhance Credibility: The Credible Input-Justified Outcome Strategy

To achieve a set of credible moral intuitions, several authors suggest stringent selection of initial judgments at the start of reasoning, in order to prevent the 'bad' ones from entering the reasoning process (Swanton 1991; Nielsen 1982b; Singer 2005). For example, Beauchamp and Childress argue for the use of the common morality

(e.g., a set of norms shared by all persons committed to morality) as a reliable source of moral intuitions that can be allowed into the reasoning process (2008: 387). We will use Beauchamp and Childress' proposal as an example when we argue against the credible input-justified outcome strategy.

Unfortunately, the attractive idea of criteria that can tidy up our messy set of intuitions has serious drawbacks. In the case of Beauchamp and Childress' common morality, the first drawback is that selection of moral intuitions—and the subsequent exclusion of those that seem not sufficiently trustworthy—leads to excluding intuitions from (possibly relevant) minority groups of agents. The remaining set may provide only a small input in the reasoning process, because norms that are shared by all morally serious persons are either of a general nature (like principles) or very few in numbers (DeGrazia 2003). Moreover, limiting the set of moral intuitions in this way complicates the task of integrating the moral wisdom we argued for in previous sections. For example, agents may have moral intuitions that are not shared by all persons committed to morality, because their intuitions stem from moral experiences that are uncommon.

The second drawback is that a thinker who is convinced of the credibility of each of his intuitions (before testing them in the light of relevant principles, ideals and so forth) will be wary of major alterations (DePaul 1993:44). The consequence may be that intuitions from the common morality become privileged elements, in the sense that they are less prone to revision than other elements (principles, background theories and morally relevant facts). This runs counter to the non-foundational character of RE (Strong 2010). Beauchamp and Childress explicitly address the question of whether a change in the common morality could occur. They state that the theory of common morality remains open to the possibility that the common morality changes. At the same time, they endorse the view that the possibility of such a change seems to weaken the idea of a common morality (Beauchamp and Childress 2008: 390–391).

Finally, the demand for credible moral intuitions might be contrary to the dynamic character of the method of RE. A significant role of moral intuitions is to fuel the thinking process of the thinker. To enrich his view, the thinker should seek to broaden the set of moral intuitions throughout the whole process of reasoning. Selection at the start of reasoning can hamper this function of intuitions. These disadvantages of the credible input-justified outcome strategy are an invitation to explore another line of thought with the aim of defeating the no-credibility objection.

10.6.2 Strive for Openness to Revision: Good Reasoning-Justified Outcome Strategy

Another approach to the no-credibility objection is to move away from the discussion about the characteristics of moral intuitions and focus on the argumentative process. This strategy is employed by DePaul (1993:39). We start from his work and

try to develop it further by arguing for the good reasoning-justified outcome strategy for moral justification in RE.

Following the good reasoning-justified outcome strategy for moral justification, the thinker starts with identifying the broadest set of relevant moral intuitions. Relevant intuitions are closely connected or appropriate to the matter at hand. In our view it is essential that the moral views of others rather than the thinker himself are taken up to enrich the initial set of moral intuitions. Empirical work designed to obtain these intuitions is usually necessary. In the subsequent process of moral reasoning the moral intuitions, principles and theories can gain or lose justificatory power. RE provides a model in which they together are examined, adjusted, accepted or expelled. The guiding principle of examination is the level of coherence among different elements. The thinker will try to achieve coherence by mutual adjustment of beliefs. In the end, only the beliefs with sufficient justificatory power are part of RE. The moral intuitions in this RE can be considered to have sufficient credibility because they were tested and confirmed in the reasoning process towards RE.

10.6.3 Requirements for the Good Reasoning-Justified Outcome Strategy

In the good reasoning-justified outcome strategy, much importance is bestowed upon the argumentative process. To achieve an equilibrium that has strong justificatory power, the reasoning and the joint attitude of the thinker should meet several criteria.

Transparency: Transparency is an ideal characteristic of RE reasoning. It should be pursued with the aim of making the reasons for a decision accessible to a wider public and open for scrutiny. This accessibility requires that reasoning is not confined to the thoughts of the thinker. He or she will have to make clear and documented steps, arguing which facts and arguments are considered and how they are weighted in the reasoning process. For example Daniels refers to transparency as a key element of a procedure for fair priority setting, the Accountability for Reasonableness framework (Daniels 2000). Clarity about facts and arguments adds to the justificatory power of RE because the normative force of the outcome of RE depends in part on the strength of the reasons that have been prominent in the process (Holm 2008: 13). With regard to moral intuitions, transparency increases the chance that unfounded or ill-argued retaining or rejecting of moral intuitions is exposed. This decreases the risk of conservatism and avoids deliberate systematization of prejudices.

Openness of mind: The thinker should avoid getting ‘caught up’ in his own intuitions by taking on an attitude of openness. This requires first that he is aware of biases and motivated to correct for them (e.g. through employing debiasing strategies; Horton 2004). Second, the thinker should seek alternative ways to interpret the

moral aspects of a case. This may lead to the introduction of new moral intuitions or to the abandonment of others. In its most extreme form this results in a radical shift in moral views. DePaul named this a moral conversion: abandonment of a large part of—or even all—initial beliefs (DePaul 1993: 41). According to DePaul, the thinker should develop his abilities and faculties for making judgments by expanding his range of experiences. This is a valuable approach, but necessarily limited because gaining in-depth experience in a moral practice is a time-consuming endeavor. The thinker should therefore in our view obtain information about the moral experiences of relevant others, for example through empirical inquiry into their moral intuitions.

Reasonableness: The notion of reasonableness is prominent in the process of adjusting beliefs. For the purpose of an RE in which the moral intuitions of relevant agents are taken up, a reasonable thinker is sensitive to the perspective of all parties involved. Moreover, reasonableness requires that the person considering the reasonableness of a claim is aiming at agreement or at finding a course of action with which everyone will be satisfied (Scanlon 1998: 33).

The good reasoning-justified outcome strategy allows NE-RE to introduce moral intuitions into the reasoning process without being defeated by the no-credibility objection. RE's dependence on credibility of the elements at the start of reasoning is replaced by a process in which beliefs can gain or lose credibility. Thus, in NE-RE, we do not seek moral justification in separate elements. Instead, we accept low credence levels of beliefs at the beginning of the quest for a justified judgment or (modest) theory. In working towards NE-RE, we depend on good reasoning and coherence for moral justification of the result.

10.7 Moral Justification and Coherence

Rawls (1971/1999: 18) speaks of RE when the thinker formulates:

> [A] description of the initial situation that both expresses reasonable conditions and yields principles which match our considered moral judgments duly pruned and adjusted.

Thus, when the thinker decides that coherence is achieved, the reasoning process comes to an end. In RE in general, coherence is a key element. In NE-RE, the notion is even more important because the thinker abandons the ideal that the belief-system he or she started with consists of elements with sufficient credence levels. However, the nature of coherence and the way people should evaluate their beliefs with respect to coherence is poorly described (Beauchamp and Childress 2008: 387; DeGrazia 2003; Rauprich 2008). Intuitively, coherence is a matter of how well a body of beliefs hangs together: "[H]ow well its components produce an organized, tightly structured system of beliefs, rather than a helter-skelter collection or a set of conflicting subsystems." (BonJour 1985: 93)

Our purpose is to use NE-RE for moral justification of our judgments and theories. Therefore, a tangible concept of coherence is crucial.

10.7.1 Broad Coherence: Consistency, Comprehensiveness and Interconnectedness

Consistency is usually put forward as the first requirement for coherence. It is obvious that inconsistency is incompatible with coherence. However, on any reasonable coherentist account, coherence is more than mere logical consistency. Bonjour illustrates the need for additional requirements with two sets of propositions, A and B.

SET A	SET B
This chair is brown.	All ravens are black.
Electrons are negatively charged.	This bird is a raven.
Today is Thursday.	This bird is black.

Clearly both sets of propositions are free of contradiction. But in the case of A, this consistency results from the fact that its components are almost entirely irrelevant to each other; though not in conflict, they also fail to be positively related in any significant way (BonJour 1985: 96).

What is needed for meaningful coherence is substantial mutual support between the elements of a set of beliefs (Rauprich 2008). Sayre-McCord (1996: 166) argues that the two properties of connectedness and comprehensiveness add to the coherence of a set of beliefs that are consistent and thus minimally coherent.. However, he gives no indication of the relations between beliefs that add to connectedness and comprehensiveness. Nonetheless, we believe that these two properties are interesting starting points for a more substantial account of coherence in NE-RE. Comprehensiveness represents a guiding ideal of RE reasoning to consider as many relevant beliefs are reasonably possible. The notion of connectedness is an invitation to focus on the relations between beliefs. Each individual belief can be connected to others in ways that add to coherence. However, connections that diminish the coherence of the set are also possible.

10.7.2 Measuring: Four Types of Coherence

Investigation of the connections between elements is a way to measure coherence in a specific set of beliefs. Following Bonjour, we call the connections between beliefs that are relevant to coherence, inference relations (BonJour 1985:96). Thagard (1998) described four types of coherence:

– Explanatory coherence is the coherence between observation and understanding. The importance of this type of coherence lies in the fact that some normative principles are tied to empirical claims. Guarini (2007) gives the example of capital punishment. The argument for a general principle that capital punishment is acceptable may depend on the deterrent effect that it has. But whether capital punishment has a deterrent effect is a largely empirical question.

- Deductive coherence is the coherence between principles and judgments. There is a positive connection if a principle and a particular judgment are likely to be either both accepted or both rejected. Positive connections between principles and judgments are constituent parts of coherence.
- Deliberative coherence is the coherence between actions and goals. The positive deliberative connection is when an action facilitates a goal. Incompatibility of an action with a goal is a negative constraint between elements (Guarini 2007).
- Analogical coherence implies supporting a conclusion in one case by comparing it to a similar case whose moral status is more obvious. Analogical arguments are rarely convincing on their own, but they can contribute to the overall coherence of a view.

These types of coherence can guide a thinker in NE-RE when he comes to a point where he has to decide whether his set of beliefs can qualify as a reflective equilibrium. Connectedness refers to the relations among beliefs that can either be strong or loose, as in Bonjour's example of propositions. However, measuring the support for a belief does not inform the thinker about the level of coherence that is necessary for reflective equilibrium. There is no clear cut-off point for a RE (Strong 2010). Nonetheless, a thinker can evaluate the inference relations between his beliefs. Some will have many positive connections, and others will only have a few. For a reflective equilibrium, the beliefs that are situated at the heart of the system should be positively connected to each other. This idea may have a foundationalist ring, because it suggests that some beliefs are more important than others. We will address this issue in the next section.

Moreover, the requirement of comprehensiveness means that even though a small set of beliefs may be free of inconsistencies, the thinker has to make an effort to keep all beliefs he deems relevant aboard the reflective equilibrium as long as possible. Thus, the thinker should not readily accept a small but consistent set of beliefs. Instead, he has to have good reasons to dismiss beliefs. In a transparent reasoning process, the thinker can justify his choices and be criticized for it by others.

10.8 Coherentism and the Power of Moral Intuitions

Coherence is key to reflective equilibrium and RE is—unsurprisingly—generally characterized as a coherentist model. Earlier we pointed out that the no-credibility objection against (NE-)RE depends on the combination of the use of moral intuitions and the coherentist nature of the model. In this section we elaborate on coherentism with the aim to further clarify our view on moral justification in NE-RE.

10.8.1 Coherentism

Coherentism is contrary to the foundationalist approach, in which theorists search for a specific, fundamental norm or principle to justify moral claims. Coherentists give up the search for fundamental principles and instead claim that a particular

belief is justified for an agent if and only if it coheres well with the other things the agent believes (Radzik 2002). With regard to the coherentist nature of (NE-)RE two points have to be clarified. First, NE-RE is not a coherence theory of moral truth. Rather, NE-RE is a model for moral inquiry that leads a thinker to a justifiable moral judgment or (modest) theory. Contrary to epistemic interpretations[1] of RE, we hold that the justification of the views attained through RE is thus not based on the claim that they are true. The objective of justification is reflective testing of all relevant considerations in order to produce a coherent moral view that boosts our confidence that we are not mistaken.

Second, the result of a moral inquiry through NE-RE should not be seen as a fully stable equilibrium. It is a justified moral view for the time being. RE is a dynamic process that goes on as one's set of convictions changes. These changes can be provoked by new experiences and ongoing reflection. A thinker should continuously strive to increase the coherence of his beliefs in order to gain stability and justificatory power as things progress (Sayre-McCord 1996).

10.8.2 Two Versions of Coherentism

For a proper characterization of NE-RE, it may be helpful to distinguish between different versions of coherentism. A general distinction can be made between pure or uncompromising coherentism and coherentism with different levels of justificatory power.

Pure, uncompromising coherentism requires that the individual elements of a set of propositions are (i) independent of foundations, and (ii) derive their justificatory power only from the relationship with other elements of the set. Coherentism in this sense holds that a belief can only be justified by coherence considerations and that it is coherence alone that justifies (Ebertz 1993). It implies that no belief has a distinguished epistemic status and that no belief has a distinguished place within a coherent set (Haack 1993: 17–18).

Coherentism with different levels of justificatory power acknowledges the possibility of degree of justification. The degree of justification of a single belief can vary due to factors other than the inference relations between that belief and the other beliefs in the set. For example, some beliefs have a distinguished initial status, and this can confer a certain weight on a belief. The justificatory power of a set of convictions depends on the weighted mutual support. In addition, some beliefs are distinguished because they are more deeply embedded in a coherent set than others (Haack 1993: 17–18).

Ebertz specifically points to considered moral judgments as having a distinct status that is incompatible with pure coherentism. He refers to the accepted view that we draw upon a moral sense when we form our initial judgments (we call them moral intuitions) about the rightness or wrongness of actions. The justificatory

[1] Epistemic interpretation is thoroughly presented and discussed in Kappel (2006).

power of these moral intuitions is not derived from their relationship with other beliefs. Moral intuitions are rather sudden appearances of a judgment about the case at hand. And even though our first moral judgments can be replaced by others, their very nature implies that they not only enter into RE reasoning by virtue of cohering with other beliefs (Ebertz 1993). Ebertz goes on arguing that considered moral judgments play a foundational role in RE. His argument is that there is an element at the heart of the idea of justification through reflective equilibrium that is contrary to uncompromised coherentism: "There is a kind of test by which principles can always be tested—the test of whether they fit the CMJ we are committed to at that point in the reflective process" (Ebertz 1993: 204). He concludes that RE is a model not of coherentism but of modest foundationalism combined with the claim that coherence between beliefs is an additional necessary condition for justification.

Our view on the matter is that each element in RE, whether a principle, moral intuition or a general theory, is tested against the others. In this process, moral intuitions have no special status, and they can be adjusted or eliminated altogether. Thus, once in the process, moral intuitions play no foundational role. Our openness to the moral wisdom of, for example, health care professionals, should not be mistaken for a shift in locus of authority in RE. The locus of authority determines which element of a model or theory is decisive when there is inconsistency or conflict in a set of beliefs. The authority in top-down methods for example, is located in theory: when a principle conflicts with a moral intuition, the principle should prevail. The opposite holds for bottom-up models: for example, in the hermeneutic approach, the locus of authority is in practical know-how. In the RE method, none of the elements has a privileged status in the reasoning process. The strength of an RE is determined by the process of reflection and the coherence of the result, not by the epistemic status of one of the elements (Van Delden et al. 2005). Expanding the range of moral intuitions does not change this. We therefore would still characterize RE as a form of coherentism. However, we endorse the view that RE is not a pure, uncompromised coherentist model. We agree with Ebertz that moral intuitions enter into the reasoning process for reasons other than their coherence with other elements alone. They may simply have come to the thinker's mind or may be derived from empirical research. This is incompatible with pure coherentism. Moreover, the justificatory power of the elements in RE is ultimately dependent on their strength in the deliberative process in which they are studied and modified in the light of other beliefs. However, this does not rule out that a moral intuition can possess certain characteristics that add to the power of the intuition. We would therefore prefer to characterize NE-RE as coherentist with different levels of justificatory power.

In the argumentative process towards RE, the level of justificatory power of beliefs can be assessed. This assessment is added to the mutual testing and adjustment of moral intuitions, principles, morally relevant facts and background theories. In the assessment, at least three weighing factors can play a role. In the next section, we clarify the use of these factors when weighing moral intuitions. Moral intuitions may be more than other elements in need of support to give them sufficient warrant. However, elements such as principles and general moral values can be assessed accordingly.

10.8.3 Assessing the Justificatory Power of Moral Intuitions

In the argumentative process towards RE, the thinker can examine the set of moral intuitions by using some weighing factors, with the aim of assessing their justificatory power. We mention three such weighing factors: durability, transcendence and experienced perception. To preserve essential aspects of RE, such as the non-foundationalist character and the attitude of openness, none of the factors we mention here are decisive. Instead, they help the thinker examine the weight of intuitions in moral deliberation.

Durability: Moral intuitions can be weighted by their durability. We are likely to have more confidence in judgments that are confirmed in a history of cases. Durability can therefore be used as a weighing-factor. This implies that, for example, the judgments that match the common morality will have extra justificatory power. However, in our view the common morality should not be the only source of moral intuitions. Moral reasoning should be open to a broader set of initial beliefs than the common morality can supply.

Transcendence: The extent to which moral intuitions are appreciated and affirmed by a community is relevant for the justificatory power of a moral intuition. The degree of justifiability is then rated by their capability for transcendence (Van den Hoven 2006:163). Moral deliberation requires that individuals transcend their personal concerns and interests in such a way that others can share their perspective. This does not imply that individuals should disconnect from the values, projects and commitments they cherish personally. However, in the process of deliberation, moral intuitions that can be shared by people from different perspectives have more justificatory power.

Again, we stress the fact that we accept this merely as a weighing factor. Some authors, for example Nielsen, proposed to trade considered moral judgments in RE for the moral beliefs that are part of the consensus in a society (Nielsen 1982a). This interpretation of transcendence is at risk of being reduced to the majority is morally right (Nielsen) or in the case of Beauchamp and Childress, to the view that criticisms on local customs and attitudes are warranted only if they maintain fidelity to the common morality. Both interpretations are contrary to the idea that, in working towards RE, a thinker should strive to consider the broadest set of moral and non-moral beliefs.

Experienced perception: Experienced perception is a characteristic of persons that can only be acquired by a long process of gaining life experience and in-depth insight in the choices that people face in life (Van Willigenburg 1991: 205). Because experienced people may be better equipped for moral judgment in a case, their moral intuitions deserve to be taken up in the initial set of moral beliefs. In RE, moral intuitions should be evaluated with respect to the characteristics of their beholder. Agents who acquired a moral sensibility regarding a specific moral case should be identified. However, valuing experience perception is not the same as assuming that this always produces correct judgments.

Just as any other moral intuition, the ones that come from experienced persons are considered as preliminary fixed points.

These criteria can guide inspection of initial beliefs. There is no hierarchy, and these features of a moral intuition cannot simply add up to a final judgment about its justificatory power. The criteria can be mutually supportive. The features need to be addressed for each moral intuition in the specific context.

10.9 In Conclusion

Ethicists involved in practical ethics cannot overlook moral wisdom if they want to connect moral judgments and theories to a specific practice. Nonetheless, normative ethics requires independence from practice to preserve its critical force. With NE-RE we aimed to achieve a middle-ground in which empirical information on moral intuitions and normative reasoning are integrated in valid ways. The NE-RE model has to reply to those who question the viability of incorporating moral intuitions in a coherentist model of moral justification. This criticism is known as the no-credibility objection against RE. We claim that NE-RE has characteristics that provide arguments against both these criticisms.

Finally, the model of RE has been criticized for its limited practical clarity. We address the issue of measuring coherence and provide criteria that can help a thinker to decide when sufficient justification for his moral view is reached.

Chapter 11
Moral Expertise – The Role of Expert Judgments and Expert Intuitions in the Construction of (Local) Ethical Theories

Bert Musschenga

Reflective equilibrium (RE) is a widely accepted method both for the construction of ethical theories and for moral argumentation. Some authors explain its popularity by referring to its affinity to common sense. Wayne Norman says that the method amounts to little more than the codification of common sense (Norman 1998). In the view of others it roughly approximates the way in which many of us tend to think when we are dealing with practical moral problems (Dworkin 1978; Jamieson 1991). RE does not have high pretensions. It does not try to find a rock-hard foundation for moral judgments, nor does it pretend to produce certainty. At most, it performs a weaker form of warrant (Van der Burg and Van Willigenburg 1998: 3). As is often true, popularity does not exclude something from being controversial. The weak area of RE is the credibility of the initial judgments or moral intuitions.

In his 'Outline for a decision procedure for ethics' John Rawls says that the principal aims of ethics is the formulation of justifiable principles which may be used in cases wherein there are conflicting interests, to determine which one of them should be given preference. The main reason to accept these principles as justifiable is that they explicate the considered judgments, which are the mature convictions of competent moral men as they have been worked out under the most favorable conditions (Rawls 1951: 187). Who are these competent moral men or, as Rawls calls them, moral judges? A competent moral judge needs (i) to have a certain degree of intelligence (but no more than a normal intelligent man), is (ii) required to know those things concerning the world about him and those consequences of frequently performed actions, which it is reasonable to expect the average intelligent man to know, and, further, the peculiar facts of the case in which he has to express his opinion, is (iii) also required to be reasonable (be willing to find reasons for and against possible options, be open-minded and aware of the possible influences of prejudice and

B. Musschenga (✉)
Philosophy Department, VU University, Amsterdam, The Netherlands
e-mail: a.w.musschenga@vu.nl

M. Christen et al. (eds.), *Empirically Informed Ethics: Morality between Facts and Norms*, Library of Ethics and Applied Philosophy 32, DOI 10.1007/978-3-319-01369-5_11,

bias), and, finally (iv) needs to have a sympathetic knowledge of those human interests that, by conflicting in particular cases, give rise to the need to make a moral decision (1951: 178f.).

The image of the competent moral judge does not return in a theory of justice (ToJ). Rawls does speak of the qualities the person making the judgment needs to have, but in quite general terms. He is presumed to have the ability, the opportunity, and the desire to reach a correct decision (or at least not the desire not to) (1999: 42). These qualities are less detailed than those of the competent moral judge in the 1951 article. They also seem less central to Rawls' account, given the fact that he does not start with these qualities, as he did in the 1951 article, but mentions them only after saying which of our judgments to take into account. The reason for the competent moral judge's disappearance is that the role of well-considered judgments has changed in ToJ. Rawls now more openly embraces a coherence epistemology according to which these judgments, as the data for which the theory has to account, themselves can be adjusted, rectified, or even rejected in a process of mutual adjustment between judgments and principles (Van der Burg and Van Willigenburg 1998: 7). The judgments that serve as an input in the reasoning process need not be as well-considered as the judgments of a competent moral judge. It is sufficient that we discard those judgments made with hesitation, or in which we have little confidence, and those given when we are upset or frightened, or when we stand to gain one way or the other. It is sufficient for them to have initial credibility.

This assumption of initial credibility has been heavily criticized by a number of authors.[1] Brandt stated that the level to which we are committed to the beliefs involved in these judgments, their 'initial credence', does not tell us anything about their credibility (Brandt 1979; see also Brandt 1990). Another well-known critic of reflective equilibrium theory is Peter Singer who first formulated his objections in 1974 (Singer 1974). In 2005 he repeated his objections, now referring to a number of empirical studies that question the epistemic value of moral intuitions being products of evolutionary processes (Singer 2005). Empirical studies also figure in the critique of other authors such as Walter Sinnott-Armstrong, who denies that we can claim non-inferential justification for our moral intuitions because they are subject to too many distorting influences (Sinnott-Armstrong 2006, 2008a). I discussed the impact of such influences on the epistemic value of moral intuitions elsewhere (Musschenga 2010b).

My focus here is on empirical studies demonstrating that there is a significant relation between the reliability of someone's (moral) judgments and the level of his relevant expertise. Although the competent moral judge in Rawls' 1951 article does not need to have more qualities than the average intelligent morally maturated person (the 'ordinary moral person'), he also shows some similarity with an expert. The competent moral judge has, according to Rawls, to know the peculiar facts of the situation in which he has to express his opinion. Ordinary moral persons are continuously confronted by complex moral problems of which they do not know

[1] For an overview of the critics see Van der Burg and Van Willigenburg (1998: 7f).

sufficient details for making a well-considered judgment. It seems that in such situations some level of expertise is needed.

My aim in this article is to examine whether the quality of a reflective equilibrium can be strengthened by requiring that the initial judgments come from moral experts. I start in Sect. 11.1 with a description of the (nature of) expertise. In Sect. 11.2, I examine whether there is such a thing as moral expertise. In Sects. 11.3 and 11.4, I discuss the relation between expertise and reliability of (moral) judgments. Section 11.5 deals with the relation between moral expertise and moral principles. In Sect. 11.6, I go into the relevance of ethical theorizing. Section 11.7 discusses whether ethics should limit itself to the initial judgments of moral experts. In Sect. 11.8, I draw some conclusions.

11.1 The Nature of Expertise

Studies within cognitive psychology have shown the superiority of experts over novices in nearly every aspect of cognitive functioning, from memory and learning to problem solving and reasoning (Anderson 1981). Chess masters, for instance, have been found to perceive patterns of play more effectively (De Groot 1965) and to have better memory for chess positions. Charness (1976) showed that expert chess players do not rely on a transient short-time memory for storage of briefly presented chess position. They are able to recall positions even after the contents of their short-term memory have been completely disrupted by an interfering activity (Charness 1976). Subsequent research has shown that chess experts have acquired memory skills that enable them to encode chess positions in long-term memory (Ericsson and Kintsch 1995). Experts in physics, mathematics, and computer programming reveal similar superior skills (Mayer 1983). Several insights have emerged from this body of research:

1. Expertise is domain specific. The special skills of an expert are diminished outside his area of expertise: "Chess experts do not appear to be better thinkers for all their genius in chess" (Anderson 1990). Apparently, the thinking of experts is 'domain adapted' (Slatter 1987).
2. Expertise is acquired through stages of development, somewhat akin to the mental development of children. According to Fitts and Posner (1967), the first is the 'cognitive stage,' where specific facts are memorized to perform the task. The next is the 'associative stage,' where connections between successful elements are strengthened. The last is the 'autonomous stage,' where the skills become practiced and rapid.[2]
3. Experts see and represent a problem in their domain at a deeper, more principled level than novices who tend to represent problems at a superficial level (Chi et al. 1988).

[2] Dreyfus and Dreyfus (1991) distinguish between five stages in the acquisition of expertise: novice, advanced beginner, competence, proficiency, expertise.

4. Experts use different thinking strategies. For instance, novices have been found to reason backwards from the unknowns to the givens in solving physics problems. Expert physicists, in contrast, reason forward using stored 'functional units' from the givens to the goal (Larkin 1979). Therefore, expertise produces more efficient approaches to thinking about problem solving and decision making (Anderson 1990).
5. The thinking of experts is more automated (Shiffrin and Schneider 1977). These automated processes generally operate in parallel and function somewhat like visual perception or pattern recognition. Novices, in contrast, rely on controlled processes, which are linear and sequential, more like deductive reasoning (Larkin et al. 1980). Because of their special abilities, expert processes are reflected by and can be studied through verbal protocols. Experts are asked to think aloud, qualitatively (Simon 1980). Although other methods have been proposed (Hoffman 1987), protocol analyses are commonly used to provide the raw data for building expert systems (Slatter 1987).

In sum, the cognitive science view is that experts within their domains are skilled, competent, and think in qualitatively different ways than do novices (Anderson 1981; Chi et al. 1988). They have skills to develop complex representations that allow them immediate and integrated access to the demands of action in current situations and tasks. These acquired skills can also account for their superior memory performance, such as recalling a briefly presented chess position (Feltovich et al. 2006).

11.2 Moral Expertise

Modern knowledge-based societies need many kinds of expertise. In every domain of expertise there are three categories of people: outsiders (laymen), novices, and experts. Most people are neither able to play chess, nor interested in that game. They are outsiders to the game of chess. An outsider who is interested, starts learning to play chess, but remains a novice. Only a few of those who do play chess are chess experts. Most people have no medical knowledge or skills. They are outsiders to the practice of medicine. Most physicians, such as general practitioners, possess broad medical knowledge and general skills. We are not used to calling general practitioners experts, but compared to medical students, they do have expertise. For us, medical experts are specialists in parts of the body, in applying special techniques or in performing complex interventions. A former neighbor of mine was an experienced dentist. When he decided to specialize in dental surgery, he became again a novice.

Are there also moral experts? Dreyfus and Dreyfus (1991) explicate (acquiring) moral competence in terms of (acquiring) mastery of certain skills. So do Narvaez and Lapsley (2005). For them all morally maturated persons are 'moral experts.' The novices are the very young children who are still at the first stage of their moral

development. The amoralists, by contrast, are clearly the outsiders. They are unable to 'play the morality game.' Although defensible, this use of the concept of moral expertise squares with common language.

In my view, however, moral experts are experts only in a particular domain of morality. A morally mature person who aspires to be a medical ethicist becomes, notwithstanding a general moral competence, a novice in medical ethics as soon as he starts his training. He then becomes a novice in a specific domain of morality, the domain of medical ethics. Roles, practices, and institutions as well as social spheres (e.g., law, politics, economics) can be seen as moral domains. In my view moral expertise is, as any other kind of expertise, domain specific. A medical ethicist cannot claim any expertise in, for instance, the domain of social security or animal ethics.

I stated that a moral expert is an expert in particular domain of morality. But we still do not know how to identify such an expert. Leaning on Rest's (1983) review of social development research, Narvaez and Lapsley (2005) have identified the characteristic skills of ethical experts.[3] These skills extend Rest's four psychologically distinct processes (moral sensitivity, moral judgment, moral motivation, and moral action) by outlining a set of social, personal, and citizenship skills. Although I am not in favor of calling ordinary moral persons 'ethical' or 'moral experts,' Narvaez and Lapsley's view of the skills of such an expert might also be relevant for identifying the qualities of the specialized, real, moral expert.

Experts in the skills of moral sensitivity, for example, are able to 'read' a situation and to determine their role in it more quickly and accurately. These experts are also better at generating functional solutions, due to a greater understanding of the consequences of possible actions. Experts in the skills of moral judgment are more adept at seeing the crux of a problem quickly, bringing with them many schemas for reasoning about what to do, and solving complex problems Their information-processing tools are more complex, but also more efficient. Experts in the skills of moral motivation are capable of maintaining their focus on prioritizing the ethical ideal. Their motivation is directed by an organized structure of moral self-identity. Experts in the skills of moral action are able to keep themselves focused and take the necessary steps to get the moral job done. They demonstrate superior performance when completing a moral action.

Narvaez and Lapsley's view of moral skills differs from that of expertise theory as developed within cognitive science. The skills that constitute a mature moral person (in their terminology, a person of good character) include motivation and volition. Experts as seen in cognitive science solely distinguish themselves by cognitive and practical skills, by their knowledge and competence in identifying and solving certain problems, or in performing certain types of action. Using Narvaez and Lapsley's terminology: experts catch our attention by their skills in ethical sensitivity and ethical judgment, not by their skills in ethical motivation and ethical

[3] Narvaez and Lapsley use the adjective 'ethical' but I prefer the adjective 'moral.' They do not discuss whether ethical (moral) expertise is domain specific.

action. I will argue that moral experts need not be moral exemplars or persons of exceptional moral integrity.

In the last two decades there has been a continuous debate, mainly within medical ethics, on the existence and the nature of moral expertise.[4] The central question in that debate seems to be whether knowledge of, and training in ethics (ethical theories and moral argumentation) makes one a genuine moral expert. The answer to that question largely depends on one's view of the goals and tasks of ethics as an academic discipline. Some ethicists find that the task of ethics is to elucidate and articulate the various relevant, sometimes conflicting moral perspectives on a problem. For example, some believe ethicists should not prescribe morally competent persons how to solve practical problems, while others think that ethics makes no sense if it cannot guide people in making decisions. Referring to the divergent answers given by ethicists to practical moral problems in medicine and health care, health professionals are inclined to deny that ethicists have expertise.

I believe that there is such a thing as moral expertise in the domain of medicine and health care. However, knowledge of, and training in the academic discipline of specific expertise is not sufficient for becoming a moral expert, and perhaps not even necessary. The right approach to answering the question whether moral expertise exists in, for instance, the domain of clinical ethics is to proceed as Wear suggests: "[…] by describing the skills and knowledge that are ingredient in clinical ethics expertise, and then presume to observe that this description goes far toward providing sufficient proof of the normative claim that it exists and is worthy of the name" (Wear 2005, p. 243). Clinical ethics experts, says Wear, "[…] need and can claim knowledge in a number of areas, including the law, institutional policy, the characteristics of the clinical context, and various codes and standard statements of agreed upon concepts" (2005: 250). More generally, a moral expert is someone who, by virtue of his knowledge, training, experience, and other 'skills of ethical judgment and ethical sensitivity' (Lapsley and Narvaez 2005), is competent enough to make justifiable judgments on issues in his particular moral domain. Part of his expertise is also that he is able to defend his judgment in a convincing manner. Moral experts are better equipped to make authoritative and convincing judgments on issues in a particular domain than novices and outsiders, but only on issues in that particular domain.[5]

I do not think that there is, or can be, moral expertise in every moral domain. Expertise presupposes institutionalized contexts with an accepted body of theoretical and practical knowledge, relevant documents, policies, laws, precedents, skills, and so on. In many countries there are moral experts in the ethics of experiments with

[4] For an overview of the discussion on moral/ethical expertise see Weinstein (1994) and Rasmussen (2005).

[5] In some circles it is thought that the practitioners within a moral domain are also the moral experts. In this view, physicians as practitioners in a particular domain of medicine are also the moral experts in that domain. I do not agree. Not all physicians have sufficient moral sensitivity and knowledge of relevant concepts, policies and protocols to qualify as moral experts. Neither is it required for moral experts in the domain of medicine to have all the medical knowledge that physicians possess (Musschenga 2010a).

animals. However, as I will argue in Sect. 11.7, in countries that lack legislation, policies, and procedures protecting the well-being and interests of animals, there can be no such moral experts. Although in the Netherlands there is a growing ethical reflection on the moral dimensions of the use of nanotechnology, I do not believe there are already moral experts in that field.

11.3 Experts and the Reliability of Intuitive Judgments

Behavioral studies of skill acquisition have demonstrated that automaticity is central to the development of expertise, and practice is the means to automaticity (Posner and Snyder 1975). Through practice, the speed and the smoothness of cognitive operations improve, which leads to a reduction of the cognitive demands of the situations, thus releasing cognitive resources (such as attention) for other, usually higher cognitive functions such as planning and self-monitoring (Feltovich et al. 2006: 53). The judgments of experts are usually not the product of deliberate reasoning but of unconscious and automatic processes. The role of automaticity in experts' judgments might also explain why expert judgments are generally more reliable than judgments of non-experts.

Since there are several kinds of automatic processes, it is relevant to know which processes underlie the intuitive judgments of experts. According to psychologist Bargh (Bargh 1989, 1996), automaticity has been invoked to explain the following process effects: (1) effects of which a person is not aware, (2) effects that are relatively effortless such that they will operate when attentional resources are scarce, (3) effects that are unintentional, occurring even in the absence of explicit intentions or goals, (4) effects that are autonomous in that they will run themselves to completion without the need of conscious attentional monitoring, and (5) effects that are involuntary or uncontrollable, even when one is aware of them. Attention, awareness, intention and control do not necessarily occur together in an all-or-none fashion. They are, to some extent, independent qualities that may appear in various combinations. Bargh (1989) argues that these automatic effects fall into regular classes: those that occur prior to conscious awareness ('preconscious automaticity'); those that require some form of conscious processing but produce an unintended outcome ('postconscious automaticity'); and those that require a specific type of intentional, goal-directed processing ('goal-dependent automaticity').

Intuitive judgments of experts belong to the third class of automaticity, goal-dependent automaticity. Goal-dependent automaticity appears in an unintended and an intended form. In goal-dependent automaticity with unintended effects, the perceiver is aware of the stimulus but not necessarily of its effects on cognitive processes; such effects nevertheless require some cognitive capacity and depend on the perceiver's goal. Thus, for example, inferring a trait from a written description of behavior seems to occur spontaneously at encoding; it occurs without intent or awareness, is subjectively effortless, and is difficult to disrupt with a

concurrent task (Bargh 1989: 20). An important example of intended goal-dependent automaticity is the skillful behavior of an expert. Intended goal-dependent automaticity occurs autonomously and outside awareness, but the output was intended by the goal of the current processes. Well-learned situational scripts or thoroughly routine action sequences typically operate autonomously, with little need of conscious control or significant attentional resources. Another type of goal-dependent automaticity is 'incubational automaticity.' This is goal-directed thought that continues after one's conscious attention has moved on to other concerns (Bargh 1989: 24). This type of thinking is also characterized as 'unconscious thinking.'

The reliability of intuitive judgments, in comparison to that of deliberated judgments, is a hot topic in social psychology.[6] In spite of the popularity of the subject, there is still not much hard evidence for the superiority of intuitive judgments. Exceptions are the studies of Dijksterhuis and his colleagues on the reliability of the process of 'unconscious thinking' or 'deliberation without attention.' This defines a process that takes place when we 'sleep on something' to get more clarity in what we want. These studies contend that the reliability of unconscious thinking is superior to both immediate judging and conscious deliberation (Dijksterhuis 2004; Dijksterhuis et al. 2006; Dijksterhuis and Nordgren 2006).

According to Hammond (1996) both intuitive and analytical thinking produce errors, although the kinds of errors produced tend to be different. Hammond states that intuition rarely results in responses that are precisely correct, because it involves the tacit aggregation of different informational cues. Errors are not likely to be large, however, because of the absence of systematic biases. Systematic biases occur in deliberate thought. A small error, such as a minor mistake in a calculation, can lead to huge errors in the final result. Errors in deliberate thought tend to have an 'all or nothing' quality.

Hogarth (2002) concedes that the major problem in assessing the evidence on the advantages and disadvantages of intuitive and deliberate systems is that few studies have been conducted with this issue specifically in mind. Most relevant are the studies on expertise. Hogarth states that one must consider (a) the trade-off and error implicit in tacit, automatic thinking and (b) the probability that a person will know the appropriate deliberate 'formula.' He assumes that the greater the complexity a task exhibits in analytical terms (as measured, e.g., by number of variables, types of functions, weighting schemes, and so on) the less likely it is that a person will both know the appropriate formula and apply it correctly (Hogarth 2002: 32). His conclusion is that deliberate thought should be preferred to intuitive thinking when analytical complexity is easy. However, as analytical complexity increases, tacit processes become more accurate in a relative sense, which means that the increasing probability of making errors in analysis eventually outweighs the bias and error in tacit responses.

[6] See a.o. Hogarth (2002) and Woodward and Allman (2007).

11.4 Experts and the Reliability of Intuitive Moral Judgments

In the previous section, I discussed the reliability of intuitive judgments in general, but what can be said about the reliability of intuitive moral judgments? There is hardly any research concerning the reliability of intuitive judgments with a focus on morality. It is improbable that the number of such studies will rapidly increase. The reason is that reliability studies presuppose a consensus on the criteria for accuracy or reliability. As with many other intuitive judgments, intuitive moral judgments are made in contexts in which there are no explicit criteria for their accuracy. However, some research is done into the reliability of unconscious thinking on moral issues. Building on the research by Dijksterhuis and his colleagues (2006), the Dutch psychologists Ham et al. (2009) investigated the possible merits of unconscious thinking for people's justice judgments. They studied justice judgments on the fairness of application procedures. Ham et al. conducted two experiments. In both, participants were presented with complex and extensive information about four application procedures that job applicants had experienced. One of these descriptions of an application procedure implied a predominantly fair application procedure, and one implied a mostly unfair application procedure. The two remaining descriptions implied neither very fair nor very unfair application procedures. Each application procedure was described by a list of 14 items, yielding a total of 56 different items. There were three categories of items: just items, unjust items, and filler items. Ten items were used to describe just elements of the application procedure (e.g., "The application procedure was clearly explained"). Ten items were used to describe unjust elements of the application procedure (e.g., "Of four administered tests, only one was examined during applicant selection"). Another eight items were justice-neutral items that were (slightly) related to social justice but did not necessarily imply a just event or an unjust event (e.g., "The applicant had to wait upon arriving"). The remaining 28 items served as filler items and were not directly related to social justice (e.g., "The company website was reasonably well taken care of"). These justice-neutral items and filler items were included in order to increase the complexity of the decision problem. After this information had been presented, some participants (the conscious thought condition) could think about their justice judgments for 3 min and then were asked to indicate their justice judgments. Other participants (the unconscious thought condition) performed a distracter task for 3 min which prevented conscious thought about the justice judgments they had to make, after which they were asked to indicate their justice judgments. The remainder of the participants were asked to make a justice judgment immediately (immediate judgment condition).

In experiment 1, participants were asked to directly compare the justice levels of the four application procedures and to indicate which procedure was the most just. In experiment 2, participants made their justice judgments comparable to the assessment of justice judgments in earlier justice research. They indicated their justice judgments on rating scales for each application procedure separately.

The dependent variable the researchers constructed in all experiments was the accuracy of participants' justice judgments. They constructed accuracy scores that indicated whether participants correctly indicated the appropriate application procedure to be the most fair application procedure, the appropriate application procedure to be the most unjust procedure, and the appropriate two other ones as intermediate justice levels. The results provide evidence for the merits of unconscious thought for justice judgments as these findings are the first to reveal that the accuracy of justice judgments increases under conditions that allow for unconscious thought relative to conditions of conscious thought or immediate judgment. The findings further indicate that unconscious thought can lead to more accurate justice judgments than both conscious thought and immediate judgment.

The findings of Ham et al. show that unconscious thinkers made the most accurate justice judgments. In the study by Ham et al. the criteria for fair application procedures were given. Studies on the reliability of the (intuitive) judgments of moral experts are possible since moral expertise, at least in my conception of it, presupposes institutionalized contexts with an accepted body of theoretical and practical knowledge, of documents, policies, laws, protocols, and precedents. For example in the Netherlands, there are committees that examine whether the decisions taken by doctors to carry out euthanasia are justifiable. A study examining the reliability of the judgments of the more experienced members of such committees compared to that of laypeople or novices should be possible, given the existing legislation and other policy documents. However, the public debate on the admissibility and desirability of human enhancement by gene selection started not long ago, and the views have not *crystallized* out.

11.5 Moral Expertise and Moral Principles

I argued that in their particular moral domain, the judgments of moral experts are superior to those of novices and outsiders. If I am right, this implies that, from the standpoint of reflective equilibrium theory, an ethical theory regarding that domain should at least, or maybe primarily, match the 'intuitive' judgments of moral experts in that domain.[7] However, there is a certain tension between moral expertise theory and the theory of reflective equilibrium. While reflective equilibrium theory says that neither intuitions (considered judgments) nor principles have priority, some adherents of moral expertise theory give priority to the intuitive judgments of experts and downplay or even reject the role of principles. Therefore, in this section I am going to discuss why we need principles for a particular domain, say clinical ethics, if there are moral experts in that domain.

[7] I do not think that this claim conflicts with Rawls' view. Rawls' aim in *ToJ* is to find principles of justice that could serve as a moral basis for designing the basic structure of society. I doubt that there are moral experts in that domain, whose well-considered judgments Rawls could have given a special place.

According to Dreyfus and Dreyfus (1991), the moral expert is not necessarily an expert in applying moral principles. Moral principles are aids for the inexperienced, for those who still need instruction. On the highest stage in the model of expertise acquisition, the expert leaves rules and principles behind and develops more and more refined ethical responses (Dreyfus and Dreyfus 1991: 237). Dreyfus and Dreyfus do not deny that experts deliberate, but their deliberation is, in most cases, based on intuitions. Even in situations that are problematic though not unfamiliar, an expert's deliberation is still based on intuitions: he deliberates about the appropriateness of his intuitions (pp. 240f). Only in a novel situation in which he has no intuition at all, must an expert resort to abstract principles like a novice (p. 247). A similar view is defended by Churchland (1996: 106, 107):

> The portrait of a moral person as one who has acquired a certain family of perceptual and behavioral skills contrasts sharply with the more traditional accounts that pictured a moral person as one who agreed to follow a certain set of rules. […] State-able rules are not the basis of one's moral character. They are merely its pale and partial reflection of the comparatively impotent level of language.

Authors like Churchland (1996) and the Dreyfus and Dreyfus (1991) believe that recent work in cognitive science and Artificial Neural Networks confirm Aristotle's views in which moral judgments require practical wisdom, gained by rich and sound experience. According to cognitive scientists, neural networks such as the human brain are capable of extracting and encoding information (knowledge) in forms with richness, fluidity, and context-sensitivity that far outstrips anything that could be supported by a set of linguistically couched action selection rules, principles, or maxims. At the heart of this view Clark says is (2000: 270), "[…] a daunting story about vectors, prototypes, high-dimensional states and non-propositional, distributed encodings."[8] This computational story fits in, according to Clark, with work from cognitive psychology suggesting that much human knowledge is organized around encoding of prototypical cases rather than via the use and storage of rules and definitions.[9] Clark does not agree with Churchland (1996) and Dreyfus and Dreyfus (1991). According to him the marginalization of what he calls 'summary linguistic formulations and sentential reason' in their work is a mistake (Clark 2000: 274). Summary linguistic formulations are not mere tools for the novice. Rather, they are essential parts of the socially extended cognitive mechanisms that support communal reasoning and collaborative problem solving, and thus are crucial and (as far as we know) irreplaceable elements of genuinely moral reason. They are the tools that enable the cooperative explorations of what he calls 'moral space': a space that is intrinsically multi-personal and whose topology is defined largely by the different, but interacting needs and desires of multiple agents and groups (Clark 2000: 274).

Neither the Kantian view of moral judgment that gives priority to general principles nor the Aristotelian view that gives priority to particular judgments captures,

[8] Clark refers a.o. to work by McClelland (1989) and Churchland and Sejnowski (1992).

[9] Here Clark refers to Rosch (1973) and Smith and Medin (1981).

according to Clark, the subtlety, power and complexity of human moral intelligence:

> Instead it is the cognitive symbiosis between basic, prototype-style, pattern-based understanding and the stable surgical instruments (for learning, criticism and evaluation) of moral talk that conjures moral understanding (Clark 2000: 279).

Clark calls moral maxims and recipes (to which I add principles) 'anchor points for moral thought and reason':

> They are the re-visitable islands of order which allow us to engage in exploratory moral discourse, approaching practical moral problems from a variety of angles while striving, nonetheless, to maintain a sense of our targets, priorities and agreed-upon intermediate positions (Clark 2000: 278).[10]

I agree with Clark that the process of thinking and judging by experts can be based on patterns and prototypes as long as they deal with familiar problems. However, many moral debates in our society do not regard (normal or problematic) familiar problems, but novel ones. Many of these novel problems require public debates and collective decisions.[11] This is where 'collaborative exploration of moral space' —to use Clark's terminology (in Clark 2000)—has to take place. This is also where moral experts need moral principles.

11.6 Do We Also Need Ethical Theories?

I argued that recognizing the role of moral expertise does not commit us to denying a role for moral principles. Principles are generally derived from, or embedded within ethical theories. Do we also need ethical theories? Reflective equilibrium theory stresses the need for a coherent set of moral principles, which in my view is the aim of ethical theories. Nussbaum (2000), who in her earlier work (1986) stressed the importance of Aristotelian practical wisdom, surprisingly answered that question positively. She gives the following definition of ethical theory: '… a set of reasons and interconnected arguments, explicitly and systematically articulated, with some degree of abstractness and generality, which gives directions for ethical practice' (Nussbaum 2000: 233f.). She formulates six criteria for ethical theories. Ethical theories (1) give recommendations about practical problems, (2) show how to test the correctness of beliefs, rules and principles, (3) systematize and extend beliefs, (4) have some degree of abstractness and generality, (5) are universal, and (6) are explicit (Nussbaum 2000: 234ff). An ethical theory is not a system of rules. Ethical theories formulate the point and purpose of rules, which enable us to

[10] In his response to Clark's article (Clark 2000), Churchland recognizes the role of discursive moral rules. At the same time he underlines that our internal representations and cognitive activities are not just hidden, silent versions of external statements, arguments, dialogues and chains of reasoning that appear in our overt speech and print (Churchland 2000: 294).

[11] See also Musschenga (2009).

determine when the point is better served not by following a rule, but by making an exception to the rule. Unlike systems of rules, ethical theories also give arguments for their conclusions (Nussbaum 2000: 236–241).

I agree with Nussbaum (2000) in her defense of ethical theories. However, if it is important that ethical theories are supported by the judgments of moral experts then ethical theories can only be local and contextual, since moral expertise is local and domain specific. Such theories have a low level of generality. In the empirical sciences it is widely accepted that theories often cannot have both a high level of universal validity and a high level of generality. A trade-off is then needed. There is no reason why this phenomenon should not also occur in ethics.[12]

11.7 Moral Expertise and Its Limits: The Case of Animal Ethics

Moral experts are in important respects similar to legal experts. Both presuppose institutionalized contexts with an accepted body of theoretical and practical knowledge, of documents, policies, laws, protocols, and precedents. Both have expertise in a particular domain, of law and morals, respectively. An important similarity is also that neither the legal nor the moral expert themselves create the policies and documents that are part of their expertise. Legal experts do not make laws. In democratic countries that is the prerogative of legislative bodies such as parliaments. For moral experts the ultimate source of the moral beliefs and values that guide their judgments and decisions are not their own moral views, but the views of society at large which are embodied in the relevant documents and policies. Moral experts can of course contribute to public debates on the moral framework within which they work, but they do not solely determine this framework. I will illustrate this by the example of expertise in animal ethics.

The subject of animal ethics is the protection of animal welfare. It is widely accepted that welfare for animals means that animals should feel well—be free from prolonged and intense fear, pains, and other negative states, and experience comfort, contentment and normal pleasures—and function well—have a satisfactory health, normal growth, normal functioning of physiological and behavioral systems, and lead natural lives through the development and use of their natural adaptations and capabilities. Pigs should have the opportunity to root in the mud; chimpanzees to live in groups; chickens to pick in the sand for seeds, worms, and so on.[13] This view underlies the Dutch Animal Health and Welfare Act of 1992. This law also prohibits, with a few exceptions, interventions that remove a part or parts of animal, such as the cutting of tails and ears. This moral framework of comprehending animal welfare plus the integrity of the animal body provides sufficient orientation for

[12] See Van der Steen and Musschenga (1992) on the trade-off between different methodological criteria in science and ethics.

[13] See, e.g., Fraser et al. (1997) and Rutgers and Heeger (1999).

those whose task it is to protect the interests of animals. Nowadays many animal ethicists find this view of the protection of animal interests unsatisfactory. They plead for a broader view on what it means to lead natural lives through the development and use of their natural adaptations and capabilities. They find that wild animals such as lions and tigers, should not be kept in zoos, factory farming should be abolished, animals should not be used for medical experiments, and so on. In their view, caring for the welfare of animals includes creating conditions in which they can have a life that accords with their species-specific capacities and adaptation patterns.

Moral experts in animal ethics need to work within an accepted local moral framework that informs the current legislation and policy documents. Nothing in the expertise of such a moral expert forces him to adopt the broader view of the natural functioning of animals, because this view is not a logical extension of the narrow view. It is informed by moral intuitions which rest on a view of animal flourishing that has its adherents both among animal ethicists and the general public. An established moral expert in animal ethics might not share these intuitions. This suggests that animal ethics should not limit itself to the judgments of moral experts.

11.8 Conclusion

My aim in this article was to examine whether the quality of a reflective equilibrium can be strengthened by requiring that the initial judgments come from moral experts. My conclusions seem to be contradictory. On the one hand, I argued that the reflective equilibrium of local ethical theories can be strengthened by giving special weight to the judgments of moral experts. The judgments of moral experts are superior to those of laypeople if they stay within the locally accepted moral framework. On the other hand, moral intuitions sometimes transcend accepted moral frameworks. In such cases it is not up to moral experts to determine whether such intuitions are relevant and should be accommodated for within an ethical theory.

Acknowledgements I am grateful to Nicole van Voorst Vader, Robert Heeger and Markus Christen for their comments on earlier versions.

Chapter 12
Social Variability in Moral Judgments – Analyzing the Justification of Actions Using the Prescriptive Attribution Concept

Erich H. Witte and Tobias Gollan

In 1958 Fritz Heider published his groundbreaking monograph "The psychology of interpersonal relations." For social psychologists it proved to be a rich source of conceptual ideas and gave rise to some of the "grand" theories of the discipline (e.g. balance theory, theory of justice, and attribution theory). From our perspective, however, the book in its theoretical richness is not yet fully appreciated (Gollan and Witte 2008), especially the 8th chapter with the title "Ought and Values". For instance, Heider (1958) conceives 'oughts' and 'values' as people's culturally shared concepts of what should be attained or done. They refer to what people consider to be "right" or "wrong" and are therefore crucial elements in moral behavior, ethical decision-making, and ethical justification. In our chapter, we adopt Heider's idea of "rights" and "wrongs" in social contexts and combine it with another of Heider's notions that has received even more attention: causal attribution. We argue below that this juxtaposition of ideas represents a logical precursor to the concept of prescriptive attribution (Witte and Doll 1995).

Our key assumption is that explaining why an action is evaluated as morally good or bad—not moral reasoning but solely moral or ethical justification—in many ways resembles explaining why an action or event occurred. In other words, we suggest that ethical justification is similar in logic to explaining causal factors. Speaking with Heider, who considers a person explaining the causal conditions of an event as one proceeding like a "naïve scientist," we regard a justifying person as one proceeding like a "naïve ethicist." Because explaining an event by its causal factors is referred to as "causal attribution," we will term the justification of an action by ought-standards "prescriptive attribution."

In the following we will (1) introduce the prescriptive attribution concept by its formal analysis, (2) report two extensions to the model that were included as a

E.H. Witte (✉)
Department of Psychology, University of Hamburg, Hamburg, Germany
e-mail: erichwitte@aol.com

T. Gollan
Research Institute for Quality in Child and Youth Support, Foundation 'Die Gute Hand', Kürten, Germany

M. Christen et al. (eds.), *Empirically Informed Ethics: Morality between Facts and Norms*, Library of Ethics and Applied Philosophy 32, DOI 10.1007/978-3-319-01369-5_12,

consequence of empirical studies, (3) present two measures that offer operationalizations of prescriptive attributions, and finally (4) report empirical results describing specific factors that influence how prescriptive attributions are construed in everyday life.

12.1 Conceptual Analysis of 'Ethical Justification'

12.1.1 Prescriptive Attribution

Before we discuss the terms 'ethical principles' and 'justification' in detail, we will outline the similarities of the causal and the prescriptive attribution concepts from a formal perspective. In order to do this, a conceptual analysis of both causal and prescriptive attribution is provided in Table 12.1. In the left column, causal attribution is decomposed into its fundamental elements (cf., Witte 1994), contrasted by the elements of prescriptive attribution in the right column of the table (cf., Witte and Doll 1995). For both types of attributions, we provide examples.

In the example below, all judgments of abortion are positive, regardless of their underlying ethical principle. This is reasonable, since the justification of behavior requires its positive evaluation. Nevertheless, deriving judgments of an action on the basis of ethical principles may also yield negative evaluations, and even applying the same principle to a specified action might yield different evaluations. In the example in Table 12.1, abortion is deontologically judged as right with regard to the universally valid principle of self-determination (R_1). At the same time, applying the deontological position in a different way may result in a negative evaluation (e.g., by highlighting the unborn child's right to live). This illustrates that ethical positions do not automatically imply a clear-cut evaluation; they are formal rules that need to be enriched with information on the action and its context.

Our focus lies primarily on the conceptual decomposition of the prescriptive attribution concept in the right column of Table 12.1: To a given action (A) which is to be justified, an ethical principle (E_i) is applied which results in a judgment (R_i) of the action as being right, with respect to the principle applied. In the example, applying the different ethical principles to abortion yields four different judgments that show why having the abortion was the right choice. These judgments can be differentiated (Js_i) according to how relevant the respective ethical principle is for the given action, so that some judgments are more persuasive and some are less. In the example, one judgment (R_1) is regarded as persuasive, while the other three are not. Note that the evaluations Js_i of the judgments as more or less persuasive or "valid" are subjective, so that the same judgments $R_i(A; E_i.)$ in Table 12.1 might also be evaluated differently. On the basis of this differentiation (and of other potential factors, to be discussed below), some of the judgments are incorporated into the final justification (J). Since, in the example, only one judgment was subjectively considered to be valid, the final justification only refers to this one judgment (R_1) and its respective ethical principle (E_1).

Table 12.1 The structure of causal (left column) and prescriptive attribution (right column). The example on the right side of each column illustrates the elements of the respective attribution

	Causal attribution (CA)		Prescriptive attribution (PA)	
1.	There is an action. A.	Monica has a dispute with her nephew, John (A)	There is an action A.	Paula chooses an abortion (A)
2.	There are causal sources driving the person's action	The causal sources lie within the acting person (CS_1)	There are classic ethical positions that justify an action	One must obey universal rules (E_1)
	CS_i	The causal sources lie within the circumstances (CS_2)	E_i.	One must consider the action's consequences for all affected parties (E_2)
		The causal sources lie within the object toward which the action was addressed (CS_3)		One must do what brings no personal disadvantages (E_3)
				One must do what seems intuitively right to oneself (E_4)
3.	There are arguments (Ar_i) that link the causal sources CS_i with the action (A)	Monica argues with John, because she is a quarrelsome person and is easily irritated (Ar_1)	Judgments of an action are based on the perceived relation (R_i) between ethical positions (E_i) and the action (A)	Since every woman (as every human being) has the right to self-determine what happens with her body, it is her right to choose an abortion: (R_1)
	Ar_i (A, CS_i)	Monica argues with John, because she had a bad day at work (Ar_2)	R_i(A, E_i.)	Since Paula does not truly want the baby, it is better to choose the abortion, because the baby's need for loving care could never possibly be fulfilled by Paula: (R_2)
		Monica argues with John, because John caused a lot of trouble (Ar_3)		It would not be convenient for Paula to have a baby now, so it was right to choose the abortion: (R_3)
				Abortion seems to be a good choice to Paula, so it was right to select it: (R_4)

(continued)

Table 12.1 (continued)

	Causal attribution (CA)		Prescriptive attribution (PA)	
4.	There is a subjective differentiation of the arguments with respect to their relevance for the explanation of the action (attribution strength: As_i)	Ar_1 is not a good explanation (because we know that Monica usually is patient and well-tempered with John (As_1)	There is the subjective differentiation of the judgment according to the importance of the ethical position for the action (judgment strength: Js_i)	R_1 is a persuasive judgment (because it draws on a universally valid principle) (Js_1)
	As_i [Ar_i (A, CS_i)]	We do not know, if Ar_2 is a good explanation or not, because we do not have information about Monica's day at work (As_2)	Js_i [R_i(A; E_i)]	R_2 is not a persuasive judgment (because it neglects that the baby's need for loving care may be fulfilled in other ways, e.g. to have the child adopted) (Js_2)
		Ar_3 is a good explanation (because John is known to be a very difficult and defiant kid) (As_3)		R_3 is not a persuasive judgment (because the interests of the baby are not considered at all) (Js_3) R_3 is not a fair judgment (because it leaves out any moral considerations that apply in this situation): (Js_4)
5.	There are also person-related explanations PE of the action, which consist of the same patterns of arguments (As):	"To me, it seems that Monica argues with John because he is a kid who is not easy to handle" (E)	There is a justification (J) of an action as "good" or "bad," "right" or "wrong"	"In my opinion, by choosing the abortion Paula did the right thing, because her prerogative to self-determine what happens to her body outranks all other considerations." (J)
	PE {As_i [Ar_i (A, CS_i)]}		J(Js_i [R_i(A, E_i)])	

In brief, the conceptual analysis yields a formalized definition of prescriptive attribution which is represented in a quintuple comprising all elements described above:

$$PA=\left\{A;\ E_i;\ R_i\left(A,E_i\right);\ Js_i\left[R_i\left(A,E_i\right)\right];\ J\left(Js_i\left[R_i\left(A,E_i\right)\right]\right)\right\}$$

Causal attribution always refers to the causal origins of actions. Thus, causal attribution can be applied to all actions that have a set of possible causal origins

(which are, indeed, all actions that can be conceived). In contrast, the scope of actions that prescriptive attribution can be applied to is restricted to specific actions and situations. These limitations result from its theoretical design as a framework for ethical justification.

First, only those actions or decisions that are taken deliberately and that the person feels *accountable for* can be justified because it is intended. Second, an *ethical* justification is only reasonable for behavior in situations that are at least in part characterized by ought requirements, in contrast to situations where oughts are irrelevant. One could also term these situations as 'moral' or 'value-laden' situations. For example, it makes no sense to ethically justify sitting on a public bus—unless an old lady must stand and you are physically in much better shape than she is. Generally, situations with *ought* requirements occur when harm to other creatures is involved. However, which situations specifically fall under a moral scope depends on the cultural context and its historical circumstances. Likewise, from a historical perspective the variability of oughts becomes evident when considering how much a human life was worth in the medieval times. Haidt et al. (1993) showed that there are culturally specific conceptions of morality (which may be interpreted as *oughts* in Heider's sense) that may even apply to situations when physical harm plays no role at all. For instance, burning the national flag of one's own country may be a serious moral transgression in only some cultures, whereas sex among siblings, even if they are both infertile and of full age, is disturbing in almost all cultures.

Thus, assuming that ethical justification only applies to actions that can generally be evaluated as positive turns out to be questionable—since the evaluation is highly context-dependent and its scope very ample. Later in this chapter, we will discuss an extension of prescriptive attribution concept that applies when the actions are negative or aggressive.

In prescriptive attribution, the question is not which causes can be attributed to the event or action, but by which ethical principles the action can be justified as positive. The ethical principles refer to "how to arrive at a judgment of the action" (cf., Table 12.1). One could wonder, however, why ethical principles are relevant for judgments at all. For example, the person from Table 12.1 could simply judge abortion to be right, because denying the woman to have the abortion would violate her right of self-determination. This view may be absolutely persuasive—until another person argues that the killing of an unborn child violates the child's right to live or the value 'sanctity of life'. Thus, the situation gets more difficult if there is more than one valuable good at stake so that trade-offs have to be made. This is the rule rather than the exception, albeit the naive ethicist frequently tries to avoid acknowledging this trade-off requirement (Tetlock 1986; Tetlock et al. 1996).

12.1.2 Ethical Positions

In the history of humankind, different argumentation patterns have emerged that serve as guidelines for dealing with complex, value-laden situations. They were

Table 12.2 A classification schema of four classical ethical positions, illustrated by exemplary statements

	Ends/Consequences	Means/Rule
Personal	Hedonism ("I try to make sure that I'm fine.")	Intuitionism ("I am sure this action is appropriate.")
General	Utilitarianism ("I believe one has to consider the consequences an action has on everyone.")	Deontology ("I believe that general principles serve as a guideline for our actions.")

Witte and Doll (1995)

identified by practical philosophy and are referred to as the "classic ethical positions." These positions provide a rationale for justifications and are (at least implicitly) used in any kind of ethical judgment. Witte and Doll (1995) reviewed the literature on moral philosophy and developed a classification scheme for ethical positions based on two dimensions.

The first facet, widely known in moral philosophy, is the division between means-oriented and ends-oriented ethical theories. The former focuses on the ethical evaluation of the process and the latter emphasizes the consequences. A second factor is the level of observation. Here, the focus can be on the individual or on society in general. Within this scheme, four classes of ethical positions can be identified: hedonism, intuitionism, utilitarianism, and deontology (see Table 12.2).

The position of hedonism holds that actions must not be enforced against the happiness of the individual; the individual's happiness should be the point of reference for ethical decisions. The striving for pleasure and conviviality had already been raised to the level of an ethical norm by Aristotle (*eudaimonia*) and Epicurus (*ataraxia*). It is important to note that following hedonism does not imply pure selfishness: Only if the individual's happiness is at the expense of others, does the hedonistic principle result in pure egoism (cf., Parfit 1984). In Table 12.1, the hedonistic principle is reflected in E_3.

When judging an action by the utilitarian principle, the points of reference are the consequences for all people who are affected. An action is better and more positive if overall outcome is for everyone, regardless of the means employed. With the utilitarian principle, even negative consequences (e.g., for single individuals) can be accepted if they are outweighed by the positive ones (e.g., for the community). Utilitarianism is closely associated with the philosophy of John Stuart Mill, and is very present in current ethical discourses (e.g. on technological impact assessment). In Table 12.1, the utilitarian principle is represented in E_2.

The means-oriented principle of deontology is often seen as the antagonist of the consequence-oriented utilitarianism. Here, the judgment of an action hinges on its congruence with moral rules, norms and values, whereas the outcome of the action is disregarded. The most famous example of deontological reasoning is the categorical imperative by Immanuel Kant: "Act only on that maxim (principle or rule) whereby you can at the same time will that it should become a universal law". In Table 12.1, the deontological principle is given in E_1.

With the fourth ethical principle, intuitionism, an action is judged on the basis of individual insight or the personal feeling that this is "simply the correct judgment." Intuitionism argues that there is no need for further arguing, but that the judgment simply should be accepted. Hence, an intuitionist argument is usually not allowed to have the status of a valid argument. Ewing (1953), however, explicated that without the principle of intuitionism all possible argumentation would be in infinite regress, since any basic assumptions could be questioned anytime. Thus, intuitionism must have its place among the principles of ethical argumentation. In Table 12.1, intuitionism is demonstrated in E_4.

In order to evaluate the viability of these categories, a questionnaire was developed, comprising 20 general justification statements (see Table 12.2 for sample items). In several studies, participants were presented a list of value-laden actions and for each were asked to indicate their agreement with the justification statements. In all studies, Varimax-rotated factor analyses revealed that the statements split into four factors, each referring to one of the four hypothesized ethical positions. Scale reliabilities ranged from $\alpha = 0.69$ to $\alpha = 0.83$ in Hackel (1995), from $\alpha = 0.61$ to $\alpha = 0.93$ in Witte and Doll (1995), and from $\alpha = 0.60$ to $\alpha = 0.79$ in Witte and Heitkamp (2005). Thus, the classification of ethical positions underlying ethical justifications received some empirical support.

12.1.3 Judgments

In the third row of Table 12.1, the first and second columns are linked to each other. In causal attribution, relating the potential causal sources (i.e., the person, the entity, or the circumstances) to the concrete situation (i.e., the action to be explained) results in different arguments, representing different options for how to explain the action. These provide a pool of possible arguments in the concrete situation, each of them having the form "action A was done, because of reason CS_i" (cf., Table 12.1).

The analogue to an argument in prescriptive attribution is a judgment. It links the concrete situation (i.e., the action) to ethical principles. This results in a pool of ethical judgments of the action to be justified, each of them having the form "action A is right (or wrong) because of ethical principle E_i" (cf., Table 12.1, third row column 3 and 4).

Of course, there are also differences between arguments and judgments. A primary example is that in contrast to an argument in causal attribution, a judgment unfolds the additional dimension of evaluation. While an argument simply links action and causal source, the linking of action and ethical principles implies an appraisal on the continuum right-vs.-wrong at the core of judgment. Thus, in prescriptive attribution the subordinate causal clause "because…" refers to why the action is evaluated as right or wrong, whereas in causal attribution it refers to why the action supposedly occurred. With respect to Heider's image of the naïve scientist, the absence of evaluation in causal attribution corresponds with the postulate of 'Wertfreiheit' in the natural sciences, which are interested in causal reasons but not

in assessments in terms of values. This, however, is the goal of the naïve ethicist and the main aim in prescriptive attribution.

However, the linking of specific actions with ethical principles and subsequent evaluations is not entirely arbitrary. As described in the section above, in causal attribution there might be not only one but many arguments that potentially explain the causal factors of an action or event. Thus, for a final explanation, people need to select the persuasive (or 'strong') arguments from the 'pool' of all potential arguments and sort out the unconvincing (i.e., the 'weak') ones. In other words, they differentiate the arguments with respect to their relevance for the explanation of the action (or their 'attribution strength', see 4th row in Table 12.1). One can say that separating weak arguments from strong arguments constitutes the 'core' of causal attribution.

Our idea is that prescriptive attribution operates in a similar way: On the basis of the four ethical principles, there may be many judgments of a given action. From this 'pool' of all possible judgments, the persuasive ones need to be selected and the unpersuasive ones to be sorted out. In other words: People need to make an evaluation of the judgment with respect to the importance of the ethical position for the action (i.e., their 'judgment strength').

12.1.4 Justification

Finally, one or more judgments are settled upon and an overall justification of the action is derived. This does not necessarily reflect solely the judgment with the highest judgment strength, but it may also be influenced by cognitive distortions or motivational biases. Consequently, judgments (and their underlying ethical principles) with lower relevance may also be incorporated. In addition to the judgment strength, as derived from an action's range of impact, we will now outline three other factors that might affect the integration of the judgments into a final justification.

First, since a justification aims to convince an addressee, the use of ethical principles is probably influenced by the perceived expectations and attitudes of the addressee. For example, if the action is socially desirable and its judgment very likely to be accepted, intuitionism might prevail in its justification. Consider a pupil who tells his friends that he skipped school on a nice summer day because he "was simply in the mood" and it "simply felt right." However, if he had to justify his absenteeism in front of his parents (who are not amused at all when hearing about it from the teacher), he would probably not make use of the hedonism and intuitionism principles but would rather try to construe a deontological ("Knowing all the content addressed in that lesson—as I did—it was okay to skip it") or a utilitarian justification ("Because I already knew all the contents, it was better for the teacher that I left the class, so that he could concentrate more on the other learners"). Note that with the shift in the mode of justification from the personal to the general focus

(utilitarianism and deontology), the perceived impact of the action is altered simultaneously from individual to interpersonal. These motivational biases in favor or against certain ethical principles can be seen as parallel to self-serving attributions in causal attribution. While in causal attribution the attribution to certain causal sources may threaten or promote self-esteem, in prescriptive attribution the use of certain ethical principles may threaten or promote the acceptance of the addressee (Keltner et al. 2006).

Second, the final justification might depend on whether the action was carried out by the justifying person herself or by someone else. A special case of the latter class of actions is if the action is not executed yet, termed 'ethical recommendation.' We suggest that our own actions are more frequently justified by ethical principles with a personal focus, while others' actions or actions that have not yet been performed are more frequently justified by principles with a general focus. A similar phenomenon in causal attribution is the so-called actor-observer difference: Here, there is a tendency to attribute other persons' actions to the person factor and not to the circumstance factor, and vice versa when explaining one's own actions.

Third, the use of ethical principles may depend on the cultural and social background of the justifying person. In collectivistic cultures, for instance, the use of ethics with a personal focus may be accepted to a lesser extent compared to individualistic cultures. More assumptions and findings on cultural determinants of prescriptive attribution are discussed later on in this chapter.

12.2 Extensions of the Prescriptive Attribution Concept for Negative and Aggressive Acts

Following public debates and political decisions in different cultures and nations, one finds actions that are not easy or that are even extremely hard to justify, because these actions are either complex or generally considered as "negative" or even "aggressive." For instance, how is it justifiable that some people have to pay more taxes than others (Witte and Mölders 2007), or how can acts of war and terrorism be justified (Halverscheid and Witte 2008)? In the course of our research we found that the concept of prescriptive attribution needed to be extended in two ways, in order to be applicable to complex and negative actions. First, in order to cover the justification of actions that relate to a particular group (e.g., taxation of the richest tenth of the population is higher than for the lowest tenth), the focus dimension of ethical principles needs further differentiation among ethics with a personal focus, a general focus, and a group specific focus (cf., Table 12.3).

The need for a second extension became evident when we examined how actions that are obviously negative and aggressive can be justified. Halverscheid and Witte (2008) found that many justificatory statements on politically motivated acts put emphasis on the violation of ethical principles by the opponent—examples are the Red Army Faction in Germany, El Qaida, and wars like the Iraq-Iran-War in the

Table 12.3 The extended prescriptive attribution model

	Ends/Consequence – oriented ethics	Means/Duty – oriented ethics
Individual level of judgment	Hedonism	Intuitionism
Group-specific level of judgment	Particular Utilitarianism	Particular Deontology
General level of judgment	Utilitarianism	Deontology

Table 12.4 Examples of indirect justification patterns

Ethical position	Justification pattern	Example
Indirect Hedonism	The well-being of a certain individual is periled by the enemy's action	"Buddenberg, the pig, allowed Grashof to be moved from the hospital to a cell when the transfer and the risk of infection in the prison were a threat to his life."
Indirect Intuitionism	The enemy's action reveals a lack of common sense	"Those who condemn these operations [9/11] have viewed the event in isolation and have failed to connect it to previous events or to the reasons behind it. Their view is blinkered and lacks either a legitimate or a rational basis."
Indirect-particulate Utilitarianism	The enemy's action poses a (potential) threat to a certain group	"We're concerned that Iraq is exploring ways of using these UAVS for missions targeting the United States."
Indirect-particulate Deontology	The enemy does not fulfill his specific duties	"We will carry out attacks against judges and state attorneys until they stop committing violations against the rights of political prisoners."
Indirect Utilitarianism	The enemy's action poses a (potential) threat to all humanity	"This enemy attacked not just our people, but all freedom-loving people everywhere in the world."
Indirect Deontology	The enemy violates norms and values regarded as universally valid	"And by the will of God Almighty, we will soon see the fall of the unbelievers' states, at whose forefront is America, the tyrant, which has destroyed all human values and transgressed all limits."

Based on Halverscheid and Witte (2008)

1980s or the second Gulf War in 1990/1991. This observation led to the assumption that actions of aggression may be *indirectly* justified by pointing at the enemy's amoral offences that have to be compensated for by taking counteractions. Due to the frequent occurrence of justifications stressing the enemy's violation of ethical principles, a model of indirect justification patterns was developed, consisting of six negative expressions analogous to the ethical positions presented above in Table 12.3. Indeed, all six indirect justifications were found in public speeches and explanations (see Table 12.4).

12.3 Empirical Measurement of the Justification

We developed two kinds of measures for assessing the ethical content of justifications, and consequently, for assessing prescriptive attributions: One is a questionnaire based on the four classic positions of ethics as four scales, measuring the subjective importance of these positions for a decision as a justification. Each of the 20 items represents a statement reflecting one ethical position that has to be rated on a five-point-scale (from 1 = not important to 5 = very important), according to its subjective relevance for justifying a given action. Representative items include: "I am concerned for my personal well-being." (hedonism); "I am sure that this is the right behavior." (intuitionism); "In my opinion, one has to consider the consequences for everyone." (utilitarianism); and "In my opinion, general values are decisive for behavior." (deontology). There is empirical evidence showing the quality of the questionnaire originally developed by Witte and Doll (1995). Since its development it has been tested repeatedly and shown to be a reliable, suitable instrument to measure ethical positions (Gollenia 1999; Hackel 1995; Maeng 1996).

The second measure of prescriptive attribution is a content analytic category system in which arguments are sorted according to their underlying ethical principles. This system also considers the quantity of these arguments . This measure is particularly suited for analyzing justifications given in speeches or written texts. So far, the content analytical system has been used both based on the original model, with four classic positions (Witte, et al. 1995; Heitkamp 2007) and based on an enriched framework, with a group specific range (Witte and Mölders 2007) and with indirect justification (Halverscheid and Witte 2007, 2008). In all these studies, the use of the classification system required a thorough training of the raters. it was applicable to all the diverse forms of written text (e.g. speeches, legal texts on taxation). Thus, the extended model with 12 (2 ethics by 3 ranges by 2 forms) categories seems to provide a good framework to capture ethical content in materials of any conceivable form.

12.4 Some Empirical Results on Determinants of Prescriptive Attributions

So far, we have described the prescriptive attribution concept formally and with respect to its operationalization. But what sort of information is used in order to select the persuasive judgments? Or, in other words, in which situations do justifications reflect one or the other ethical principle, and why? In the remainder of this chapter, we present those factors we consider to be the main determinants of prescriptive attribution, illustrated by some empirical results.

Table 12.5 Range of impact of actions and the associated preferred ethical principles in justification

		Range of impact of the action		
		Individual	Interpersonal	Social
Preferred ethical principle	Intuitionism	+	+	–
	Hedonism	+	–	–
	Utilitarianism	–	–	+
	Deontology	–	+	+

+ indicates that the condition applies, – indicates that the condition does not apply
Adapted from Witte and Doll (1995)

12.4.1 Characteristics of the Action to Be Justified

According to Witte and Doll (1995), information about the action's *range of impact* is crucial for the selection of the underlying ethical position. With regard to this range, they distinguish three types of actions. In *individual actions*, the actor and the recipient of the action are one and the same person, and no other people are affected (e.g. a man tries to help himself). *Interpersonal actions* affect both the actor him- or herself and other identifiable people (e.g., a man tries to help a friend). Finally, *social actions* affect both the actor and the general public (e.g., a man tries to help the public). The difference between interpersonal and social actions is that in the former case, the recipients are identifiable others, while in the latter they are unidentifiable. Witte and Doll (1995) made predictions about which types of actions are preferably justified by which ethical principles (see Table 12.5).

Individual actions are expected to be justified by ethical positions with a personal focus. Since individual actions only affect the actor, considering the benefit of the broader community (utilitarianism) and following universal norms, values and principles (deontology) may play a less important role. Therefore, ethical positions with a personal focus(hedonism) and arguing with personal intuition and feeling (intuitionism) appear to be adequate strategies of justification, and thus are expected to be used frequently.

In *interpersonal actions* not only the actor but also other persons are affected. Hence, these situations require the coordination of potentially differing interests. Judgments on the basis of the interests of the individual, as posited in the hedonism principle, are probably not socially approved in these situations. On the contrary, deontology may offer a valuable point of reference for the evaluation of behavior, since it provides generally accepted rules and norms for the negotiation of interests. Due to the fact that the rules and norms associated with interpersonal situations might often be well-established, internalized, and highly salient, people may also be able to judge right from wrong without much conscious effort. Thus, intuitionism is expected to be a preferred mode of justification too. Since the range of impact of interpersonal actions is rather small, a reference to the mean benefit for all might not be reasonable, so that utilitarianism is probably an uncommon mode of justification.

Social actions affect a broader community. In these situations, the mean benefit for all people, as stated in the utilitarian principle, might be an important point of reference for the evaluation of actions. Likewise, deontology might be a preferred mode of judgment for social actions, since its congruence with rules, norms and values can be understood as an indicator of whether the action is good or bad for the community. Hedonism, however, is not expected to be a frequently used principle, because orientation toward an individual's own pleasure or the pleasure of specific others might interfere with the goals of the wider community. Similarly, intuitionism is not expected to be applied to social actions, since that position is not necessarily obvious to the actor if effects of the action on the community are received well or not and judged as right or wrong. Therefore, relying solely on one's own intuition might not yield persuasive ethical judgments.

Witte and Doll (1995) presented a set of 18 different actions to 60 students from West-Germany and asked them to rate the relevance of 20 ethical justification statements for each. All actions were a priori classified according to their range of impact. As predicted, individual actions were primarily justified by the hedonism principle and, to a lesser extent, by intuitionism. Interpersonal actions were justified by intuitionism, but contrary to expectation, not by deontology. Social actions were justified with the utilitarian position, but surprisingly with hedonism and intuitionism as well. Contrary to expectations, neither in justifications of interpersonal nor of social actions did deontology play a significant role.

12.4.2 Cultural Determinants

The authors also present evidence for the assumption that the use of ethical content in justification depends on the cultural background of the justifying person. They asked a total sample of 1,300 people from East- and West-Germany about the justification of divorce, i.e., an interpersonal action. The West-German participants used the deontological principle to a much lesser extent than did participants from former East-Germany (Cohen's $d = -.51$).[1] In contrast, West-German participants applied the hedonistic principle to a much greater extent ($d = .60$). Thus, in more collectivistic cultures ethics with a general focus were used more intensively whereas in more individualistic cultures, personal positions were preferred. Similarly, Hackel (1995) showed that in justifications of individual, work-related actions, East-German participants referred to the deontological and utilitarian principles to a larger degree than did West-Germans (Cohen's $d = .81$ and $d = .48$, respectively). Further evidence for the impact of cultural background is reported by Maeng (1996). She compared Germans and South Koreans concerning their use of ethical principles in justifications of interpersonal actions and found that in the Korean sample

[1] This is a standardized difference of means. In social sciences a $d = 0.50$ is a medium effect size, $d = 0.20$ a small and $d = 0.80$ a large size.

deontological justification played a larger role and hedonistic justification a lesser role in comparison to the German sample. Similar results were obtained in unpublished data by Gollan et al. (2011). They asked Brazilian and German students to make decisions in six everyday dilemmas and to justify their decisions both in their own words and by filling in the Ethical-Positions-Questionnaire (EPQ). They found differences in the ethical content between the two cultures, which were moderated by the form of measuring. In the EPQ, Germans used ethical principles with a personal focus (hedonism, intuitionism) whereas Brazilians applied those with a general focus (utilitarianism, deontology). However, content analysis showed that when producing justification themselves, Germans preferred teleological ethics (hedonism, utilitarianism) while Brazilians preferred means-oriented ethics (intuitionism, deontology).

There is also an influence of the justification pattern by professional identity, which can be considered an independent sub-culture. Gollenia (1999) found that economists prefer hedonistic positions, and that physicians and jurists favor deontological and utilitarian positions to justify germ line therapy.

Halverscheid and Witte (2008) theoretically extended and empirically analyzed the ethical positions framework by also focusing on indirect justifications. They analyzed the content of speeches and declarations of various governmental institutions and terrorist movements in both Western and non-Western countries. Aggressive acts were primarily justified by indirect justification, i.e., by referring to an ethical transgression by the opponent. Comparing direct justification (39.85 %) with indirect justification (60.15 %), the data show a greater use of the latter practice. Furthermore, there is a large difference between the Arabian and the Western cultures in justifying war and defining terrorism: Both the direct and the indirect utilitarian argumentations seem to be typical for the Western groups, however, the negative expression of particularistic utilitarianism, (i.e., emphasizing the bad consequences of the enemy's action for a certain group), appears more often in the justifications of the Arabian parties. Likewise, the group-specific expression of negative deontology is highlighted considerably more often by the Arabian group than by the Western group (Halverscheid and Witte 2008).

In summary, there is some empirical evidence that the prescriptive attribution pattern is sizably influenced by the cultural background of the justifying person.

12.4.3 Personal Determinants

The data presented by Gollan et al. (2011) illustrate that in addition to cultural background, personal characteristics also influence the way people justify their actions or decisions. In addition to assessing ethical decisions and justifications, personal values were also examined with the Portrait Values Questionnaire, (Schwartz et al. 2001). It was found that persons self-rating high on self-enhancement values made heavier use of the hedonism principle in their justifications, whereas persons rating

high on self-transcendence relied more on utilitarianism. Both groups had a broader focus on the wider social community. Finally, as would be expected, deontological justifications were primarily used by persons with conservative values.

12.4.4 Situational Characteristics

Witte and Heitkamp (2005) show that there is also an influence of social roles when it comes to justifying an economic decision. In their study, participants were asked to first engage in a group discussion and then to decide and justify whether the production of a mobile-phone company should be transferred out of their country, from the perspective of one specific role randomly selected from the following options: external consultant, member of the supervisory board, member of management, labor union representative, employee of administration, and politician. There was a significant interaction between the roles and the importance of the ethical positions in the justification of the decision. Furthermore, the results clearly supported the expectations concerning how the roles would affect the justifications of the decisions. Thus, group discussions of participants with different roles, such as is common for ethics committees, may lead to predict a justification pattern with a narrowed rationality. The rationality could be improved by a systematic group facilitation method using the four ethical positions as a baseline, providing a guideline to discuss a problem with ethical content in a systematic way (Heitkamp 2007).

12.5 Discussion

Because humans are social beings, they constantly need to coordinate their actions and to align their behavior with the rules and norms that are defined by their social contexts. This coordination is only possible if they are able to reason and argue about why a behavior is to be judged right or wrong. In this chapter, we derived a theoretical concept of how these justifications are obtained, introduced two measures that can be used to operationalize these justifications in research, and briefly reported some empirical findings. Justifications are inferred in a way similar to reasons: The right- or wrongness is attributed to basic underlying ethical principles just as causality is attributed to underlying causes. In both kinds of attribution, there is neither a "normatively right" nor "a perfectly correct" solution. They are both subjective construals shaped and biased by quality of the action (e.g., individual, interpersonal, social), situational factors (e.g., roles), personal characteristics (e.g., preferred values) and cultural context (e.g., collectivistic vs. individualistic culture). This is why people engaging in attribution are not assumed to proceed like scientists or ethicists, but like *naïve* scientists or ethicists. In most cases, they try to come up with a solution for the required attribution that is, for them, reasonable and

convincing enough in the specific situation. This solution does not necessarily reflect the truth or ultimate morality; rather it is shaped by their personal background and (possibly self-serving) needs.

The main sources of information for construing both kinds of attributions (i.e., the "material") are characteristic of the action or event that is to be justified or explained, respectively. In causal attribution, for instance, it matters whether or not the event is seldom, or whether an action was carried out by only one or more actors. Similarly, in prescriptive attribution, it matters whether the action's scope is individual, interpersonal or social, or whether it is an action considered positive (e.g., prosocial), negative or even aggressive.

In justification, the evaluation as right or wrong mandatorily refers to a basic ethical principle: People base their ethical thinking and arguing—at least in a subtle way and often without being conscious of it—on these ethical positions that have been known in practical philosophy for thousands of years. Apparently, practical philosophy reflects what people think and vice versa. Similarly, the research process that was outlined in this chapter is also bidirectional in nature. In a first step, the prescriptive attribution concept (cf., Table 12.2) was derived on theoretical grounds, but later on it was developed further based on empirical findings (cf., Table 12.3). The prescriptive attribution concept, in its revised form, is therefore an example of an empirically informed theory of ethics.

What is the benefit of this concept for research and practice? It seems possible to operationalize the main positions of moral philosophy by developing a questionnaire and a content analytic category system. With these instruments ethical justification can be measured as prescriptive attributions in the form of rated subjective importance (questionnaire) or frequencies (content analysis). Both measures enable researchers to obtain empirical data on ethical justification and to empirically test hypotheses, for instance concerning the dependence of the justification pattern on the kind and quality of action, on culture, role, personality and mode of group discussion. For practitioners, knowing the effects of these variables may enhance understanding for the argumentation of the other party in a conflict, it may help in discussing difficult ethical decisions in a more rational way following ethical standards, and it may help circumvent conflictous or aggressive acts by an enhancement of the own ethical position. In this way, empirically informed ethics has the potential to have a substantial impact on ethical practice beyond the field of research.

Part V
Practicing Ethics in the Real World

Chapter 13
Becoming a Moral Person – Moral Development and Moral Character Education as a Result of Social Interactions

Darcia Narvaez and Daniel Lapsley

It is commonly assumed that humans do not begin life with moral character or virtue. Most documented societies through history considered infants to be unformed persons, not yet moral members of society, "humanity-in-becoming" who have "watery souls" (Fijian) (Sahlins 2008: 101–102). This person-becoming view fits well with human sciences today, as a child's development is viewed as the unfolding and co-construction of a complex dynamic system. At first, the infant is co-constructed by other complex, dynamic systems—caregivers. The personality that is formed is very much dependent on this early formation, which is largely beyond the control of the individual. However, over time, the individual takes on more choices about her or his own character development within the framework of subsequent social experience and enculturation.

13.1 Early Experience[1]

As a dynamic system, initial conditions of human development matter greatly (Churchland 1998). In fact, how one begins life may be of utmost importance to the emergence of virtue (Herdt 2008). Early experience plays a key role in the development of all body and brain systems and so it necessarily has an influence on subsequent moral functioning (Narvaez and Gleason 2013). From conception, if not before, the quality of brain and body systems are influenced by caregiver behavior, affecting such things as immune system receptors and ratios, brain transmitter quality and stress response, all of which relate to physical and mental health outcomes

[1] The focus here is on the first few years of life. Of course, there are other sensitive periods and other experiences that play roles in moral development. But the first years of life establish thresholds for physiological and psychological functioning that are difficult to change later.

D. Narvaez (✉) • D. Lapsley
Department of Psychology, University of Notre Dame, Notre Dame, IN, USA
e-mail: dnarvaez@nd.edu

M. Christen et al. (eds.), *Empirically Informed Ethics: Morality between Facts and Norms*, Library of Ethics and Applied Philosophy 32, DOI 10.1007/978-3-319-01369-5_13,

long-term (Grosjean and Tsai 2007; Davis and Sandman 2010). Perinatal and post natal experience can influence the mother-child relationship by contributing to establishing a positive highly-responsive relationship, or not (Bystrova et al. 2009; Klaus and Kennell 1976/1983). The nature of this early interaction affects the child's life outcome significantly (Felitti et al. 1998). For example, children with inconsistent or non-responsive caregivers may develop emotional dysregulation that becomes the foundation for further psychopathology such as depression, aggression, compromised social abilities and lifelong anxiety (Cole et al. 1994; Davidson et al. 2000; Henry and Wang 1998; Panksepp and Watt 2011; also see Schore 2013). Let's take one of hundreds of possible examples. When care is responsive and subverts infant stress, the child develops a good tone in the vagus nerve, which is implicated in the functioning of digestion, respiratory and cardiac systems as well as self-regulation capacities (Porges 2011). Vagal tone also influences compassionate response. Those with poor vagal development tend to be less compassionate (Eisenberg and Eggum 2008). Warm, responsive parenting is longitudinally related to the development of agreeableness, conscience and prosociality (Kochanska 2002), characteristics of adult moral exemplars (Walker and Frimer 2009). Thus, as others have pointed out (e.g., Tomkins 1965), early experience has import for moral functioning in adulthood.

Triune ethics theory (TET; Narvaez 2008) describes how early experience can influence the neurobiological underpinnings of moral functioning, identifying three moral orientations that emerge from the evolved strata of the brain (MacLean 1990) and that are shaped by early experience: Safety (reflexive self-protection), Engagement (relational attunement) and Imagination (reflective abstraction). Individuals can be influenced by the situation to adopt one of the ethics (e.g., Safety in threatening situations), but individuals can also develop a dispositional orientation toward one or another based on experience during sensitive periods.[2] When an orientation is activated, it influences perception, including which rhetoric is attractive and which action possibilities (affordances) are salient. For example, when one feels threatened, vision narrows toward actions that facilitate the reestablishment of a sense of safety (Rowe et al. 2007; Schmitz et al. 2009). When an individual uses an orientation to take socially-relevant action, trumping other values in the moment, it becomes an ethic. For the individual, acting for self-protection with aggression or withdrawal "feels" like the right and moral thing to do.

The safety ethic relies on the extrapyramidal systems, basal ganglia and lower limbic system units that function to protect the organism. These are available at birth but can be conditioned by early experience to be over or under-reactive. When children do not get what their brains and bodies evolved to expect, they develop a stress-reactive brain (Henry and Wang 1998). This leads to a more self-protective orientation to the social life and fosters a dispositional Safety Ethic for socio-moral functioning. In this case, from early "undercare" the individual loses free will,

[2] Generally speaking, sensitive periods in brain development include the first 5 years, early adolescence, early adulthood and during therapy. Although thresholds for many systems are established early, there is opportunity for change during these other sensitive times. It is not yet known how the three ethics differ in malleability during these periods.

forced by conditioning (as the brain reacts without control to situations as threatening) to rely on suboptimal emotional circuitry or even more primitive brain systems for social interaction (Narvaez 2013b). The brain/body system stays 'on alert' to some degree, keeping self-protective systems online. This activation subverts the more relaxed states that are required for positive prosocial emotions and sophisticated reasoning. Fearful vigilant states can render the Engagement and Imagination ethics mute.

According to TET, mature moral action requires control of stress reactivity and a heightening of the prosocial aspects of brain function that underlie the engagement and imagination ethics. The Engagement ethic is well developed when early care is good. Good early care matches the evolved developmental niche (EDN Narvaez and Gleason 2013), representative of social mammalian characteristics that emerged over 30 million years ago (extensive breastfeeding, constant touch, natural childbirth, prompt response to needs, multiple adult caregivers and extensive maternal social support, free play). The EDN environment fosters well-functioning physiological systems (e.g., vagus nerve mentioned above) as well as prosocial systems (e.g., higher limbic system with connections to the prefrontal cortex), hormones such as those related to bonding and attachment (e.g., oxytocin), allowing the individual to reach out and attune to others (Narvaez 2013b). For example, emotional intelligence, the ability to get along skillfully with others, is higher in children raised with a mutually-responsive-orientation with the primary caregiver (e.g., Kochanska 2002). The individual is able to adopt an intersubjective stance with others through "limbic resonance" (Lewis et al. 2000), leading to greater emotional engagement, self-regulation and skilled sociality (Schore 1994).

When care is ideal, the systems underlying the imagination ethic (e.g., prefrontal cortex systems) are also well established allowing for appropriate abstraction that does not detach from emotion and typically maintains a prosocial connection (communal imagination). Under conditions of undercare or toxic behavior during sensitive periods (e.g., neglect in early childhood, binge drinking in adolescence; Bechara 2005; Lanius et al. 2010), the imagination ethic can be damaged resulting in disassociation from emotion (detached imagination). Moreover, trauma during sensitive periods can strengthen connections to the self-protective orientation (vicious imagination), adding to a self-centered moral orientation.

Recent studies in the first author's lab show that early care practices matter for moral functioning in early childhood. She and her colleagues are studying the evolved developmental niche (EDN) for young human offspring which emerged with the social mammals over 30 million years ago with slight variation among humans (extensive, on-demand breastfeeding; nearly constant touch; responsiveness to the needs of the child; multiple adult caregivers; free play; social support; and natural childbirth). The EDN components influence health and wellbeing with known physiological mechanisms (e.g., stress reactivity; neurotransmitter development; Narvaez et al. 2013a). The first author and colleagues are showing that they matter for moral development as well. Each of these practices relates to early childhood moral development (conscience, self-regulation, empathy, cooperation). For example after controlling for education, income and general responsivity, maternal

touch patterns at age 4 months predict fewer behavior problems at age 3 (Narvaez et al. 2013b) and a higher amount of maternal positive touch in early childhood correlates with empathy at age 3 in both the USA and China (Narvaez and Gleason 2013). These findings suggest that practices representative of the ancestral human mammalian milieu may be important, above and beyond responsivity alone, for fostering sociomoral development.

But does early experience relate to moral functioning later in life? The first author has also been exploring how early experience (e.g., attachment) influences ethical identity and moral action in college students. She has developed measures of identity for each of the TET ethics based on Aquino and Reed's (2002) moral identity measure. Research (Narvaez et al. 2011a) on college students is showing that proxies for early life experience (e.g., attachment) predict the personality traits of agreeableness and openness. These personality factors predict engagement and imagination identities (i.e., preferred goals for self). Engagement identity predicts action of helping the less fortunate beyond agreeableness. Those with safety identities are more dishonest and are more likely to want to impose their values on others whereas those with engagement and imagination ethic identities are more likely to live according to their core values (e.g., buy products, choose activities), than those with a safety identity, which suggests a more integrated personality. Longitudinal studies must be done to verify the linkages.

13.2 Self-co-construction from Social Experience

As noted above, early life provides the environment for developing (or not) the well-functioning emotion systems and self-regulatory capacities that underlie social interactions and support the emergence of virtue. In fact, human developmental theory emphasizing the importance of early life in shaping moral personality (Kochanska 2002) matches up well with Aristotelian theory, providing insight into the development of moral personhood. Aristotle describes the nature of virtue in terms of habituation. By exercising or practicing virtue, individuals acquire virtue. One becomes virtuous through practicing virtue under the guidance of a mentor until one can mentor oneself (Urmson 1988). Aristotle's formulation fits well with contemporary psychological theories of learning and virtue development (Bransford et al. 1999; Hogarth 2001; Narvaez 2006).

Although the notion of habits has been controversial within psychology, new theories provide integrative approaches that avoid these problems (see Lapsley and Narvaez 2006 for a discussion). Social cognitive accounts of moral personality interpret the dispositions of habits and virtues as social cognitive units (schemas and prototypes) that emerge from and are transformed by immersion, repeated experience and guided instruction (Lapsley and Narvaez 2004). Using an apprenticeship model, Steutel and Spiecker (2004) suggest that Aristotelian habituation can best be understood as learning-by-doing that involves regular and consistent practice under the guidance of a virtuous tutor. Habits developed in this way lead to dispositional

orientations that occur automatically without reflective thought (Steutel and Spiecker 2004). Similarly, Narvaez (2005; Narvaez and Lapsley 2005) suggested that the formation of moral character reflects expertise development. According to Narvaez (2006), moral character is fostered by multiple levels of social influence including caring relationships, cultural climates, and a supportive community in a type of moral ecological context (Bronfenbrenner 1979). Within this complex set of social influences, moral character development is a matter of perfecting interactive skills (in perception, sensitivity, reasoning and judgment, focus and action). Indeed, understanding virtues as socially-mediated skills is an argument also made increasingly by virtue theorists (Jacobson 2005; Stichter 2007a, b). Coached practice of a skill leads to increasing intuitive responsiveness that permits rapid, automatic judgments and behavioral responses to relevant contingencies (Bartsch and Wright 2005; Dreyfus and Dreyfus 1991; Narvaez 2010a). The automaticity of social skills can account for the tacit qualities often associated with Aristotelian "habits". These habits correspond to social cognitive schemas or behavioral components whose frequent activation becomes overlearned to the point of chronic automaticity (Lapsley and Hill 2008; Narvaez et al. 2006).

If we move beyond early life, childhood involves additional social experiences that influence the development of the moral person. What are the developmental sources of moral habits and chronicity? Lapsley and Narvaez (2004) suggest that moral chronicity is built on the foundation of generalized event representations that comprise early socio-personal development (Thompson 1998), the "basic building blocks of cognitive development" (Nelson and Gruendel 1981: 131), internalized working models of what one can expect of social experience. So for example, children taught to pay attention to their impact on others ('how does your sister feel after you took her toy?', 'how can we share the single cookie?') learn to frame their social lives with this sort of awareness. Prototypic working models are progressively elaborated in the early conversations with caregivers who help children review, structure and consolidate memories in script-like fashion (Fivush et al. 1992).

Another type of internal working model that guides behavior is proposed by attachment theory (Bowlby 1988). This approach integrates the emotional aspects of social experiences with caregivers. Attachment has to do with implicit, 'felt,' experience more so than conscious explanation of experience. Triune Ethics Theory, described earlier, brings these ideas into the moral domain, emphasizing the underlying neurobiology of moral internal working models. The child internalizes emotional memories as part of the self. The topics and emotional frameworks the caregiver uses in helping children organize their lives become routine, habitual and automatic the longer they are practiced. The socio-emotional patterns in routine relationships become expectations for the social life. When the parent references norms, standards and values, they encourage the formation of chronically accessible social cognitive schemas (e.g., 'What should you say when you receive a gift?'; Lapsley and Narvaez 2004). As the self develops, the child integrates these patterns of experience into autobiographical memory, facilitated by parental conversation, interrogation, emphasis and focus. The moral self is part of this package of

experience, deeply influenced by the parents and other caring relationships. Reflection on past success through the rehearsal and review of events, reactions, and so forth, is important for future moral action because the implicit understanding of certain brain systems is integrated with the more conscious conceptualization of other brain systems. Parents help children reflect and build moral representations of their lives.

Becoming a moral person is a lifelong enterprise. In the USA, it is much more difficult to be virtuous that in the contexts of most of human genus history (prior to agriculture) because of poor childrearing in relation to basic needs, multiple pressures towards vicious behavior (e.g., self-centered consumption), disregard and enslavement of animals and the natural world, violent media infused with humor (which has a greater influence on children to imitate), shifting or unclear goals, roles and duties (for detailed discussion, see Narvaez 2013a, b).To discuss the full flower of moral personality, we must move beyond early childhood, and even the college years. We must examine the nature of moral maturity (Narvaez 2010b).

13.3 Mature Moral Functioning

When we think of mature functioning, we often think of advanced moral reasoning and the ability to be impartial in making decisions. But we know now that emotions are essential for good cognition (Greenspan and Shanker 2004) and without them, decisions are often faulty (Damasio 1999). We also know that reasoning is only weakly linked to action (Thoma 1994).

We can take another tack in examining mature moral functioning by using the framework of expertise, which involves implicit and explicit understanding, integrating intuition and deliberation. Moral exemplars often exhibit the characteristics of experts. Although the knowledge and skill advantages the expert has are still being uncovered, experts are distinguished by certain characteristics. They have many more or less automated responses including perception that allow them to see patterns and opportunities that novices miss (Chase and Simon 1973; Chi et al. 1988). They are better at selecting appropriate schemas and having them readily available for action (Spiro 1980). They have rehearsed action responses to high levels of automaticity (Ericsson and Smith 1991). In other words, expertise is a combination of perceptual attunement, complex understanding, motivation for excellence and effectivities that provide the capacity to take action given the affordances of the situation.

What kind of knowledge is expert knowledge? Schooling and literacy has focused us so much on reasoning and conscious thinking (in contrast to holistic, creative thinking), that we have begun to emphasize them in our childrearing practices too. We often forget that most of our mind and actions proceed nonverbally without deliberation or conscious awareness and that emotions are foundational to adaptive functioning (Bargh 1989; Panksepp 1998). Intuition has become a large focus of recent psychological research, replacing a focus on conscious processing,

as psychology has come to appreciate how much implicit processes govern human action (Reber 1993). One critical facet of this research is the necessity to distinguish between naïve and well-education intuition (Narvaez 2010b). Expertise blends both deliberation and well-educated intuition together (Hogarth 2001). This is true in the case of moral functioning as well (Narvaez 2010b).

Although reasoning has often been the focus of moral psychological development and moral maturity (e.g., Rest et al. 1999), if we expand on Rest's (Rest 1983; Narvaez and Rest 1995) component model of moral behavior, we can see that there may be other aspects to consider in terms of mature morality. The expanded model identifies five sets of processes: reasoning and judgment, sensitivity, motivation, action skills, and perception. Because of its dominance in research, we start with moral reasoning.

Moral reasoning: For decades, research in moral development focused on the naturalistic development of moral reasoning. Moral reasoning sophistication develops with age and education. In adults, moral reasoning sophistication is related to real-life behaviors such as democratic teaching style, professional clinical behavior, attitudes towards human rights and at the same time is distinguishable from intelligence, political attitudes and religion (Rest et al. 1999; Narvaez et al. 1999a; Thoma et al. 1999, 2009). However, contrary to Piaget and Kohlberg's suppositions, everyday experience is not sufficient to reach the highest levels of moral reasoning development. Narvaez and Gleason (2007) proposed that moral judgment retains characteristics of being both a developmental variable (which everyone develops to some degree) and a domain variable (which requires extensive, deliberate study). As such it bears resemblance to other domains. For example, virtually everyone is familiar with some aspect of music or even skilled in some fashion—as with singing, and yet musical expertise requires specific and prolonged practice beyond everyday familiarity. Sloboda (1991) contrasts the tacit musical expertise of novices, a type of receptive, recognition-based expertise, with the explicit or productive expertise of expert musicians.

These two forms also are evident in moral judgment. Moral judgment skills are in use daily for everyone and provide a base of tacit knowledge (see Narvaez and Bock 2002; Rest et al. 1999). Some are moral judgment novices (with less stimulating experience), who have fewer conceptual strategies for solving social problems, adopting simpler, more actor-centered options or ones that maintain social norms. Other lay people have receptive moral judgment expertise which enables thinking about organizing society-wide cooperation according to moral, impartial, public, reciprocal, criticizable principles (see Rest et al. 1999, for detailed discussion). Productive moral judgment expertise requires prolonged and focused experience in a particular domain, leading to for example, original contributions to philosophy or federal court opinion, or community problem solving. In short, everyday living does not usually bring about productive expertise in moral judgment. Focused deliberative experience is required such as graduate study in moral philosophy, community leadership or social activism. Moral decision making with minimal experience will be largely nonverbal and receptive expertise whereas

productive expertise involves theoretical or explicit knowledge that allows experts to make arguments and explain their reasoning.

Expertise in moral judgment may not be sufficient for establishing moral expertise. Why is that? It is because the "heart" may not be involved. Moral judgment can occur in an emotional vacuum. High functioning autistic individuals may receive high scores because they are superior memorizers of rules and systems, but they fail in everyday virtue, the moment-to-moment social and moral functioning that requires exquisite emotional intelligence. Such a detached morality cannot be a demonstration of moral excellence. As Aristotle pointed out, a virtuous action is one that is performed in the right way, with the right feelings and for the right reasons. Also from the studies of nominated exemplars, reasoning like a philosopher is not a necessary condition for moral exemplarity (Colby and Damon 1992). However, if one thinks of famous moral leaders like Gandhi and Martin Luther King, Jr., superior moral reasoning was part of the package. In short, formal education and reflection with regard to moral judgment is neither necessary nor sufficient to develop moral expertise, but it has the potential to be helpful.

Reasoning and judgment are often studied as if they are intellectual capacities. Logical analysis about the right thing to do detaches one from the present situation and works mostly in offline situations, when one is not in the middle of being emotionally focused on completing goals. However, moral experts keep an emotional focus as they apply automatized reasoning and judgment when involved in domain problem solving. For example, Monroe (1994) notes that rescuers of Jews in World War II made statements like, 'what else could I do—they were human beings in need' whereas non-rescuers were more likely to say things like 'what could I do—I was one person against the Nazis.' The habitually prosocial individuals acted spontaneously. Moral behavior involves perception, interpretation, motivation and action skills as well as judgment. These can all be integrated into classroom academic instruction, using a novice-to-expert pedagogy (Narvaez 2006, 2009; Narvaez and Bock 2009; Narvaez and Endicott 2009; Narvaez and Lies 2009).

Moral sensitivity: Moral sensitivity involves noticing the needs of others, having empathy for them, and generally noticing the need for moral action. These aspects of sensitivity depend largely on social emotions and right-brain capacities: seeing the big picture; linking the situation to prosocial emotions; determining one's potential role (Schore 1994; McGilchrist 2009). Sensitivity also includes interpretive capacities or the ability to foresee the consequences of particular courses of action or inaction in terms of concrete outcomes and reactions from others. For example, professional education programs that sensitize students to their moral responsibilities to patients and clients demonstrate increased awareness of issues, options and consequences (Rest and Narvaez 1994). But moral functioning also includes motivation, caring about the outcomes.

Moral Motivation or Focus: The third component of motivation or focus has a habitual component but also a 'here-and-now' component. Motivation, identity and personality are central characteristics of those who take moral action for others (Lapsley and Narvaez 2004; Narvaez and Lapsley 2009). For example,

agreeableness and conscientiousness are both characteristic of exemplars (Walker and Frimer 2009). Those with explicit moral identity goals are more likely to spontaneously process social information with moral categories. For example, Narvaez and Lapsley (Narvaez et al. 2006) used a primacy-of-output measure of moral identity where the initial response to a question (e.g., name characteristics of people you like) indicates a chronically used category for social information processing. Those with higher or lower moral identity were compared on two tasks. In one task, participants read sentences about a person in a role (e.g., "The plumber always meets his obligations and keeps his word") and were later primed to remember the sentences with a word either representing the role ("pipes") or the disposition ("responsible"). Half the sentences represented non-virtuous dispositions. Those with higher scores on moral identity were more likely to make a spontaneous trait inference when primed with dispositional cues than with role cues.

In their second study (Narvaez et al. 2006), participants read stories about characters who did or did not help. Those with moral identities (moral chronics) were quicker responding to probes that represented negative evaluations of story characters who did not help when requested (e.g., "selfish"). These studies showed that chronic moral identity affects social information processing.

In terms of 'here-and-now,' mature moral actors have greater sensitivity to prioritizing moral action at any given moment. But they understand that they may miss opportunities, just as a person helping one homeless person may miss the opportunity to help another. Mature moral actors develop habits that facilitate their moral actions. Habituated empathic concern is one such habit and can entail structured practices such as automatic bank account deductions to the food bank. Mature moral actors realize that when the time comes they may be otherwise distracted and build in safeguards for moral action (Trout 2009).

Moral Action: The last component comprises implementation and follow-through on a selected moral action. It requires extensive practice but is deeply linked to the other components. Perception of possible actions is an aspect of sensitivity to the situation in terms of affordances. Perception is shaped by experience and influences which stimuli reach higher order centers (Neisser 1976). Perception and action judgments are integrated into effectivities matching personal capacity built from extensive domain-relevant experience, to the possible actions (affordances) in the situation (Feltovich et al. 1997; Shaw et al. 1982). Along with concern for others, moral exemplars display more effectivities for particular actions, agency or self-efficacy in their domain that distinguishes them from others (Frimer and Walker 2009; Monroe 1994; Walker and Frimer 2009).

Moral Perception: Moral perception initiates the processes that lead to moral action: What does the individual notice or not notice? However, moral perception is influenced by other components. For example, those with more sophisticated moral judgment skills (as measured by the Defining Issues Test), are better at noticing and recalling sophisticated moral reasoning in stories (Narvaez 1998, 1999, 2001; Narvaez and Gleason 2007). They are better at discerning the intended moral theme in a story (Narvaez et al. 1999b). Similarly, those with moral identities are more

aware of moral violations and moral behavior in story characters (Narvaez et al. 2006). Triune Ethics Theory suggests that moral perception shifts depending on the mindset that is active. In a recent study, the first author's lab finds that those with high engagement ethical identities see the photo of a crying baby as closer to them than those with low engagement identities, suggesting that identity affects visual perception (Narvaez et al. 2011b).

Mature moral expertise is found ultimately in the integration of component skills (perception/sensitivity-judgment-focus-action links). In a way we are moving with a trend among psychologists and philosophers to view virtue and moral personality as sets of implicit skills. Along with moral developmental psychologists (Narvaez and Lapsley 2005), moral philosophers recently have been shifting to a view of moral virtue as comprised of skills that can be honed to high levels of expertise (Annas 2011; Zagzebski 2006). How does this expertise develop?

13.4 Expertise Development Through Relational Coaching and Community Immersion

How do individuals become experts? Through guided immersion in informative environments ("kind" environments, Hogarth 2001). Their training is focused and extensive, taking about 10 years or 10,000 h of practice (Chase and Simon 1973). Expertise development can be sped up with a mentor who points out the pitfalls of particular actions and the benefits of others, a benefit not available to a novice who is testing out actions and problem solving alone. The mentor helps to coordinate deliberative understanding with intuition development. When experts with formal training are learning to solve problems in their domain, they usually do so in the context of explicit theory and explanation. Thus, early on they are able to explain the actions they take. Gradually, however, with practice and experience, their decision making processes become more automatic as well. In fact, most experts become unable to explain their decision making processes (e.g., Kihlstrom et al. 1996).

It must be pointed out that *moral* expertise is socially grounded and has its roots in early experience. With a skills focus, we resurrect a view that has been marginalized in contemporary developmental science, that the primary parental influence on morality begins in early life—in the welcoming physiology of the mother, the experience of relationship with caregivers, and in the developing implicit understandings of the meaning and effect of emotions, relationships and reciprocity. Unlike other forms of expertise, moral functioning is intended for the social life so for optimal development it requires immersion early and often in the social life of the community where virtue is applied. Unlike engineering or medical diagnosis, the moral expert must have well-functioning social skills that underlie moral capacities. For example, in older normal children, it is apparent that reflective thinking is grounded in "lived emotional experience." Those with more adaptive emotional intersubjectivity with caregivers in early life are better able to think out problems, and demonstrate greater social skills, moral reasoning, and intelligence (Greenspan and Shanker 2004: 233).

13.5 Mature Moral Functioning Day to Day

Higher order deliberation is always situated in the interface of person and context. John Dewey offers fresh insight into the nature of ethical functioning as imagination (1930/1984). He "replaces obsolete notions of perspective-free rationality" and "sedimented moral criteria" with "flexible, rule-sensitive situational inquiry" which includes "moral perceptiveness, creativity, expressiveness, and skill" (Dewey, quoted by Fesmire 2003: 5). Instead of the "traditional assumption that reasonings and actions can be measured by an ahistorical standard," the pragmatists William James and Dewey emphasized "reason's ineliminatively temporal, aesthetic, evolving, embodied, practical, and contextual character;" "rejecting both foundationalism and subjectivism, the classical pragmatists transferred the burdens of reflective life to situated, emotionally engaged intelligence" (Fesmire 2003: 52).

Pragmatist ethics like Dewey's rejects rigid abstractions and emphasizes flexible responses to "the ordinary life-experiences of inherently social, embodied, and historically situated beings" who face hourly encounters "too unique" for classification by a sedimented rule or principle or intuition (although they can offer some guidance); "situations do not come in duplicates" (Fesmire 2003: 59; see Dewey 1922/2000: 167–168). Generalizations, whether applied in medicine or the ethical life, are quackery and they gag intelligence (Dewey 1929/1984: 221). "If morality were reducible to following rules or codes, high-functioning autism would be the moral ideal" (Fesmire 2003: 72). Mental operations involving conceptual systems, inference, meaning and language rely on a cognitive unconscious developed from an embodied mind's sensorimotor experience (Lakoff and Johnson 1999). In fact, companionship, responsive care in early life leads to greater intelligence, imagination and moral capacities (Greenspan and Shanker 2004).

On a day to day level, the work of moral functioning is to coordinate reasoning, facts, intuitions, reflection on past success, current goals, affordances (and multiple other aspects that impinge on our behavior) in the situation. Deliberative reasoning and intuition work hand in hand. The trick is to know when to trust intuition and when to deliberate. Both systems are goal driven but it is impossible to deliberate on many actions/decisions so one must make sure that intuitions are appropriate. Well-educated intuitions that develop from experience emerge from complex and sophisticated understanding whereas naïve intuitions that arise with no experience can often be misleading (Hogarth 2001). Misapplied intuitions are context-specific principles that are over-generalized to apply in other situations (Baron 1998). Intuition is not precise but approximate, so its errors are usually slight. On the other hand, although the deliberative system can be more precise, its errors are huge and damaging (Hogarth 2001). Further, deliberating on intuitive process can result in less optimal performance (Beilcock and Carr 2001).

Intuition and reasoning are both susceptible to "truthiness" (it feels right so it must be right) and require deliberative supervision of their accuracy within an open context (Narvaez 2010a). To counteract truthiness, Hogarth (2001) recommends that individuals take an hypothesis-testing approach—test, verify, get peer review.

Generally, an expert deliberator is able to attend to the first impression and then confirm it with deliberative steps of examination to verify intuition.

13.6 Conclusion

Generally speaking, childrearing practices have extensive and deep effects on the psychological and biological foundations in the child's body and brain (Narvaez 2013b). These occur in large part from experience during sensitive periods and affect moral capacities (Narvaez 2008, 2013b). Thus, caregivers, usually parents, have a great deal of influence on moral development. Over the life course, moral knowledge shifts from tacit to explicit (moral judgment expertise) yet at the same time moral expertise generally becomes more automatic (spontaneous action). Individuals have a say in who they become by selecting environments and activities that foster particular intuitions and expectancies. Greater capacities in moral sensitivity, judgment, motivation and action increase with focused practice, whether through post-baccalaureate education or community-based experience. When combined with action capacities and efficacy, communal imagination (which builds on prosocial capabilities) represents humanity's highest moral capacities. The development of such capacities requires a supportive social environment.

Chapter 14
Ethical Leadership – How to Integrate Empirical and Ethical Aspects for Promoting Moral Decision Making in Business Practice

Markus Huppenbauer and Carmen Tanner

14.1 Introduction

Corporate ethical scandals, the financial and economic crisis of the past several years, and examples of misuse of power by prominent leaders have shocked the business world. They have not only called into question the role of the economic system design at large, but have also raised questions about the role of leaders in influencing ethics and ethical behavior in organizations. It is hardly controversial to state that aspects of management and leadership are crucial in determining the strategical direction and daily operations of an organization. Few would doubt that leaders are key figures in shaping ethical conduct. However, little is known about what constitutes ethical leadership. What are the relevant competencies leaders should acquire? The behavior of leaders and the extent to which they behave with moral integrity has also become a topic of high interest in the media and public discussion. Yet, many people believe that ethical leadership is simply a matter of having good character or having the "right values. "Although character and values are obviously important, the concept of ethical leadership is far more complex than those factors and there is little consensus on what precisely characterizes ethical leadership.

The field of ethical leadership can be divided roughly into two realms of inquiry. Psychology (and economics) is concerned with descriptive (or empirical) ethics as

M. Huppenbauer (✉)
University Priority Program Ethics, University of Zurich,
Zollikerstrasse 117, 8008 Zurich, Switzerland
e-mail: huppenbauer@ethik.uzh.ch

C. Tanner
Department of Banking and Finance, Center for Responsibility in Finance,
University of Zurich, Plattenstrasse 32, 8032 Zurich, Switzerland
e-mail: carmen.tanner@bf.uzh.ch

M. Christen et al. (eds.), *Empirically Informed Ethics: Morality between Facts and Norms*, Library of Ethics and Applied Philosophy 32, DOI 10.1007/978-3-319-01369-5_14,

to how individuals "do" behave, while ethics is concerned with the normative implications of leadership and the question as to how individuals and organizations "should" behave. We argue that both empirical and normative approaches are important in the research and discussion of ethical leadership. Hence, this chapter explores the possibilities for cross-fertilization between psychology and ethics.

The purpose of this chapter is threefold. Our first goal is to highlight some typical features of the empirical and normative approaches to exploring ethical decision-making and behavior, and to sketch out how ethics and psychology can learn and benefit from each other. We argue that empirical research without normative reflection is "blind". Since empirical leadership research usually has normative implications, normative reflection is necessary to identify and understand those implications. On the other hand, normative reflection is "empty" without empirical insights. We believe that studying ethical leadership using an interdisciplinary approach helps to advance our knowledge of what constitutes or should constitute ethical leadership and how it can be promoted. Second, based on previous empirical research and our own perspectives, we wish to shed light on some important components and competencies of ethical leaders. Even though most scholars agree that all forms of leadership should be based on some ethical foundations (Bass and Steidlmeier 1999; Kanungo 2001), discussions still differ in regard to what should be expected from an ethical leader. While discussing important characteristics of ethical leaders, we also aim to discover some unresolved key questions about ethical leadership. We argue that these questions reflect important points of intersection between empirical and normative approaches, and points where it appears beneficial that normative reflection comes in. These questions and their implications for the practice of moral behavior and leadership will be discussed in the final section.

14.2 The Relation Between Empirical and Normative Research

Both ethicists and psychologists are concerned with identifying the qualities of moral behavior, yet they pursue distinctive goals. Ethicists usually evaluate and justify the quality of actions through reflective deliberations (Singer 2000) referring to abstract standards of ethical conduct and a moral point of view. Their focus is on normative goals and the question of what ought to be done. In contrast, psychologists examine people's beliefs, values and actual behaviors in specific contexts, and test assumptions about the mechanisms involved in ethical decision-making and behavior through empirical research. Their focus is on descriptive goals which aim to discover what actually is. Despite these differences, calls for a dialog and a closer relation between the empirical and normative approaches in moral research and business ethics are often heard (e.g., Singer 2000; Weaver and Treviño 1994; Waterman 1988). Our goal in this portion of the chapter is to contribute to this

empirical-normative dialog in the context of ethics and ethical leadership by emphasizing how the two can be mutually beneficial.

One problem of empirical research in the realm of ethics is that its concepts of moral values, norms or behavior typically rest upon consensual beliefs and common views of morality. This is also the case in leadership research. As such, the research concepts represent "social constructions that reflect the value and paradigms of leadership at a particular time and place" (Ciulla 2006: 21). Though empirical studies often build upon prior interviews with experts and practitioners, the concepts implemented in experiments and surveys are based on characteristics people consider to be ethical in leaders. However, generating knowledge about what people claim to be moral does not tell us if these claims are normatively appropriate. In other words, providing descriptions about what *is*, does not automatically reveal what *should be* (Waterman 1988; Fraedrich et al. 2011: 240). When, for example, Kohlberg claims that the highest stage of moral thinking is expressed in terms of a deontological moral perspective, this description shifts into prescription (cf., Miner and Petocz 2003: 15). In philosophy this slide is considered problematic and is called the "naturalistic fallacy" (Dunfee and Donaldson 2002: 41). Interdisciplinary collaboration is therefore important to psychology since it helps to take into account the normative implications of the concepts used, and to reduce the risk of merely deeming common sense to be sufficient for deciding which standards and behaviors are ethically adequate. Moreover, as Miner and Petocz (2003) argue, any psychological investigation must acknowledge the importance of meta-ethical positions and specific moral theories for ethical decision-making and then "consider how they might affect the processes and outcomes of decision-making" (p. 14).

On the other hand, in examining which processes and factors determine moral judgment and behavior, psychology has provided solid insights about human functioning which should also be essential for understanding normative ethics. For instance, one insight, for which there is converging empirical evidence, is that moral judgment and decision-making are often based on automatic intuitive reactions rather than reflective deliberations (Haidt 2001). Studies have shown that externally induced or even hypnotically induced gut feelings (such as flashes of disgust) do causally affect moral judgments, supporting the view that such gut feelings or moral intuitions serve as information when evaluating moral transgressions (e.g., Wheatley and Haidt 2005). There is, in addition, much evidence that the nature of automatic-intuitive or more deliberative-reflective processing is highly contingent on personal and situational factors (e.g., Fazio 1990). Overall, empirical research has made a valuable contribution to normative ethics and encouraged further discussions by forcing acknowledgement of the role of moral intuitions and automatic processes in ethical decision-making (Kennett and Fine 2009; Treviño 2009).

Should normative ethics account for these empirical findings about what determines moral judgment and decision-making? In the context of business ethics, empirical sciences provide information about actual problems and conflicts with which business agents contend. They generate knowledge about individual differences in judgment, behavior, values and attitudes based upon the people who make decisions. This is important for ethics for two reasons. First, without this empirical

"material" normative ethics would lack a sound basis for its reflection. Hence, it is an important task for ethics to clarify and formulate systematically what people think (Miller 2008: 93f.). Second, normative approaches without certain knowledge of people's specific beliefs and values, which serve to guide their judgments and behaviors, risk being irrelevant to any attempted application or implementation in daily life and professional settings. In either case, normative ethics proceeds in a critical way, asking whether or not people's given moral judgments and decision-making are right from a moral point of view. This form of critical analysis may, of course, end in doubts about the legitimacy of empirically ascertained moral judgments and decisions.

In sum, we believe that fostering a empirical-normative dialog is beneficial for advancing moral theory and research. Empirical research is blind without normative reflection since its concepts fail the critical analysis of whether they are normatively appropriate, meaning whether the norms and values people use to guide their judgments and actions are indeed right. In turn, normative reflection is empty without empirical knowledge and risks being of little practical value without acknowledging the relevance of what is, and why it is. To generate an empirical-normative dialog, the purpose of the next section is to identify some intersections within ethical leadership research where the bringing together of "is" and "ought" issues may be crucial and useful. To this end, we start with a short overview of the empirical literature on ethical leadership and then define some core features of ethical leaders. In so doing, our goal is to identify those (or at least some) key areas of ethical decision-making and ethical leadership research where it may be important to employ normative reflection.

14.3 What Are the Characteristics of Ethical Leadership?

14.3.1 Empirical Research on Ethical Leadership

Understanding the constituents of ethical leadership has captured the interest of researchers and practitioners alike. In particular, one large empirical research line has looked to develop frameworks explaining the process of ethical decision-making (for a comprehensive overview, see Treviño et al. 2006). While this research has made important contributions to understanding and predicting ethical decision-making by describing how individuals actually think and act when faced with ethical situations, it says little about what the essential characteristics of ethical leaders are. In this regard, leadership research and, more recently, research into moral intelligence (see also Chap. 7 of Tanner and Christen, in this volume) seek to identify and develop moral competences of ethical leadership and conduct.

As to leadership research, various approaches have emerged over the past decades. Although all of them tap into ethical aspects in some way, such dimensions actually play a relatively implicit or indirect role. For instance, embedded in the

charismatic or transformational leadership paradigm is the idea that "transforming" leaders are those who feel responsible for achieving the good for their organization and society, and who inspire followers to elevate their levels of motivation and morality (Bass and Steidlmeier 1999). Conceptualizations of authentic leadership typically assert that (moral) authenticity is achieved when individuals act in concert with an internalized moral perspective (e.g., Gardner et al. 2005). Still other models have highlighted either the importance of a leader's ethical values that guide choices and behaviors (e.g., Resick et al. 2006; Russell 2001), or the role of a leader's moral virtues (e.g. Manz et al. 2008; Solomon 2003).

With an explicit focus on ethical leadership, seminal work has been performed by Brown and colleagues (Brown et al. 2005; Brown and Treviño 2006). They define ethical leadership as "the demonstration of normatively appropriate conduct through personal actions and interpersonal relationships, and the promotion of such conduct to followers through two-way communication and decision-making" (Brown et al. 2005: 120). As to this conceptualization, ethical leadership involves promoting normatively appropriate conduct through role modeling and interpersonal relationships. The authors also suggest some ethical ideals that they deem to be (normatively) right: ethical leaders should be perceived as trustworthy, fair and concerned about others; they should set clear ethical standards and use rewards and punishments to promote ethical conduct. Consistent with these suggestions, Brown and colleagues have built an Ethical Leadership Scale (ELS) designed to assess whether a leader can be characterized as fair, trustworthy, or caring, and whether he or she makes an effort to communicate or demonstrate ethical behaviors (Brown et al. 2005).

In a similar vein, one author of this chapter and her colleagues have taken an action-based approach, developing an Ethical Leadership Behavior Scale (ELBS) (Tanner et al. 2010) that is based on specific behaviors reflecting concrete manifestations of ethical values (such as fairness, respect) across occasions and situational challenges. This approach shares with others the idea that moral norms and values are essential in guiding and promoting ethical conduct, but remains distinct from them by calling for more focused attention on whether and how moral values are reflected in behavioral patterns. It is commonly known that values and good intentions are not always implemented in actions. Of course, there may be many good reasons why leaders with moral intentions may choose not to act ethically, including that of avoiding unpopularity or preserving their own career (May et al. 2003). Tanner and colleagues (2010) and other scholars (Ciulla 1999) therefore emphasize the importance of leaders acting upon moral standards and values consistently, on a regular basis, and despite potentially unpleasant consequences, in order to earn the attribute of an "ethical leader."

With the concept of moral intelligence, a quite different approach to moral or ethical leadership has recently emerged. In the past, researcher and practitioners alike acknowledged that beyond cognitive skills, emotional intelligence and social intelligence represent additional advantageous capabilities (e.g., Goleman 1995; Salovey and Mayer 1990). In the aftermath of recent business scandals, researchers and practitioners alike have now started to turn their attention more thoroughly to

the moral competencies that business leaders should have or acquire. With an explicit emphasis on moral skills, Lennick and Kiel (2005) introduced the term moral intelligence to capture a new facet of intelligence. Tanner and Christen (Chap. 7 in this volume) define moral intelligence as the individual's capacity to process and manage moral problems. Current research and discussion is engaged in identifying and assessing the key elements and abilities of moral intelligence (Lennick and Kiel 2005; Martin and Austin 2010; Narvaez 2005; Chap. 7 by Tanner and Christen, in this volume). Scholars working in this area or the domain of moral expertise commonly assert that individuals need multiple abilities, but their approaches differ in terms of which skills and subskills are crucial. An elaborated model by Narvaez (2005), for instance, posits that moral experts need skills in ethical sensitivity, moral judgment, moral motivation and moral action. In general, we believe that the approach of moral intelligence with its emphasis on various moral skills can expand our understanding of ethical leadership in useful ways.

In the following section, we wish to highlight a few core characteristics of ethical leadership, building upon the research and development of leadership and morality. Our goal is also to discover points of intersection between empirical and normative approaches by identifying some key questions about ethical leadership that demand normative reflection. In this chapter we will focus on just three such points of intersection which we deem to be highly relevant (of course, we do not claim to be exhaustive).

14.3.2 Defining Ethical Leadership

We will structure the elements of ethical leadership using two categories which Treviño and colleagues (2000) termed the aspect of the "moral person" and the "moral manager." According to these authors, moral person refers to traits, characteristics and motivations of leaders. Yet, as Treviño et al. have emphasized, ethical leaders are not just moral persons, they are also moral managers in that they "lead" and influence followers to develop ethical conduct. This dimension of moral manager represents the leader's efforts to influence the ethical or unethical behaviors of followers. In what follows we focus on psychological literature but of course similar definitions are presented by business ethicists in philosophy (cf., Bowie 1999; Price 2008; Solomon 2009). With regard to aspects of the ethical person, we deem the following facets to be fundamental.

Committed to ethical values: Psychologists usually assert that values, typically defined as stable beliefs about desirable states or conducts of behaviors (Schwartz 1992), are standards that serve to judge and justify actions and have the potential to energize and regulate behavior (e.g. Verplanken and Holland 2002). Since standards and values guide choices and behaviors, ethical values appear to be at the root of ethical leadership (Lord and Brown 2001). Schmidt and Posner (1982) therefore asserted that managerial values are the "silent power" in personal and organizational life. Lennick and Kiel (2005) have emphasized that ethical leaders distinguish

themselves from other leaders in that their thoughts, decisions and actions appear to be guided by moral principles and values. Lennick and Kiel (2005) have used the metaphor of the "moral compass" to refer to this set of moral standards, values and beliefs which serve as a reference in all matters of right and wrong. Although having a moral compass is relevant, it is not sufficient to initiate action. A leader may sometimes have quite clear ideas about what should be done, but he or she may then lack the motivation to pursue it. Literature suggests that commitment to ethical values is such a crucial motivational source that it leads individuals to prioritize moral goals over other ones and to strive for desirable moral ends (Narvaez 2005; Rest 1986; Chap. 7 by Tanner and Christen, in this volume).

Given these findings, ethical leaders are expected to be committed to ethical values that serve to guide their thinking, decisions and actions. We believe that this aspect is the first intersection at which normative reflection should occur. More specifically, asserting that ethical values play a crucial role leads to the ultimate question: which values can we consider essential? Which moral principles should leaders convey? These are normative questions that cannot sufficiently be resolved on an empirical basis by common sense and consensual beliefs about what people consider to be relevant. We will discuss this topic more thoroughly in Sect. 14.4.

Endowed with ethical competencies: Here, we refer to an additional set of personal moral competencies that are likely to facilitate ethical leadership. First, we expect leaders to be ethically sensitive, recognizing and identifying ethical issues when they arise in practice. Individuals may not always be aware that they are facing an ethical issue. As Treviño & Brown accurately pictured, decisions rarely arrive with waving red flags announcing that they are ethical issues (2004: 70). The relevance of this point is obvious: if leaders do not recognize the ethical nature of a problem, no moral judgment or decision-making process is initiated (Narvaez 2005; Rest 1986; Bleisch and Huppenbauer 2011; Chap. 7 by Tanner and Christen, in this volume). We therefore consider moral sensitivity to be another key feature of ethical leadership.

Second, ethical leaders need problem-solving capabilities. Once an ethical problem has been identified, the next challenge consists of finding viable ways to cope with it. Drawing on prior work and current research, effective problem-solving entails reasoning skills. From a moral philosophical perspective, reasoning demands the capability to critically and impartially reflect on ethical dilemmas and to give good reasons and proper justifications for possible solutions (Maak and Ulrich 2007: 383ff, 480ff.). Leaders are frequently required to justify their decisions not only within the organization, but also towards stakeholders and society at large (e.g. Freemann et al. 2010). While moral psychology has intensively studied the development from childhood to adulthood of moral reasoning processes since Kohlberg (1984) and the ways in which individuals think about ethical dilemmas and justify their decisions, normative ethics teaches the application of multiple frameworks as method of choice when faced with dilemmas where values conflict. Yet, as noted, ethical problems are often rather complex and confront individuals with great uncertainty regarding possible alternatives and consequences, competing values and incompatible courses of actions, pressures from outside, etc. Individuals have to

cope with emotional stress, especially when strong beliefs or convictions are at risk or when decisions have threatening implications (e.g., employees may be harmed) (Hanselmann and Tanner 2008; Luce et al. 1997). In addition, ethical problems rarely offer obvious solutions concerning which course of action is most ethical. This requires sound and viable reasoning skills.

We suggest that this type of discernment is another key area where an exchange between empirical and normative viewpoints is useful. Acknowledging the complexity of ethical decision-making is a critical step with respect to how a leader endowed with reflective capabilities should proceed in order to come up with a reasonable and justifiable solution. What are the criteria for appropriate and effective ethical decision-making and a reasonable solution? Again, we will address these questions later, in Sect. 14.4.

Note that reasoning and reflection typically require conscious cognitive or deliberate processing efforts. What if leaders work under conditions that limit their capacity for controlled processes, such as time pressure or high mental workload? Empirical evidence has demonstrated that under such conditions, which limit the capacity for extended reflection, individuals tend to rely on intuitive judgments (e.g., Fazio 1990; Marquardt and Hoeger 2009). Moral psychology and moral theory, mainly in the Kantian tradition, have focused on the conscious and deliberate aspects of moral judgment (e.g., Kohlberg 1984). This research, however, has underestimated the role of intuitive and affective processes in (moral) decision-making—a critique that was, among others, highlighted by Roberts (2003) and Nichols (2004), as well as by the social intuitionist model of Haidt (2001). According to Haidt, people often base their moral judgments on quick flashes of affectively-laden approvals or disapprovals ("gut feelings") which tell us that something is right or wrong (Haidt 2001; Monin et al. 2007). Meanwhile, an impressive body of research points to the fact that automatic and affective processes assert a much more powerful influence on judgment and decision-making than was previously believed (for an overview, see Loewenstein and Lerner 2003).

It seems obvious that, when under conditions that encourage intuitive rather than deliberate reasoning judgments, ethical leaders are expected to come up with the proper intuitions. Yet, when acknowledging that intuitions often play an essential and demonstrably causal role in decision-making, important questions arise as to how intuitive and reflective capabilities are or should be related. Is it acceptable, even desirable, to allow intuitions to affect ethical decisions or not? Are there "good" or "wrong" intuitions? What are the features of proper moral intuitions? The answer is not simple. There is a large body of psychological research demonstrating that intuitions and choices can easily be influenced by subtle, but otherwise irrelevant factors such as mood, problem descriptions, or the presence of others (Loewenstein and Lerner 2003). This provides little confidence about the relevance of intuitions. Other research, in contrast, supports the view that intuitions and affect contribute to better decision-making because they provide vital information about aspects of the current situation or about past experiences (Damasio 1994; Baumeister et al. 2007). This is another point of intersection for a dialog between empirical and normative approaches (see Sect. 14.4).

Third, we expect ethical leaders to act upon moral values, consistently and persistently, regardless of the presence of external obstacles. Ciulla (1999: 169) pointed out that "leaders sometimes lack the ability or the moral courage to act on their values". But, leaders' values only matter to organizations and followers if they convey their values and beliefs through "visible" actions. Ethical leaders are therefore expected to act in accordance with ethical standards. Even more, we wish them to behave so on a regular basis. Fundamental to our conception is that consistency between words and deeds must be demonstrated repeatedly, across time and situations (Tanner et al. 2010). This is based on the idea that the more a leader maintains an ethical stance over time and situations, and the more predictable and transparent his or her behavior, the more likely observers will be to characterize the leader as credible, trustworthy, or possessing integrity (Tanner et al. 2010).

Whether or not leaders act on their values is also influenced by their moral courage, the state of mind that enables one to pursue what is considered right, despite potentially unpleasant consequences (e.g., threat to career survival, financial costs, social pressures; Sekerka and Bagozzi 2007). We believe that ethical leaders are more likely than others to display moral courage and take a stand, even when it is costly (see also Solomon 2003).

We now turn to the aspect of ethical managers, which focuses on efforts to influence the ethical conduct of followers (Treviño et al. 2000; Palazzo 2007). In general, there are several ways leaders can affect the ethical behavior of workers, including communication practices, performance compensation practices, ethical training, or codes of ethics, etc. (e.g. James 2000). We focus here primarily on two mechanisms that have been revealed to be important: role modeling and reinforcement by rewards and punishments.

Being a role model: Brown and colleagues (Brown et al. 2005; Brown and Treviño 2006) emphasized that ethical leaders should promote normatively appropriate conduct via communication of clear standards and intentional role modeling. According to social learning theory (Bandura 1986), leaders influence the behavior of followers through modeling. Bandura demonstrated the relevance of vicarious learning, suggesting that individuals do not only learn through their own, direct practice, but also by observing others' behavior and its consequences. The specific mechanisms involved are observation, imitation and identification. That is, by observing ethical leaders, followers may come to identify with those models, internalize their values and standards, and imitate their behaviors (Brown and Treviño 2006). Thus, having ethical leaders as role models can promote ethical conduct. Obviously, a leader's capacity to be an ethical role model is also based on that leader's ability to act upon ethical values, as noted above. We propose that only by engaging habitually in ethical behavior can leaders come to be seen as ethically credible models by the workers. We conclude that ethical leadership also entails leaders becoming models of ethical conduct by engaging in ethical behaviors (Brown et al. 2005).

Reinforcing ethical conduct: In order to generally promote ethical conduct, it is essential for each organization to have a kind of feedback system that reinforces the achievement of ethical goals. There is an extensive theoretical and empirical

literature indicating that organizational (formal or informal) rewards and punishments affect ethical behavior (see e.g., James 2000; Metzger et al. 1993). Above, we have highlighted the potential of vicarious learning. Followers also learn and adapt their behaviors through direct experience and its benefits and costs. Undoubtedly, due to their status and power to influence the organization and outcomes of others, leaders are an important source of reinforcement. Ethical leaders are therefore expected to set ethical expectations for followers and to hold them accountable by giving direct feedback on employees' conduct. In essence, ethical leaders should ensure that unethical behavior is punished, while ethical behavior is rewarded (Ciulla 1999; Treviño et al. 2003). The organizational structure established by leaders combined with informal organizational factors such as the corporate culture, are key elements in promoting ethical conduct (James 2000).

However, implementing a useful and effective reinforcement system is not simple. Some ethical lapses may go undetected, and others may not be the result of willful intent. In order to apply rewards and punishments, a leader must monitor and control follower's behavior. However, prior research suggests that too much control can undermine followers' work motivation or, when perceived as a threat to freedom, augment their resistance (i.e., reactance; Brehm 1966). Thus, it is not simple to ensure an "ethically balanced" system (James 2000) that does not inadvertently discourage ethical conduct.

To summarize, drawing on prior literature and building on our own work, we conceptualize ethical leadership as entailing: (a) adherence to ethically upright values and (b) endowment with ethical competencies. The latter entails subcompetences, such as ethical sensitivity, ethical problem solving skills (including proper reflection and intuition), and the ability to act in accordance with ethical values across time and various settings. Furthermore, ethical leaders should have the ability to influence and encourage employees to behave ethically (Ciulla 2006). For this reason, they should (c) serve as role models for employees and d) use rewards and punishments to promote ethical conduct.

14.4 Interdisciplinary Research into Ethical Leadership: Intersections with Normative Reflection

In Sect. 14.3 we identified and selected a number of key areas and questions where an integration of psychology and normative ethics appears to be important to improve our understanding of ethical leadership. We focused on three sets of questions:

1. Which values can we consider as essential in the context of ethical leadership? Which moral principles should leaders convey?
2. What are the essential features of ethical decision-making and reasonable solutions?
3. Is it acceptable or even desirable to allow intuitions to affect ethical decisions or not? Are there "good" or "wrong" intuitions?

14.4.1 Which Moral Values and Norms May Be Essential for Ethical Leaders?

There is no doubt that ethical norms and values play a crucial role in economic and management contexts. But ethical leaders need to know which moral norms and values are considered as essential. This question has become increasingly significant with the advance of globalization. Three points should be addressed here.

First: Throughout the world, a large number of countries has signed the "Universal Declaration of Human Rights." This indicates that the respect for human rights is nearing a consensus in the international community. Companies and ethical leaders can therefore draw on this global moral framework to adopt important values and standards. Even though the claim that human rights are grounded in universal moral principles has provoked highly controversial and lengthy philosophical and theological debates (cf., Dunfee and Donaldson 2002; Beauchamp 2010), it is reasonable that companies and leaders apply this moral framework to their own values, goals and actions when faced with tangible ethical problems. Indeed, the "Universal Declaration of Human rights" and its subsequent covenants represent the greatest normative consensus achieved on this topic within the international community. In light of this level of agreement, any respectable company must ensure that its legitimate pursuit of profits does not lead to 'collateral damage' in terms of human rights (Leisinger 2006: 15).

Furthermore, beyond the official consensus on human rights, a multitude of universally recognized norms, values and virtues exists for ethical leadership (e.g., Ciulla 2003; Price 2008; Solomon 2009). They include integrity, responsibility, compassion and forgiveness, as well as respect, honesty, integrity, caring, encouragement, courage and fairness, to name just a few. Norms, values and virtues of this nature have been well researched, both empirically and interculturally (for an overview, see Resick et al. 2006).

The problem is not so much a lack of awareness of these norms, values and virtues, but the fact that they are usually formulated so generally that they fail to provide orientation for specific actions. Questions often arise regarding how they are to be interpreted and implemented in individual contexts. For this reason, Beauchamp (2010) states how important it is "that we engage in specification: the process of reducing the indeterminate character of abstract norms and generating more specific action-guiding content. All general norms must be specified for particular contexts" (Beauchamp 2010: 260). This is true, not only for moral protagonists making ethical decisions and engaging in ethical reflection, but also for empiricists conducting research in this field. Empirical studies are bound to an analogous specification process if their conclusions are to be of any use, when they operationalize norms, values or virtues (e.g., Tanner et al. 2010: 229). For example, the value "respect" is operationalized by Tanner et al. (2010) in the context of the following two behaviors by leaders: "Insults coworkers while others are present" or "Includes employees in decisions that affect them." It is important to find appropriate specifications and

interpretations if research is to be fruitful, and this task is one that can only gain from an intersection between psychology and ethics.

These interpretations and specifications can, of course, lead to divergences and tensions, especially against a background of culturally differing interpretation patterns. It is therefore important to know precisely the contextual circumstances of a moral debate when striving for good ethical decision-making (cf., Bleisch and Huppenbauer 2011: 18–31). Since, in addition, it is likely that individual prejudices, biases or group interests may have entered into the abovementioned processes of interpretation and specification (cf., Rawls 2005: 58), critical reflection is a must.

Second: Another problem exists in the issue of which moral theory should underlie ethical decision-making. Moral theories (e.g., deontology, consequentialism, ethics of virtue or contractualism) serve to evaluate and substantiate the ethical legitimacy of actions, norms and values (Audi 2010). They provide fundamental normative criteria. Depending on which theory is adopted, different judgments and decisions can result. Various authors advocate pragmatic and pluralistic dealings with moral theories (e.g., Goodpaster 2002; Crane and Matten 2010; Miner and Petocz 2003). They prefer not to rely on one theory alone but on different theories. An important basis for this position is everyday moral life. Often decision makers use consequentialist as well as deontological reasoning to arrive at ethical judgments (cf., Sparks and Pan 2010: 413). In line with this, Goodpaster (2002) employs four principal "normative lenses" (or "avenues"). With them, he systematically questions the interests, rights, obligations and virtues of all those involved and affected (Goodpaster 2002: 127ff.). The "normative lenses" he uses correspond to fundamental normative criteria representing important voices in the ethical debate. In a modified guise they take up the abovementioned moral theories, aiming to achieve an adequate and potentially complex "insight of the moral point of view."

Third: So-called "bottom-up" approaches are being used more and more in applied ethics instead of the classic "top-down" approaches. Rather than taking abstract moral theories as a starting point, opting for one theory and then applying that theory to real situations, moral principles are instead critically reconstructed on the basis of different areas of practice and moral experiences, as well as intuitions, and then used within the framework of ethical decision-making. Particularly in this context, then, it makes sense to speak of "empirically informed ethics" (cf., Musschenga 2005). In the classic work on this methodology (Beauchamp and Childress 1979), autonomy, beneficence, non-maleficence and justice are cited as four universally recognized and therefore consensual mid-range principles. Regarding the normative criteria used, this ethical approach is therefore pluralistic. An analogous pluralism can be found in more recent works aiming to provide decidedly practical and viable methods for ethical decision-making (cf., Mepham 2008; Weston 2008; Bleisch and Huppenbauer 2011). Pluralism does not mean that the moral norms and values relevant to ethical leaders (such as fairness, respect, honesty, integrity) are arbitrary. Pluralism means that their justification and application to specific situations can occur within the framework of different moral theories.

14.4.2 What Are Essential Features of Ethical Decisions and Reasonable Solutions?

From a normative perspective, ethical competence undoubtedly includes certain reflective skills (Maak and Ulrich 2007: 383ff., 480ff.). From a certain distance and with a certain neutrality, ethical leaders have to be able to recognize moral issues, then to analyze and incorporate them in a reasonable decision (Bleisch and Huppenbauer 2011; Chap. 7 by Tanner and Christen, in this volume). In addition, ethical leaders are expected to justify their decisions, with sound arguments within their own company and in discussion with external stakeholders. What "reasonable" solutions are, how such arguments and justifications are to be structured, and how decisions are to be reached, all constitute crucial questions. Three points of interest with regard to this situation will be addressed in the following section.

First: As far as the meaning of "reasonable" is concerned, we find ourselves entering the terrain of moral philosophy. Taking all the information relevant to a problem (empirical facts, the interests of stakeholders, legal contexts, etc.) as a basis, it is sufficient for tangible problems to demand that controversial issues are processed in a manner that is intersubjectively comprehensible. This does not mean that a consensus must emerge. It simply means that reasonable persons have to be able to comprehend the decisions reached. In a famous formulation by John Rawls, reasonable persons are those who are able to "draw inferences, weigh evidence, and balance competing considerations" (Rawls 2005: 55). Since the use of such logical and argumentative means does not incorporate mathematically precise procedures and rules, differences of opinion are inevitable, as Rawls himself makes clear. However, it appears to be important with regard to moral practice that the interests and concerns of those affected and involved are fairly taken into account during the decision-making process (Dunfee and Donaldson 2002).

Second: Ethical leaders need to know how to arrive at well-structured and comprehensible results. Goodpaster (2002), for example, suggests a five-step method to cope with moral problems: (1) Describe the key factual elements of the situation; (2) Discern the most significant ethical issues at stake; (3) Display the main options available to the decision; (4) Decide among the options and offer a plan of action; (5) Defend your decision and your moral framework (Goodpaster 2002:128; see also Bleisch and Huppenbauer 2011 with an analogous model). Other authors have presented different methods of ethical decision-making (cf., Miner and Petocz 2003; Payne 2006; Maak and Ulrich 2007). Which of these methods is used is of less importance than the fact that ethical decision-making proceeds in a well-structured way.

Third: As stated earlier, when discussing the leader's problem-solving capabilities (Sect. 14.3.2), leaders often rely on intuitive judgments to address commonly recurring situations because a lack of time and resources inhibits them from carefully applying methods of ethical decision-making. Nonetheless, from an ethical point of view, a retrospective critical analysis of intuition-guided behavior is recommended to assess its adequacy. But even when sufficient time is available, it is

important to acknowledge that methods of ethical decision-making do rarely produce unambiguous and reproducible results (Palazzo 2007). Ethical decision-making is thus a process that must be continually reassessed.

14.4.3 What Relation Exists Between Moral Intuitions and Ethical Reflection?

This question brings us back to the abovementioned difference between empirical research and philosophical-normative reflection. On the one hand, empirical research can study how moral intuitions and reflection are interrelated in real ethical decision-making situations. On the other hand, philosophical reflection can (possibly based on these empirical findings) establish norms for how ethical decision-making should take place.

As mentioned in Sect. 14.2, psychological research has promoted acknowledgment of the role of moral intuitions and automatic processes in ethical decision-making (Kennett and Fine 2009; Treviño 2009). Meanwhile, a number of philosophical (and theological) authors have also advocated metaethical and methodological positions, according to which moral intuitions are an important component of ethical decision-making. In fact, the problem is not the moral intuitions themselves, but the question of the nature of the role they should play in processes of ethical decision-making. Looking at van Thiel and van Delden (2010: 189; see also their Chap. 10 in this book), one can first define moral intuitions very generally as "beliefs that a person comes to hold without extensive deliberation." On this foundation they then present an interesting model for how, within a theory of "Reflective Equilibrium," empirical findings can be used during decision-making: "The thinker who wants to produce a reflective-equilibrium has to consider empirical elements together with normative principles and background theories. In this process, the thinker aims for coherence among all relevant considerations" (van Thiel and van Delden 2010: 193). Empirical elements refer to the moral intuitions of practitioners who have gained a wealth of experience in their specific contexts: "People who work and live in a certain moral practice have experiences that are generally not found among those outside this practice" (Van Thiel and van Delden 2010: 187). Seen from this perspective, this model also makes it clear that moral intuitions are not simply emotions and affects which occur randomly and then disappear again, like anger, annoyance or rage: "These experiences amount to specific moral wisdom, which can be defined as expert-level knowledge and judgment in the fundamental pragmatics of life" (Van Thiel and van Delden 2010: 187).

Due to its focus on acquired competencies, this definition of moral wisdom can readily be linked to approaches from an ethics of virtues (Solomon 1992, 2003, 2009), yet is also in line with psychological approaches. According to several authors, intuitive decisions are highly accurate when they are "expert-like" (Dane and Pratt 2007; Hogarth 2001). As Narvaez and other scholars posited, moral

experts are similar to other experts. They differ from novices in that they have more complex, domain-relevant and chronically accessible mental structures, which trigger effective responses (Dane and Pratt 2007; Narvaez 2005; Lapsley and Narvaez 2005).

Van Thiel and van Delden (2010) argue that no intuitions need advance discarding from the ethical decision-making process. This stems from their definition of intuitions as expert knowledge. As the result of experiences, these intuitions do of course contain manifold influences, as well as reflections about the experiences in question (cf., Musschenga 2009: 608). Since they wish to have a basis for ethical decision-making that is as broad as possible, and an ethical judgment that is supported as broadly as possible, they rely on as much expert knowledge as possible (Van Thiel and van Delden 2010: 198f.). From an ethical perspective, this does not mean, however, that all the intuitions brought into play by experts are inherently correct. As already mentioned, prior research has revealed that intuitive judgment choices can easily be influenced by subtle but otherwise irrelevant factors such as mood, problem descriptions, presence of others (Loewenstein and Lerner 2003). Furthermore, it is probable that specific prejudices, biases and group interests may have entered into moral intuitions (see also Musschenga 2009). The important task of reflection and critical deliberation is thus to adopt a critical stance towards moral intuitions: "In the … process of moral reasoning, moral intuitions, principles and theories can gain or lose justificatory power" (Van Thiel and van Delden 2010: 198). In short, it is not self-evident that every expert intuition is ethically justified. This can only be judged as the result of an empirically enriched deliberation process.

Despite empirical research having demonstrated that human beings do not regularly draw on critical analysis and reflection, this does not imply that reflection is not needed; quite the contrary! Leaders should use reflective competencies at least in difficult and controversial situations. Lack of time is not a sound argument in most instances. Since moral questions usually address important issues, they should be processed with the same degree of earnestness and expertise as other important business issues. We do not intend to imply that protagonists have to be in a constant state of reflection, there are certainly many situations where intuition-based decision-making is clearly appropriate. Key occasions for reflection arise when conflicting interests are held by company stakeholders and intuitions are not helping to resolve the turmoil.

Since stakeholders are likely to have divergent interests and moral intuitions, not only the moral intuitions of the leaders themselves should be integrated in the process of ethical decision-making, but also those of the relevant stakeholders. Unfortunately, intuitions and convictions held by different stakeholders are sometimes directly and irreconcilably opposed (Leisinger 2006: 19). To deal with such situations, obviously, recourse to intuitions is not enough. Reasoned communication is needed between those involved, and in order for this to succeed, reflective and critical competences are required: "Reasoning skills may not be necessary for finding the right answers to moral problems, but you cannot participate in collective debates without having them" (Musschenga 2009: 609).

In conclusion, this chapter was designed to highlight some typical features of the empirical and normative approaches toward exploring ethical leadership and ethical decision-making. Based on a respect for the unique disciplinary foci, while remaining critical, we tried to sketch out some areas of intersections where ethics and psychology can learn and benefit from each other. Of course, more work and ongoing dialog are needed to develop further forms of integration. We believe, however, that attempts at interdisciplinary collaboration in the development of business ethics are, in the long term, beneficial for researchers and practitioners alike.

Chapter 15
The Empirical Turn in Bioethics – From Boundary Work to a Context-Sensitive, Transdisciplinary Field of Inquiry

Tanja Krones

15.1 Introduction

The debate on facts and values in (bio)-ethics is also a debate on the contribution of the social sciences and psychology to bioethics and vice versa. This debate has recently reached a new state of reflection. It started with indifference in the early 1970s, when both ethics (philosophy, theology, law) and the (social) sciences (especially medical sociology and medical and social psychology) began to penetrate the field of biomedical science and practice from its margins.[1] A phase of some interest, debate and cooperative efforts followed, when both disciplinary fields bloomed and became institutionalized in the late 1970s and early 1980s.[2] The first critique of bioethical reasoning was uttered by the social sciences in the 1980s and 1990s, predominantly not expressed in bioethical but social science and theory of science journals (cf. Hoffmaster 1994). At that time, bioethics was not only established as an important scientific field outside the US, but also as a political endeavor of a pool of experts taking part, and positions in, biomedical and political institutions and debates. Today we witness a fundamental and central scientific debate on a practical, theoretical and epistemological level in the social sciences, philosophy and bioethics. This debate entails a thorough reflection of the contributions of: the social sciences to the core project of bioethics; ethics to the discussions in the social sciences; and both social sciences and bioethics to one of their (many) aims they have

[1] William Cockerham, for example, did not mention the term 'ethics' in the subject index or even once even in its fourth edition of *Medical Sociology* in 1989.

[2] This was fostered by research surrounding the human genome project on ethical, legal and social implications, but also in other fields of biomedicine, such as intensive care. See for example the work of Chadwick et al. (1992), Chadwick (1987) and Wertz and Fletcher (2004).

T. Krones (✉)
University Hospital Zurich, Institute of Biomedical Ethics, University of Zurich, Zurich, Switzerland
e-mail: tanja.krones@usz.ch

M. Christen et al. (eds.), *Empirically Informed Ethics: Morality between Facts and Norms*, Library of Ethics and Applied Philosophy 32, DOI 10.1007/978-3-319-01369-5_15,

in common which is to analyze, reflect on and (I would stress) improve theory and practice of medicine and health care.

In this article I first embed the debate on empirical ethics in recent theory of science discussions on early "linear" versus late reflexive modern thinking (Part II). I consider the movement of a truly common project of descriptive-normative ethics as part of these reflections that is visible in many fundamental epistemological discussions in philosophy, science and politics. Core features of these debates are the analyses of origins, content and use of presumably self-evident boundaries between facts and values, philosophy, science and society. This 'boundary work' is also highly visible in the debates on empirical ethics (Part III). Origins, content and use of arguments of an 'orthodox', linear, philosophically dominated model of applied bioethics and of an epistemic and technical social science against a 'heterodox' empirical-normative transdisciplinary model of a pragmatic, "phronetic"[3] bioethics are considered (Part IV). I then sketch theory (Part V) and praxis (Part VI) of a context sensitive bioethics that takes the diagnosis of reflexive modernity, pragmatism and pluralism seriously. Through this analysis I hope to convince boundary workers, that a reflexive, transdisciplinary approach to bioethics is more fruitful than continuing old disputes in order to advocate own sinecures.

15.2 Bioethics as a Child of Reflexive Modernity

The debate on the empirical turn in bioethics is part of a wider reflection in theory of science on modernity and late reflexive (or: post-) modernity. This debate influenced the whole twentieth centuries' theory and philosophy of science. For some, the labels and concepts of this discourse now seem to be no more than catchwords. Yet, central arguments of this discourse, such as the fact-value distinction, rejection of metaphysics and positivism on the one hand and the fear of relativism on the other, are core arguments in the debate on the empirical turn in bioethics. That is of no surprise.

Bioethics itself is a child of this time and debate. Proponents as well as opponents[4] of late reflexive post-modern thinking in science and ethics share a historical diagnosis: In post-war society, modern certainties such as progress through human

[3] The reflections on a "phronetic" versus an "epistemic" science, a differentiation first made by Aristotle in his critique of Plato in Nicomachean Ethics, are now found in social science (e.g. Flyvbjerg 2001) and also in bioethics (e.g. Engelhard 1999). See also the discussion below.

[4] Besides many French sociologists and philosophers, such as Jacques Derrida, Francois Lyotard, Michel Foucault, Paul Ricoeur, Paul Valéry and Pierre Bourdieu and the German sociologists and philosophers Ulrich Beck, Odo Marquard and Wolfgang Welsch, I consider Richard Rorty and Judith Butler as the most thorough and differentiated philosophers of modernism and post-modern thinking as a pluralism of rationalities and truths; See Richard Rorty's early book on epistemology (1979) and his later book on practical philosophy (1989). In regard to ethics, Judith Butlers Adorno's lectures (2003) describe some of the most important features of a late modern self. As some of the most important and differentiated opponents of late/post-modern thinking who share

techné and scientific advances, basic modern principles such as rationality, unlimited economic growth, certitude in knowledge, law and order, and basic differences and boundaries between nature/culture, humans/animals, men/women, lay people/experts, life/death, theory/praxis are increasingly eroding. Instead of certainties, ambivalences and risks, the focus on unintended consequences and the insight that we always produce non-knowledge while producing knowledge, have become central subject matters in theory of science, knowledge and sociology. The core topics in bioethics exactly deal with these ambivalences that are highly visible in biosciences and medicine. The debates are especially heated in this field because biology has become fundamental in western images of human beings.[5]

The reaction towards this diagnosis in epistemology, however, differs. At the beginning of the twentieth century young protagonists of positivism and analytic philosophy assumed that objective truth, or making sense of the natural and social world, could be achieved by epistemic scientific rigor. These assumptions were dismissed in the dispute on (neo)positivism and logical empiricism which started before and continued after World War II.[6] After the dispute had calmed in the 1970s, none of its participants believed in a naïve "fairy tale of an objective observer" (Von Uexküll and Wesiack 1998: 32), a firm basis of knowledge, a clear distinction between facts and values, or the unimportance of intellectual contributions of any kind of scientists to the ethical domain. Yet, the topic came back on stage again in the mid-1990s, when the diagnosis of late modern uncertainties and ambivalences was taken seriously in the philosophy of science. In the so called 'science war' between (social) science 'post-modernists' and (natural) science 'realists',[7] some physicists and biologists have again defended a positivistic science as verification of objective truth. Similarly, in (bio-)ethics the fear of relativism as an 'anything goes' mentality to deal with the 'cacophony' of health professionals' and patients' values and beliefs (Macklin 2000) underlies argumentations against a central contribution of non-positivist (social) sciences to the core normative project of ethics. Many philosophers and (natural) scientists engaged in the debate on an empirical ethics still, and again,

---

the diagnosis of late modern uncertainties in epistemology but defend some early modern features in ethics, I do consider Jürgen Habermas (e.g. 1987) and Benhabib (1992).

[5] I depict the notion of biology as meaning in Part VI. For this debate see for example Franklin (1997), Nelkin and Lindee (1995) and Haraway (1989).

[6] For an overview see Dahms (1994) and Von Uexküll et al. (1996). Main protagonists of the dispute on positivism were Neurath, Carnap and (in the second phase) Popper on the side of analytic philosophy and critical rationalism, Horkheimer, Adorno and (in the second phase) Habermas on the side of critical theory.

[7] On the side of the 'realists', Gross and Levitt (1994) defended objective science against the, as they see it, 'irrational postmodernists'. The science war became even more fierce when the "Sokal Hoax" took place: The publication of an article of physicist Alan D Sokal in a high impact social science journal on "Transgressing the boundaries: Towards a Transformative Hermeneutics of Quantum gravity" later revealed to be a bogus article by Sokal himself resulting in the assumption that most social science is bunk. This was answered by social scientists like Shulman accusing Sokal and other natural scientists of being "pre-kantian shamans repeating the mantra of particle physicists" (Flyvbjerg 2001: 1). For a debate on this book and reaction of social scientists see Flyvbjerg (2001) and Ashman and Bahringer (2001).

believe that we have to defend early modern certainties, like in Newtonian physics and Kantian ethics of the eighteenth century, against 'relativism': to resurrect a-historical, a-social norms and principles of an independent normative ethics sphere on the one hand, and a-historical, a-social objective facts of a positivistic science on the other, which are engaged in the search and finding (science), or defending, (philosophy) normative and factual truths. Several religious movements see the crisis of modern certainties as a chance to resurrect even pre-modern beliefs, such as creationists or other Christian or Islamic fundamentalists. Other philosophers, theologians and scientists, such as Albert Einstein, Thomas Kuhn, Gianni Vattimo, Richard Rorty, Anthony Giddens, Michel Foucault or Judith Butler take the diagnosis of new uncertainties seriously into account on an epistemological and ethical level. From their epistemological point of view, these uncertainties were not only recently produced by science and technology, but they were inherent in the whole project of enlightenment. These certainties are considered as early modern prejudices, grown out of an understandable quest for certainty directed towards science and philosophy after "god was dead". Concepts like truth, rationality, and objectivity are not dismissed, but are considered as important questions for which we find socio-historically different, plural answers. In late reflexive modern thinking, starting with skepticism of the "lost generation" after World War II and described as a late reflexive or post-modern era since the 1980s, the modern disenchantment of the world (Weber 1918) is followed by a late modern disenchantment of natural and social science (Bonß and Hartmann 1985) and re-enchantment of the social world (Rorty 1984). A central feature of these reflections and analyses is the deconstruction of former self-evident boundaries established by "boundary workers" in early modern thinking.

15.3 Boundary Work

As protagonists in philosophy, sociology and anthropology have analyzed since the beginning of the twentieth century (e.g. Malinowski 1925/1975), science has become the new modern meta-narrative for explaining, making sense of and ruling the entire world. Its legitimacy was not only established because of the successful endeavor of enlightening society through disenchanting the world, but through scientific policy as boundary work. The term boundary work was first introduced by Gieryn (1983) and has become influential in theory of science in sociology and (bio)-politics.[8] In this train of thought, drawing discursive boundaries between science and non-science, and philosophy and society, is a more or less a conscious strategic policy, an "ideological effort" (Gieryn 1983: 783) to define an exclusive subject of inquiry as a field free of interest and ideology. Scientists and philosophers claim themselves as experts of this field in a search for truth, only subjected to

[8] See for example Bogner (2005) as a very comprehensive description of boundary work in the field of biopolitics from a sociologist and technology assessment point of view. See also Jasanoff (1990).

epistemic rationality, and thus guaranteed for certainty. Another, wider reflection on the importance of demarcation and boundaries in meta-narratives (such as god, nature, truth and rationality) is the archeology of knowledge and history of truths, a theory of science undertaken, for example, by Berger and Luckmann (1969), Foucault (1971/2004), Kuhn (1962) and Bourdieu (1984/1988). According to these philosophers and scientists, scientific boundary work is not merely a conscious strategy but expresses internalized perceptions of the world linked to "truth politics" (Foucault 2007) and adopted while being socialized as an academic. In this view, priests of pre-Socratic times, philosophers of ancient Greece, theologians of Middle Ages, and scientists of modern age all established different rationalities and social truths. The specific interpretations of what is right and relevant by respective legitimate guardians of truth are no longer questioned they are 'epistemic doxa', established in the political history of a scientific or societal field. These boundaries are resuscitated and stressed, when putatively self-evident responsibilities are at stake. The analysis of the use of ('mere' rational) arguments in scientific and philosophical discourse as (partly conscious) means to defend an independent field of knowledge has become important in post-war philosophy of science, an age many characterize as reflexive modern age. This reflection is important in order to distinguish the analytical or practical usefulness of arguments and their goal to contribute to further (reflexive) enlightenment from their use as a strategic and often unproductive means to defend one's discipline against other disciplines and against other societal forces. In the discussion on the empirical turn in bioethics, core boundary work arguments include the fact-value distinction corresponding to the distinction of normative from descriptive ethics and the argument that if we blur this boundary we will end up in relativism and destroy a meaningful ethics or social science.[9]

15.4 Roots of Boundary Work Arguments in Empirical Ethics: Old Dualisms and Disputes

Nothings seems to be more obvious for many (bio)-ethicists, psychologists and sociologists than the existence of what is called the "is-ought" or "fact-value" distinction, often subsumed under the naturalistic fallacy argument. Let's first give two prominent examples, one from philosophy and one from sociology:

Neither descriptive nor analytical-metaethical inquiry can establish what is morally good, right or required in a particular case. They cannot extrapolate from the 'is' to the 'ought' without destroying normative ethics (Pellegrino 1995: 162).[10]

[9] For a description of boundary work from the view of a philosophical bioethicist, see Herrera (2008). For a good example of describing this boundary work in bioethics in the view of a social scientist, see Rapp (2000).

[10] Other prominent examples of taking the naturalistic fallacy (although critically thinking about it) as a main principle of a meaningful ethical endeavour are Beauchamp and Childress (2001: 2) and Sugarman and Sulmasy (2001: 6–11).

Science today is a 'vocation' organized in special disciplines in the service of self-clarification and knowledge of interrelated facts. It is not the gift of grace of seers and prophets dispensing sacred values and revelations, nor does it partake in the contemplation of sages and philosophers about the meaning of the universe (Weber 1918).[11]

The logic used in the debate on the fact-value distinction in bioethics and in a positivist (social) science entails the following:

1. There is a clear and important distinction between the factual world (science) and the normative sphere (ethics, philosophy).
2. Science usually illuminates facts; yet, facts about the social sphere are different from facts of the natural world, the brute facts. To come to conclusions and prescriptions of what should be done we need what is defined as normative ethics, relying upon but being independent of the description of the social and natural world. If we build our normative argument on facts, we commit a naturalistic fallacy.
3. There is a clear and defended disciplinary boundary between the two fields of (social) science, including psychology and sociology on the one hand and philosophy, including ethics, on the other.
4. Ethics in the view of social science is not a sphere of scientific thinking but is either politics (Max Weber) or should be clearly grounded in (social and/or scientific) facts (Laurence Kohlberg).
5. It is not the main task of the (social) sciences (in the view of scientists) nor do (social) scientists possess the capability in the view of (philosophically trained) bioethicists to fundamentally contribute to questions and reflections in the normative sphere, or to work on and provide prescriptions or guide the behavior of people.

In order to evaluate the fact-value distinction, it is useful to shortly recapitulate from where it comes.. I do not discuss current meta-ethical positions but depict historical roots of meta-ethical reflections and constructs of the fact value distinction and of the notion and concept of the naturalistic fallacy argument. I do this in the sense of Foucault's archeology of knowledge (Foucault 1973) to illuminate the use of these arguments in the body of discourse, not only influenced by rationality but by politics of truth and power.

The main contribution to the fact value distinction comes from Descartes in his *Discours de la méthode* published in 1637. He established science through separation of science and philosophy from theology via his famous distinction of Mind and Matter, Body and Soul, both only minimally interacting via an anatomic structure, the pineal gland. This Cartesian Dualism and Rationalism established science as a verification of truth via logical analysis and reflection on the one hand and observation of facts on the other. Hume's empiricism reflected upon this dualism and accepted the existence of the two spheres, but not the assumption that human

[11] As a psychologist on the naturalistic fallacy see Kohlberg (1971: 151, 222), cited in Biller-Andorno (2001: 22).

reason is capable of finding truths on a logical, rational matrix. His philosophy led him to a fundamental skepticism: Our truths are habits, adopted because our experience is grounded in empirical sense-data of our perception, and these truths do not give us the chance of believing something else. The causalities that we observe are thus only coincidences-also between the facts we observe and our morals. He reflected upon the arguments used to establish what is right or wrong in the Principles of Morals (1751) and on the is-ought distinction in A Treatise of Human Nature (1739–1740), what was later (with a slightly different meaning) by G.E. Moore defined as the naturalistic fallacy (see below). Many use 'is' arguments for directly establishing truths of what ought to be- which is a causal inference that cannot simply be applied. But Hume also strongly rejected what is sometimes called a normativistic fallacy: To prescribe what is right or wrong out of logical reasoning from a sphere independent of experiences in its very normative heart of reasons; a mistake ascribed by Hume to rationalism and to natural philosophy (Hume 1751). Kant in his early years was very much influenced by French rationalism, the Cartesian dualism and its impact on the philosophical school of his teachers. The German rationalism stressed the capability of reason to find truths in physics, as well as in metaphysics and ethics. Hume, he depicts, woke him up from his dogmatic slumber, a philosophical dream of an omnipotent human reason, and developed his answers as to how philosophy could be resurrected in a world of Newtonian certainties and irresolvable philosophical disputes. He turned philosophy into a skeptical meta-science on the one hand and a philosophical ethics without fundamental skepticism through defining scope and limits of reason and determinism. In his *Kritik der reinen Vernunft* he integrated Hume's skepticism into the theory of science and transformed it: We can find real truths, that is, not only coincidences but causalities and determinants of facts, because we are rational beings. We use both, our reason and our senses to find scientific truths, but only in the factual world, not in metaphysics and ethics. This sphere, described in Kant's *Kritik der praktischen Vernunft*, is defined as an empire of freedom without determinism. Since man possesses reason, he (that is, every human being, but, to be honest, he almost excluded women in his thoughts) possesses human dignity and the right to determine himself, what is morally right or wrong—no one else, not the church, not science. Through this combination, Kant thought that he had not only saved science and ethics, but philosophy as a meta-discipline, and also the importance of religious beliefs in a sphere of freedom (a thought, heavily influenced by his pietistic family). Kant, Hume and Descartes were not only admirable philosophers and scientists, but also good boundary workers: Descartes and Hume struggled to free science from medieval theological domination, and Kant not only defended philosophy against theology, but also against science and strong rationalism, empiricism and skepticism.

Although all of them were boundary workers; all of them also made strong efforts to overcome the boundary when struggling for practical solutions, made pedagogical and political attempts to 'apply' their thoughts in societal processes (such as Kant in his less known last big oeuvre, the *Anthropologie in pragmatischer Hinsicht* of 1798).

The assumptions of Descartes and Kant on the fact-value distinction were challenged in the debates depicted above. But their train of thought of two spheres of social life, facts and ethics, especially Kant's definition of ethics as a sphere of freedom from facts, free from social and scientific determinants, are fundamental for the opponents of a central normative contribution of (social) science to the ethical debate. Not only Kantians but also many Principlists and Utilitarian thinkers tend to define their most general principles, be it utility, the 'four principles' of Beauchamp and Childress, or a Kantian version of autonomy and dignity, as features in a sphere free from factual determinants that is rooted in a universal logic [although first being empirically (that is: socially-historically) established] of not being subject to criticism by observation and object of social and historic change. If there is an attempt to use an observation (such as: autonomy is not a central feature of common morality in many cultures) to criticize the principles applied, the naturalistic fallacy is often used as a counter argument, for good reasons, but in a way that G.E. Moore has not meant it to be used. Moore was an admirer and critic of Kant, one of the main representatives of the Cambridge version of analytic philosophy and is often considered as being the founder of modern meta-ethics. In his famous book *Principia Ethica* (first 1903) he analyzed and criticized (like Kant before) the main contemporary school of thoughts, idealism and naturalism/empiricism/utilitarianism in ethical reasoning. The answers given to what is 'good' (mostly not defined as good things, like Moore did, but as morally good human conduct) by Moore's colleagues were widespread and various. Good is something because (and only because) it maximizes lust or utility (hedonism, utilitarianism), is a representation of pure and real nature (naturalism), follows god- or self-given laws and rules (theology, deontology) or serves the true self or freedom of human beings (phenomenology, existentialism). Moore's answer was simple and followed moral intuitionism: Good is something because it is good (Bishop Butler's "everything is what it is, and not another thing", he cites at the beginning of *Principia Ethica*). Good is a quality like yellow, which we do perceive but are not able to logically define.

However, the naturalistic fallacy was not understood by Hume or Moore as an invitation to use a set of 'Ought to Is' claims (a normativistic fallacy) as the one and only way to analyze societal morality. Both philosophers linked their conception of ethics to intuitions people have while reflecting on, or rather while experiencing 'the good'. Intuitions, like moral theories, can of course be wrong. But that is exactly the point of Hume and Moore: We can never logically deduce what is good or morally required from biological or other natural (brute) facts nor from pure theory, and we should be fundamentally skeptical if someone claims that he or she has found the Holy grail of ethics. As fallible beings we can only find good plausibility reasons for our ethical reflections and moral decisions, and these can be derived from proven theory as well as from proven practices, that are intrinsically related to each other.

If the naturalistic fallacy is not used as a boundary work argument (when it is not asserted that the boundary between facts and values is real, or that in ethics a-social, a-historical normative theories, informed by facts, judge practices, defined as "linear ethical reasoning"- from theory to practice and not vice versa by Lindemann Nelson 2000) it can serve as a very useful tool to reveal many oversimplifications in

social and natural science and philosophy. Biology or psychology do not necessarily contain normative values (naturalistic fallacy in a narrow sense), historical insights do not directly tell us what is right or wrong in contemporary society (historical fallacy or presentism), nor does the law tell us the truth (juridical fallacy), nor ethical experts (normativistic fallacy), nor societal majority opinion (*argumentum ad populum*).[12] According to Moore, we should include as many considerations and observable facts as possible to find our solutions, and they should always serve as an invitation to prove them unfounded. The naturalistic fallacy argument is very useful for analysis in ethics, to counter simple, putatively logical, clear inferences from ought to is or is to ought. It cannot be used, however, as an argument to clearly separate the descriptive from the normative in finding solutions, to favor only one theory in ethics (or science), or favor normative theory and experiments of thought instead of practical knowledge or empirical findings. Some bioethicists do this when they dismiss plausibility reasons from sources other than established philosophical normative theories on biomedicine. This is in fact part of the boundary work of the disciplines to carve out the scientific territory of valid descriptions and prescriptions in the field of medicine and to assert certainties 'against relativism' that cannot be asserted.

15.5 Crossing of Boundaries: The Theoretical Basis of a Descriptive-Normative Transdisciplinary Approach

In many areas of philosophy and science, we witness a crossing of mental and disciplinary boundaries. In his programmatic article on a context sensitive approach to (bio-)ethics, Musschenga (2005: 467) states that "in medical ethics, business ethics and some branches of political philosophy (…) the literature increasingly combines insights from ethics and the social sciences." Musschenga describes the development of empirical ethics as a descriptive-normative enterprise with a common goal: to increase the context-sensitivity and the validity and usefulness of ethics in practical dilemmas. According to Musschenga, this goal can be aimed for from any meta-ethical position. He sees a movement towards more context-sensitivity in a recent development of broad contextualist theories, including two theoretical strands: the coherence approach and the epistemic contextualist approach, but also in the more traditional approach of applied ethics.

The acknowledgement of facts (contexts) in the applied ethics approach seeks to "mak(e) ethics context-sensitive" (Musschenga 2005: 473). The meta-ethical position of the 'applied ethics' model is a linear ethics model, in which "moral theories, informed by facts, judge practices" (Lindemann Nelson 2000: 12). In applied ethics, facts and theory are not directly interacting nor is practice interacting

[12] See this discussion also by Sulmasy and Sugarman (2001), who excellently discuss these fallacies but still have one favored way of looking at the blackbird and who, to my mind, uphold the fact/value boundary.

with theory. The interaction is thus one way- linear, from theory to practice. The theoretical assumption is that prescriptions are made through translation (Birnbacher 1999)[13] of moral theory, its most basic norms and principles, into practice rules. Contexts (empirical facts) shape practice rules in many ways: practice rules have to be aware of limitations of cognitive capacities in information processing of persons in dilemmatic situations. In order to have an impact, they must be sensitive to social convictions and motivations. They have to ally with psychological and social theories of action in order to be 'feasible' and to take into account the moral reasoning of "average human beings" (Musschenga 2005: 475). They also have to be aware of, and demand the empirical surveillance of the possible misuse of unintended consequences of ethical practice rules in practice. Contrary to the second model of broad contextualist theories, in the applied ethics model, basic principles themselves are not subject to descriptive-normative inquiry. The most basic principles of moral theory remain decontaminated by practice, and are thus prevented from fundamental reflection and change. The descriptive and normative sphere, facts and values stay separate as far as the most abstract level is concerned, be it the utility principle of utilitarianism or the categorical imperative of Kantian ethics.

Broad contextualist theories, as defined by Musschenga, consider context not only as a field of application of moral theories, but as an important source of morality. The first contextualist model is coherentism: the aim is to reach a state of wide, reflective equilibrium not only in the Rawlsian sense of moral principles and beliefs on different levels of abstraction, but also in contextualized facts, principles and beliefs. The morality of the context is taken much more seriously into account and also leaves the most basic principles open to socio-historical change. Although Beauchamp's and Childress' Principles of Biomedical Ethics are very close to this coherence model of broad contextualist theories, the foundation of their four principles in their definition of the single universal common morality (instead of the morality of the context) keeps their most basic level of theory, their "most basic moral data" (Beauchamp and Childress 2001: 385) more resistant to change compared to the concept of broad contextualist theories in Musschenga's sense. With regard to their material ethics, the content of the principles, Beauchamp and Childress remain in the 'applied ethics model' they themselves thoroughly criticize.

The second model of broad contextualist theories, the epistemic contextualism, is even closer to practice and takes the contextual practice more seriously. Their aim is to reconstruct (not only to identify) the internal morality of practice. Musschenga considers this approach as the best framework for a likewise descriptive-normative ethics, although its meta-ethical presuppositions have to be articulated in a more systematic way in his and in my view:

First, one has to acknowledge that broad contextualist theories as well as many contemporary philosophers and (social) scientists are influenced by the pragmatic

[13] Birnbacher is a philosopher and ethicist, belonging to an utilitarian trait of thought, which he describes as an indirect, sensitive utilitarianism, taking 'traditional' (social, cultural) norms besides the utilitarian view into account (cf. Birnbacher 2006).

turn in philosophy. John Dewey's work on logic and philosophy of science and ethics are the basis of a non-foundational model of late modern ethics and science. Dewey replaced the classical static concept of logic, truth, reality and structure by a genetic one. He deconstructed the concept of logic which is supposed to be prior to scientific inquiry, as an a-historical eternal science. This model of logic, he states, was a methodology adequate for science in ancient Greece as a classification of substance and forms that were static in an eternal cosmos, in which change and measurement of substances were excluded as objects of science (Dewey 1938/1986). For him, logic is a tool, not a truth, to reach warranted assertability. Dewey plead for a pragmatic view also in ethics, considering theories as hypotheses and as tools rather than truths, and for a cultural naturalism, incorporating social science research into ethics. Thus logic and science as theory and practice of inquiry and ethics as theory and practice of morality have to be understood as culturally and socially rooted disciplines. Scientists and philosophers are not partly divine ideal observers but belong to the human species, bound to limits of epistemology and knowledge and to their own social contexts. The diagnosis of a genetic –changing truth, of a positional epistemology was widely accepted in theory of science, influenced by John Dewey's work on Logic. However, in ethics, a "hankering for certainty, born of timidity and nourished by love of authoritative prestige" is still in place, a belief that "has led to the idea that absence of immutably fixed and universally applicable ready-made principles is equivalent to moral chaos" (Dewey, as cited in La Folette 2000: 416). Broad contextualist theories consider ethics as such a pragmatic philosophical and social science. Different to an applied, linear, interdisciplinary model of bioethics, context and theory are closely related. There are complex interactions: Being, the 'is' influences consciousness and the 'ought', but human creativity also allows new interpretations of reality and new insights into formerly unmarked spaces. Some unconscious conditions of knowledge and unintended consequences can be disenchanted. Theory has an effect on practice, and sometimes causes revolutionary Kuhnian paradigm shifts. These shifts, as well as some new life experiences, open the view and refocus it onto other, previously neglected aspects, thus reshaping theory. Philosophical theory, as a tool, is therefore directly connected to action. Different philosophical, psychological and sociological theories consider different parts of the action process. Some focus on the conditions of actions, others on the consequences of actions, and still others on the action itself. In a pragmatic ethics, theories are not considered as algorithms, as "proclamations of something or someone outside us" (Dewey, as cited in La Folette 2000: 419) but as precious knowledge that might serve as tools to help us dealing with moral dilemmas.

Second, epistemic contextualism incorporates the diagnosis of reflexive modernity, the acceptance of plurality of rationality and truth into its model of ethical reasoning. Close to Wittgensteinian, (neo-) Aristotelian thoughts and Gadamer's hermeneutical tradition, epistemic contextualism sees different societal moralities and rationalities interacting in ethical problem solving. Internal moralities of practices (such as moralities in medicine) and moralities external of a practice (such as political or philosophical reasoning) interact and influence each other. In these interactions, epistemic contextualism draws on an epistemic and phronetic

philosophy and science. The critique of Marx towards Hegel that the thing of logic might not be the logic of things or, in other words, that practice has a logic which is not that of logic (cf. Bourdieu 1977), is stressed by protagonists in contemporary philosophy and social sciences such as Nussbaum (1986) and Flyvbjerg (2001). Both, like Hans Georg Gadamer, Pierre Bourdieu, Michel Foucault and others before them, draw on the Aristotelian concepts of prudence and of practical wisdom (*phronesis*) besides episteme and techné in ethics and social science, as they are disciplines that are no longer strictly separated from each other. For Flyvbjerg episteme, *techné* and *phronesis* all represent highly intellectual knowledge and skills that serve different purposes and belong to different spheres, but that are all connected to truth. Whereas an epistemic science, a deductive analytical enterprise resulting in a predictive causal theory, is an apt methodology for natural science dealing with context-independent, dead objects, and an object world, where social science and ethics deal with *self-reflecting* humans and a subject world, of which scientists themselves are part of. This sphere is only partly understandable in epistemic terms. It is also not sufficiently analyzed or guided by techné as a craft/art oriented towards production and goals based on a practical, instrumental rationality that underlies many economic considerations. Not episteme or *techné* but *phronesis* as a way of dealing with human beings informed by rich experience, value rationality and acknowledgement of different context, is the most important epistemological concept for ethics and the social sciences. Although theoretical (episteme) and instrumental (techné) knowledge are important for these disciplines, they only inform a primarily phronetic approach that is based on experience with values and circumstances, and enriched by concepts of power and conflict. Flyvbjerg connects these thoughts with the model of human expertise as described by Dreyfus and Dreyfus (1988, 2004). Dreyfus and Dreyfus summarize studies that observed that very experienced people (e.g. chess or football players, physicians or paramedics) abandon rule-based thinking and behavior on which novices and less experienced professionals rely. Skilled performers' and experts' thinking and behavior are adapted to various situations, and are "intuitive, holistic, and synchronic, understood in the way that a given situation releases a picture of problem, goal, plan, decision, and action in one instant and with no division into phases. This is the level of true human expertise. Experts are characterized by a flowing, effortless performance, unhindered by analytical deliberations" (Flyvbjerg 2001: 21). To reach this stage of expertise, one has to gain profound experience. Experience that is first guided by and influenced by cognitive rules which are then incorporated in skilled habits, adopted by and during experience. For pragmatists and social psychologists, morality and moral behavior is also such a habit,[14] which is to my mind, very important for a descriptive-normative approach, for example in the field of clinical ethics.[15] Remember the critic of instant intuition I mentioned, which was the concept

[14] For the depiction of morality as a habit, drawing on Dewey's work on human nature and conduct, see La Folette (2000). For moral habits in social psychology see also Hewstone et al. (2007).

[15] For the discussion of this model in the field of clinical ethics, see the discussion in Sect. 15.6, especially Steinkamp et al. (2008).

of intuition on which Moore and Hume relied, when explaining 'the good'. Intuition in the sense of *phronesis* of human expertise in dealing with social, moral dilemmas is not understood as an instant feeling of people who were formerly not confronted by such a situation and who are not experts. Intuition in the sense of Dreyfus and Dreyfus is an a-rational (not: irrational), highly skilled performance, and a rich source of morality out of which an epistemic contextualist gains quite pure water.

Modern Science and Philosophy both heavily relied on mere epistemic, rule based theorizing. This is also the case in the linear applied model of bioethics and in many social science theories and practices. Bent Flyvbjerg analyses the attempt of social science, and most ethical theories relying on an epistemic, rational rule based train of thought, to become an epistemic science like natural positivist science.. This attempt is, to his and to Richard Bernstein's mind, based on the "Cartesian anxiety" (Bernstein 1985: 16) that I also depicted in Parts II, III, and IV: the "fear of ending in relativism and nihilism when one departs from the analytical-rational scientific tradition that has dominated Western science since Descartes" (Flyvbjerg 2001: 25). Bioethics, as Moreno (1999) sees it, is a form of naturalism. And he is right. Bioethics is (still) a mostly biological naturalism as a part of epistemic scientism, explaining and seeing the world in epistemic scientific and philosophical terms and trying to regain some essential certainties through combining epistemic scientific and philosophical expertise that are in danger to be lost in the process of reflexive modernization. It is not true that we had no empirical ethics in the beginning of our discipline. But, in the realm of facts, there was an obvious hierarchy in the classification of facts as being important or irrelevant to consider in bioethics. No matter which debate we look upon in the cutting edge issues of humanity and medicine, where we, as bioethicists, are involved in and in which we are often no longer heretics but opinion leaders: The alliance between mere biological features of humanity and philosophical reasoning in bioethics discussions on a general level has been very strong, as I will depict below.

Third, according to its basis in a pragmatic logic and ethics, a reflexive late modern thinking relies on epistemic and phronetic social science. The research question and contexts determine theories and methods and not vice versa. In empirical research, most social scientists today apply a multi-method strategy, depending on the research question, and consider the reflexive, interrelated epistemological position of researchers of any kind in society. For some research questions, such as problem solving of moral dilemmas in clinical ethics, a qualitative, ethnographic approach is usually more appropriate than epistemic, deductive approaches. Some bioethicists and social scientists argue, however, that only contexts and cases examined through qualitative ethnographic methods can save the life of ethics (e.g. Hoffmaster 1992). But this is just another piece of boundary work, that of heretics who try to replace the orthodox way of finding solutions to dilemmas simply by another absolute way. Also guilty of this are clinicians who conflate empirical ethics with 'evidence based ethics', defining the predominant methodology of evidence based approaches in medicine, epidemiological quantitative data, as the only way of empirical research, and its' results as the only relevant normative basis of ethics. This was aptly criticized by Goldenberg (2005). Yet, e.g. for examining slippery

slope arguments often used in the bioethical and political debate about dilemmas at the beginning and end of human life, deductive, epistemic quantitative methods are more appropriate than qualitative data.

Although the theoretical and empirical framework of a descriptive-normative context sensitive ethics thus needs to be eclectic, it can be described as comprising the following elements of ethical inquiry, closely related to Renée Fox's work in biomedical ethics cited above, which she programmatically summarized in her speech given at the Lifetime Achievement Award at the American Society for Bioethics (Fox 2008; see also Krones 2008a):

1. A transdisciplinary approach of normative-ethical analysis;
2. a conception of theories (norms, principles) as heuristics, which usefulness has to be proven in practice;
3. taking the morals of people (moral intuitions, attitudes, intentions and actions), the 'daily doing of ethics', its preconditions and consequences as a central topic of circular inductive, deductive and abductive (deducting from induction, reproofing of inductive results) forms of inquiry;
4. a conception of human beings as social actors whose actions are both self-determined and shaped by psychological, biological, technological and social forces, as included in the concept of autokoenomia instead of pure auto- or heteronomy, described by Sarah Hoagland, the bio-psycho-social concept of body and disease as first described by George Engel, and the duality of structure and technology as coercion and enabling as described, among others, by Pierre Bourdieu and Anthony Giddens;
5. a generation of prescriptions in participative discourses with ethical norms, principles and values in mind but with the possibility of criticizing orthodox predominant norms, and
6. a fundamental fallibility of consensus and solutions that should always be open to discussion and change.

Such an approach seeks to contribute to both epistemic (Kant 1781/1974, 1798/1983) and phronetic (Flyvbjerg 2001) questions, the phronetic questions being (1) Where are we going? (2) Who gains and who loses, by which mechanisms of power? (3) Is this desirable? (4) What should be done? and Kants epistemic questions (1) What am I able to know? (2) What should I do? (3) What can I dare to hope? and (4) What is the human being?

And now: practice!

15.6 A Descriptive-Normative Transdisciplinary Approach at Work

In this section, four examples of a transdisciplinary approach are given, in which I combine my own experience with other published work in the sense of a transdisciplinary descriptive-normative approach: the debates on the status of the

preimplantation embryo, slippery slope arguments in regard to prenatal diagnosis, decisions made in ethics committees and the situation of practical clinical ethics. In these examples, epistemic and/or phronetic elements are important and different methodological (philosophical and social science) tools applied.

15.6.1 The Status of the Preimplantation Embryo

How we reproduce ourselves has always been both a matter of paramount and practical societal relevance and of the highest personal concern. But with the preimplantation embryo, an even more fundamental aspect of conception is structuring the current body of discourse. Anthropologists have been quick to discover that "beliefs about conception are inseparable from questions about what it is to be human, how a human comes into being and the "miracle" of this creation" (Franklin 1997: 207). In other words, as Malinowski first described in his ethnographic studies, what a society believes about conception can reveal what it believes about everything else and about kinship and gender in particular. Thus, existing images of conception and reproduction reveal what a society believes about gender roles, kinship and genealogical connections, which in turn influence how we interpret what is often conceived as the scientific (putatively objective) observation of the reproductive process. A vast philosophical bioethical literature has been produced on the 'the status of the preimplantation embryo'. In most bioethical contributions to debate, the biological entity of the early human embryo is described and categorized, sometimes in evocative iconographic imagery, in one of two ways. First, by many deontologists and Christian ethicists, as the first (potential) stage of a new human being or even a (early) child with inalienable rights due to its possession of human dignity from the moment of fertilization onwards. Second, by many utilitarian thinkers, as merely a collection of cells that has no high moral status, in which case its manipulation and destruction is unproblematic. Contributions from transdisciplinary approaches in this debate are an excellent example to demonstrate that the thing of logic or science is sometimes not the logic of things. In regard to the beginning of human life, the status of the embryo is discussed in a way, Irma van der Ploeg has aptly described as a deletion and purification pattern of biomedical and philosophical hermeneutics (Van der Ploeg 2004). Biological states and stages (fertilization, syngamy of the two genomes, development of the primitive streak) were scrutinized by biologically well informed philosophers and philosophically well informed scientists and directly linked to moral arguments. The fact that embryos are stemming from somebody and that there are women and men closely involved, bodily, emotionally and existentially in the process of coming into being and in the development of the 'biological' entity embryo, was often considered either as an epiphenomena or as a contamination of the putatively mere theoretical normative and scientific constructs of the status of the embryo as the main aspect to consider in reprogenetics. These connections are of high importance, as several studies and discussions in care ethics (Haimes and Williams 1998; Edwards 1993;

Wiesemann 2006), among these our own work (Krones et al. 2006), demonstrate. We first conducted qualitative research of couples in IVF treatment, high genetic risk couples, and several professional groups, among these ethicists and human geneticists. The embryo itself, contrary to bioethical and biopolitical debates, was not an important topic in most of these interviews. The embryo was seen in connection with the mother, and reflections directed towards the future child, as in other qualitative studies in the field of reprogenetics (Edwards 1993; Franklin 1997). Descriptions were different depending on the status of the embryo per se (more as an abstract object), or the own experienced (IVF couples) or envisioned (other groups) embryo (more as one's own child) was considered. We went on and initiated large quantitative surveys among these groups (more than 800 experts, 500 couples) and a representative survey (n = 1,000) of the German population. We used the categories mentioned in the qualitative interviews. With regard to the beginning of human life, four main categories were considered as important: conception, nidation, fourth month and birth. For the majority of IVF, high genetic risk couples, the general population, gynecologists and pediatricians, nidation, the bodily connection of the early embryo to the mother's uterus after the embryo had been implanted, was considered as the most decisive point in time. (e.g. German population: conception 20.8 %, nidation 46.7 %, 4th month 20.2 % and birth 6.4 %). For human geneticists (45.2 %) and ethicists (65.5 %), conception was the answer most frequently chosen in the representative surveys. Whereas in all groups, religious feelings highly influenced the view that the beginning of human life starts with conception, for human geneticists we did not find such an influence. Contrary to ethicists, human geneticists were most positive as regards to preimplantation genetic diagnosis, and stem cell research. We interpreted these results to mean that both human geneticists and ethicists see conception as the crucial process that makes human life begin as the coincidence of normative and scientific epistemic essentialism. For human geneticists, it is a logical fact not based on the grounds of religious or normative values, that conception is the decisive point in time that human life begins. Many ethicists combine the scientific logic (unique DNA after conception) with normative ascriptions of human life and normative value to the early embryo. Professionals working with pregnant women and children, as well as IVF couples and the general population consider the interactive bodily process of the mother and her embryo as more important. Interestingly, another representative survey on these issues was conducted at the same time in nine European countries, among these Germany. However, unlike our study, the categories used to describe the beginning of human life and the preimplantation embryo were not established through qualitatively ascertained in vivo codes, but were deductively derived from categories provided by scientific and bioethical reasoning (Solter et al. 2003). Accordingly, the survey applied the following categories for the beginning of human life: (1) the moment when the egg and the sperm unite; (2) 2 weeks after conception when different tissues can be distinguished; (3) 3 months after conception when growth of the fetus begins and (4) at the time of birth. The findings of the study are very different from ours in terms of the beginning of life. In the survey of 1,500 Germans, most interviewees said that egg-sperm fusion was the crucial category (38.3 %), followed by 3 months (31.5 %),

2 weeks (15.7 %) and birth (9.0 %). As described above, according to our data, the majority of the German population clearly sees nidation after implantation, being understood as the first bodily connection with the mother taking place about 2 weeks after conception, as the crucial process transforming the biological entity embryo into human life. This view can be interpreted in line with the biopsychosomatic model, which stresses the intersubjective aspect of the beginning of human life. The selection of 2 weeks, defined as the period of time when different tissues can be distinguished, ignores, "deletes" in the above cited sense of Van der Ploeg, the simultaneous and most important process of implantation and nidation, the connection with the mother. Two weeks per se (which is a very important date in regard to discussions up to when embryo research can be considered as permissive, and in which the terminus pre-embryo was defined in UK debates on stem cell research) is not considered as decisive by the Lebenswelt; the connection with the body of the mother after 2 weeks is decisive. As a result of our study, we discussed German legislation and the predominant bioethics discourse, in which the embryo is mostly considered 'on its own', as problematic.

15.6.2 Slippery Slope Arguments in Prenatal Diagnosis

In several areas, among these beginning and end of human life, slippery slope arguments are often used as counter-arguments against developments challenging normative legislative borders. Examples are the debates on physician assisted deaths or on prenatal testing and selective abortion. Implicit eugenic tendencies are postulated to be inherent in these behaviors that may lead to even more eugenic thoughts and behaviors in society. These arguments were central to the statements of the German national ethics commission and the commission on law and ethics in modern medicine of the German Bundestag in their statements on prenatal and preimplantation genetic diagnosis, and were also expressed by Habermas drawing on Hans Jonas' work (Deutscher Bundestag 2002; Nationaler Ethikrat 2003; Habermas 2001). Musschenga, among others, sees a clear need for empirical testing of slippery slope arguments as a hypothesis, also for those defending a linear applied ethics model. He draws on the distinction between a logical, conceptual and an empirical, psychological version of the slippery slope argument. The first version asserts that one cannot draw a relevant logical, conceptual distinction between (acceptable) action A and (unacceptable) action B. The second argument hypothesizes a causal relationship: If one allows acceptable action A, this will causally lead to action B. In regard to the neo-eugenics argument this would, for example, mean that acceptable motivations and actions of prenatal screening and abortion cannot be distinguished (or are the same as) unacceptable ones, and that eugenic motives underlie all decisions and actions of selective abortion. The causal hypothesis of the slippery slope argument predicts that if prenatal screening and selective abortion are offered and used, eugenic tendencies, societal attitudes towards disabled people will become more negative. The so called expressivist argument contains elements of

both, the logical and the psychological argument, asserting that there is an implicit or explicit message sent out by those offering and using prenatal diagnostics, which leads to more negative attitudes towards people with disabilities (Asch 1988; Wendell 1996).

These arguments can be tested. We measured reproductive history including use of prenatal diagnosis and attitudes towards disabled persons in 150 high genetic risk couples, 150 couples with no genetic risk and the general population (n = 1,000) (Krones et al. 2005; Krones 2006, 2008b). In regard to prenatal diagnosis, we found no association between its use, former selective abortion and negative attitudes towards persons with disabilities. However, in the general population, if interviewees expressed very positive attitudes towards prenatal diagnosis and abortion, they were more likely to have negative attitudes towards persons with disabilities. Our analysis therefore, does not support the expressivist argument in so far that users want to send out a neo-eugenic message. Yet, in society, very positive attitudes towards prenatal diagnosis are interwoven with negative attitudes towards people with disabilities, which supports the conceptual version of the slippery slope argument. To answer the causal relationship between more frequent use of prenatal diagnosis and neo-eugenic tendencies, cross sectional surveys like ours cannot be proof. Yet, in a secondary data analysis using a population based data in Germany, the results did not support this hypothesis. The spread of prenatal screening in the last decades (although there is a decline in the last few years) was accompanied by less negative attitudes and behaviors in the German population towards people with disabilities (Van den Daele 2003). On the basis of these results, we argued that the high relevance ascribed to slippery slope arguments by the most influential governmental committees are not warranted, but that pedagogical efforts have to be strengthened to counter attitudinal tendencies in society that link extensive use of prenatal diagnosis with negative attitudes towards people with disabilities.

15.6.3 Decisions in and of Ethics Committees and Commissions

Much bioethics work is done in commissions and committees, such as Institutional Review Boards (IRBs) or national ethics committees, which means, in groups. These groups formulate votes in which they often try to come to a consensus. Sometimes they work under high time pressures or heavy workloads and have a high responsibility on a local or national level. In groups with divergent attitudes, one should expect from rational choice theories and discourse ethics, that group processes lead to careful weighting of pros and cons and average moderate decisions. Empirical research in social psychology indicates that this is often not the case. Groups show phenomena such as polarization. The pre-existing tendency of opinions before a group process is started is often stressed, and opinions are shifting towards the more extreme position, especially under high pressure of time and

responsibility.[16] According to other pieces of socio-psychological research, reflections in groups are often not done in an analytical systematic way, led by goals of contributing to truth, rightness and truthfulness as Habermas postulates in his definition of the ideal discourse (Habermas 1981), but are instead influenced by other motivations. One is impression management, a behavior used by group members in order to belong to the majority opinion of the group, among other mechanisms.[17] As shown in a study of IRB decision-making, risk benefit assessment is often not done in a systematic way (van Luijn et al. 2002). In our studies on reprogenetics, taking place between 2000 and 2004, we used deliberative polls and discourse analysis to find out which way is considered desirable in regard to permitting or prohibiting preimplantation genetic diagnosis (PGD, and who gains and who loses in the debate by mechanisms of power). We found in our surveys that an overwhelming majority of all groups, even the most critical ones (ethicists and midwifes) voted for a legalization of PGD in Germany (e.g. obstetricians 97 %, high genetic risk couples 89 %, general population 88 %, ethicists 68 %).[18] The vote of the most influential committee in German biopolitics in 2000, the Commission on law and ethics of modern medicine was three out of 18 (16.6 %) who voted for a very cautious permission of PGD, all others voted were against it. One could argue that presumably members of this committee were better informed than participants in surveys. Although we constructed a deliberative poll, using information of pros and cons of PGD to inform interviewees before asking attitudinal questions, the better information basis of the governmental commission in regard to bioethical arguments can certainly not be denied. But high genetic risk and IVF couples, obstetricians and human geneticists are also well informed—not so much by bioethical arguments but by their direct experience with dilemmatic situations. And even if we take the argument of divergent votes on the basis of better or less information into account, the difference between the clear votes for legalization of PGD in all groups surveyed, and the clear vote of the governmental commission against it was striking. Group polarization might have taken place in this committee and explain this gap. As part of our discourse analysis, we further asked which groups were over- or underrepresented in the discourse on PGD and also conducted a content analysis of the press. The group that was most clearly considered as underrepresented was the group of directly affected high genetic risk couples. In the press, most articles dealing with PGD (647 in five newspapers) and expressing an opinion voted against PGD. Contrary to that, the majority of letters to the editors were positive. We summarized our results as a democratic deficit in the former discourse on PGD and as a gap between the official opinions uttered in the public domain. Goffman (1959) would say the private opinions of the German population, professional groups and affected people were 'on stage' and on back stage. Here of course, one can apply the argumentum ad populum version of the naturalistic fallacy argument, which was also done in the discussion of our results (Bauer 2005). To assert that one has to transform majority votes

[16] Already shown in the 1960 and after, among these Moscovici and Zavalloni (1969).

[17] Already shown in the 1950s and after, among these Asch (1952).

[18] See above. Another survey in Germany came to the same results (Meister et al. 2005).

directly into ethical or legal prescriptions is certainly criticizable. But one has to at least explain and deliberate upon this obvious divergence of where it might come from and think about the danger of what happens if the rationality and logic of the official bio-political and bioethical debates differ from the rationality and logic of professionals, directly affected patients and the general public.

15.6.4 Clinical Ethics

Clinical ethics is the branch of bioethics most closely related to practice. By clinical ethics, I define the profession that is contributing to problem solving in real cases in hospitals and clinics. One could assume from the history of bioethics that this enterprise is shared by formally trained bioethicists (philosophers, theologians, lawyers, social scientists) and experientially trained clinical experts (physicians, nurses, social workers etc.), contributing to a better deliberation of cases and improved patient outcome. This is not the case. In a survey of ethics consultation in United States hospitals (Fox et al. 2007),[19] individuals performing ethics consultation services were mostly clinical experts (physicians, nurses, social workers), of which only 5 % were trained in a degree program or other formal education in bioethics, and most had learned by doing ethics under supervision by an experienced ethics consultant or only by 'doing ethics'. Less than 4 % of ethics consultants were other groups, consisting of philosophers and theologians. In universities, commissions and among authors of bioethical articles, however, the proportion is vice versa. Is there also stage and backstage in medical ethics with different protagonists acting on stage and backstage-or on different stages? How can these results be explained? And how far deep are they problematic? E.g. Steinkamp and colleagues (2008) draw on the Dreyfus and Dreyfus model of human learning, also underlying Bent Flyvbjerg's model of phronetic versus epistemic approach to science.[20] Whereas at the beginning of human learning epistemic rule based learning prevails, the human experts have incorporated their knowledge while gaining rich experience in mastering or failing in various situations. To deal with ethical issues in commissions or deliberating upon fictional or already solved cases in university teaching sessions is different from being involved in problem solving of real clinical cases in a timely fashion. Since contexts are different, not only practice but also appropriate theories (tools) to deal with these contextual problems are different. In the model of the four principle approach of Beauchamp and Childress, its contents and procedural method were not primarily developed in the contexts of the clinic, but in their highly expe-

[19] I would like to thank Evan DeRenzo for making me aware of this publication and of Steinkamp et al. (2008).

[20] The authors see the Dreyfus and Dreyfus's model as an apt description of "the moral thinking of non-ethicists" (p. 180). I do not think they are right. Clinical ethicists involved in real cases also deal with ethical issues in a non-analytic way-contrary to non-clinical ethicists in the academic field, as I depict here.

rienced work in commissions and university teaching. For clinical ethicists, this method is often considered as useful, but rather at the beginning of the clinical ethical enterprise as ethics first aid, instead of gaining more phronetic experience through dealing with cases (Pullman 2005). Authors highly involved in clinical ethics are developing different frameworks, including the four principles as normative facts, but combine them with different deliberative and behavioral procedures close to clinical decision making and practice (Fletcher et al. 2005; Fins et al. 1997; Richter 2007). In clinical ethics, the psycho-social dimension of decisions is, due to practice, placed at a more central position; it is more a phronetic bio-psycho-social ethics than an epistemic bio-ethics.

According to the model of Dreyfus and Dreyfus, it is of course problematic to start practice without epistemic information. Medical students, clinical ethicists and scholars in commissions need some formal education before they start practicing and some more formal education when they start practicing and become increasingly responsible for their actions as experienced practitioners. Also as skilled experts on the highest (a-rational) level of human learning, new insights from epistemic science have to be incorporated into expert habits on a rule based leaning basis. By starting clinical ethics without important philosophical, juridical and empirical knowledge, one is less likely to reach the main goal of clinical ethics, a better deliberation of cases and improved patient outcome. That this is often not the case in the US and elsewhere (Singer et al. 2001), might first be due to lack of formal teaching of bioethics of clinical ethicists revealed in the survey cited above—but second also due to lack of experience with real cases for those at the front of the academic bioethics debate. Theories can change and improve practice, and practice can change and improve theory.

15.7 Conclusion: Embracing a Transdisciplinary Context-Sensitive Ethics

The concept of a transdisciplinary, context-sensitive, descriptive-normative ethics is unequivocally an answer to the quest for certainty that might be considered to relativize the unity of ethics. Some colleagues will not accept, but vigorously reject this version of ethical inquiry. Yet, I do not see how we can come to a better conclusion on the basis of insights from a contemporary theory of science perspective. Of course, it is much easier (and often more popular) to divide the world into good and bad, black and white, right and wrong, angels and devils. It is very tempting for leaders and their followers to believe in the fiction that a human being or a group of human beings has exactly and really found the truth and knows where the decisive borders are. The kind of solution found by a context-sensitive bio-psycho-social ethics, that takes the diagnosis of reflexive modernity seriously into account, strongly defends the modern Kantian *Sapere aude*! Have the courage to use your own mind and reason! Although it also reminds us that our final quest for certainty will not be satisfied, and the philosopher's stone will not be found by humans in the end.

Part VI
Critical Postscript

Chapter 16
Ethics and Empirical Psychology – Critical Remarks to Empirically Informed Ethics

Antti Kauppinen

16.1 Introduction

The question of whether ethics should be empirically informed has a rhetorical ring to it—how could it be better to be uninformed? Exciting developments in a number of disciplines studying human beings, from psychology and cognitive science to biology, offer hope that ethics, too, could make steady progress were it to hitch its wagons to the train of science. So it is no surprise that some want to erase what they see as outdated and old-fashioned disciplinary boundaries, and no bigger surprise that others react by reaffirming traditional methodologies or by retreating to the grand journals of old. My instinct is on the side of caution in this debate, but I will refrain from grand pronouncements. Disciplinary border skirmishes seem to invite the greatest sin in writing—being boring. In contrast, particular arguments that aim to make concrete progress with existing questions by exploiting a novel methodology can be stimulating even when they go wrong.

So what I will do in this paper is discuss six attempts to draw on psychological discoveries in metaethics and normative ethics. I will focus on psychology, since it is the branch of science that seems to be most closely relevant to ethics. The line between the two disciplines is also particularly porous, which is indicated by the fact that psychology was among the last sciences to gain independence from philosophy. For reasons of space and coherence, I cannot engage much with work inspired by other disciplines, although I believe at least some of the lessons learned from psychology will generalize.

As a general background, I will sketch two opposing philosophical outlooks—one might almost call them philosophical *temperaments*. It is important not to caricature these positions. Moral philosophers have never claimed that empirical facts play no role in ethics. Ancient and Early Modern ethicists and moralists

A. Kauppinen (✉)
Department of Philosophy, Trinity College Dublin, Dublin, Ireland
e-mail: kauppina@tcd.ie

M. Christen et al. (eds.), *Empirically Informed Ethics: Morality between Facts and Norms*, Library of Ethics and Applied Philosophy 32, DOI 10.1007/978-3-319-01369-5_16,

certainly did not shy away from a variety of empirical claims, and though Hume and Kant in very different ways argued for principled limits of what empirical knowledge can do, they did also draw on a particular understanding of human nature in their ethical works. It is true, however, that in the twentieth century, as the human sciences developed their own empirical methods, philosophers did come to focus on questions that could not be settled by empirical research. I will call the view of that emerged *Armchair Traditionalism* and sum it up in two main theses:

1. In *metaethics*, empirical facts are only relevant for causal explanations of particular moral judgments and the capacity to make moral judgments.
2. In *normative ethics*, empirical facts are only relevant for deriving judgments about particular cases from non-empirical principles and for practical recommendations.

Roughly, then, psychology, social sciences, and biology can tell us why and how people make moral judgments, but not what those judgments are or what if anything makes them true. They can also supply material for minor premises in ethical arguments—it is perhaps *a priori* true that creatures capable of pleasure and pain deserve moral consideration, but whether fetuses are sensate creatures is an empirical question. And insofar as ethics is practical, it needs to issue recommendations that are actually useful to people, which means they depend not only on moral facts but also facts about people. For example, even if utilitarianism is the true moral theory, it is going to depend on facts about human beings what decision procedure they should employ to best approximate actions that maximize utility (see e.g. Railton 1984). There is no doubt that if we are interested in promoting moral behavior and moral thinking, or in designing environments that foster moral development and engagement, we need to look to empirical psychology (for concrete suggestions, see e.g. Chap. 7 by Tanner and Christen, this volume; Chap. 13 by Narvaez and Lapsley, this volume). But that is it: the role of empirical facts is *marginal*, not *essential* or *fundamental* to ethical inquiry.

In making the case against armchair ethics, John Doris and Stephen Stich say:

> It is not possible to step far into the ethics literature without stubbing one's toe on empirical claims. The thought that moral philosophy can proceed unencumbered by facts seems to us an unlikely one: There are just too many places where answers to important ethical questions require—and have very often presupposed—answers to empirical questions. (Doris and Stich 2005, 115)

On one interpretation, this claim is not as such incompatible with Armchair Traditionalism. After all, the latter does allow for empirical answers to play a role in causal explanations and derivative judgments, which are responses to "important ethical questions." But Doris and Stich have in mind something more. They think that empirical evidence can settle or at least contribute to resolving metaethical debates and weigh directly against normative theories, such as virtue ethics. This is often because existing metaethical and normative theories make unnoticed and unsupported empirical presuppositions.

I will call the type of view that rejects Armchair Traditionalism in this way *Ethical Empiricism*, distinguishing between bold and modest versions of it as follows:

1. Metaethics
   (a) Bold Metaethical Empiricism: questions about the nature of moral judgment or facts can be answered via empirical study.
   (b) Modest Metaethical Empiricism: empirical results are an important source of evidence about the nature of moral judgment or facts.
2. Normative ethics
   (a) Bold Normative Ethical Empiricism: normative ethical questions are empirical questions.
   (b) Modest Normative Ethical Empiricism: empirical results are an important source of evidence about non-derivative moral truths and/or the empirical presuppositions of normative theories.

Both bold and modest versions of Ethical Empiricist theses reject Armchair Traditionalism. An increasing number of moral philosophers, including contributors to this volume, appear to subscribe to Ethical Empiricism at least in its modest forms. This is not surprising, given the general popularity of methodological naturalism in philosophy, and the initial plausibility of the theses. Yet to properly evaluate Ethical Empiricism, we need to look at concrete arguments and see whether they support the methodological claims.

So without further ado, I will begin with some psychological arguments in metaethics, and then examine the use of psychology in normative ethics. I will be making reference to various papers in this volume, but my discussion will range more widely. My conclusions will of necessity be tentative. Even if no sound argument supporting Ethical Empiricism can be found among the existing efforts I consider (and there are many I have no space to address here), there could always be a different one. The field of empirically informed ethics is still young. But it may be that we can draw some general morals from looking at why the existing proposals fail (or succeed).

16.2 Empirically Informed Metaethics?

Metaethics asks questions about the nature and status of moral thought and talk: Does it purport to represent moral facts or not—that is, are moral judgments cognitive or non-cognitive states? Are there moral facts, and if so, what kind of facts are they? How, if at all, do we acquire moral knowledge? Are moral demands the demands of reason? What does it take to be a moral agent, or a morally responsible agent? These questions are semantic, ontological, epistemological, and broadly metaphysical or conceptual. Some seem clearly out of reach of empirical

science—surely no experiment could settle whether the norms of practical reason and morality coincide. But it is less obvious whether armchair methods suffice for others.

One of the core questions of metaethics, in particular the branch that I like to call philosophical moral psychology, is whether moral thoughts purport to represent the way things are, or whether they are directly action-guiding non-cognitive states, or perhaps some sort of hybrid of cognitive and non-cognitive states.[1] Answers to this question are highly significant for other metaethical issues, such as the nature of moral agency, the function of moral language, and the possibility of moral knowledge. Since this question concerns a crucial feature of moral thought, it is a good test case for the potential relevance of psychological discoveries.

How do we go about answering the question? Consider the traditional armchair argument for non-cognitivism. According to it, when we reflect on moral practice and the distinctive point of moral thinking and language, we discover *a priori* that an intimate link to motivation is essential to moral judgment, since otherwise morality wouldn't be action-guiding in the way it is. This view, which comes in many varieties, is known as *moral judgment internalism*. The next step on the argument is that when we reflect on the nature of psychological states, we learn *a priori* that a mind-to-world direction of fit (tendency of content of the state to match our evidence of the way things are) is essential to belief, and that motivation or action-guiding requires a world-to-mind direction of fit (tendency for the state to move us to change the way things are to match its content) (see e.g. Smith 1987). So, we have an *a priori* argument to the effect that moral judgments cannot be (ordinary) beliefs, and hence consist in some type of non-cognitive or hybrid state. Counterarguments have the same structure—for example, if amoralists, people who make moral judgments without being moved by them, are conceptually possible, the moral judgment internalist premise of the non-cognitivist argument is *a priori* false (and *moral judgment externalism* is true).

This armchair debate has persisted for decades without consensus resolution, although arguably significant advance has been made. The same, of course, could be said about any number of major philosophical debates, so this is not a specific reason to reject the armchair method in moral psychology. But it does provide some motivation to look for an additional source of evidence. I will examine two different attempts to use empirical evidence in resolving the dispute.

16.2.1 From Surveys of Ordinary People to Conceptual Truths

Proponents of the philosophical movement known as experimental philosophy have taken to the streets (or classrooms) to present people with philosophically

[1] When philosophers talk about non-cognitive states, they mean thoughts that do not purport to represent the way things are, and hence cannot be true or false. Paradigmatic examples are desires and affective states. Psychologists often use the term 'cognition' more broadly.

interesting scenarios and elicited judgments (often called 'intuitions') about them. When this method is applied to the case of moral judgment, the argument goes in something like this way, using an argument for internalism as an example:

1. Some philosophical debates concern the extension of ordinary people's concepts, such as the concept of moral judgment.
2. Ordinary people's responses to thought experiments reveal/provide evidence about the extension of the folk concept of moral judgment.
3. The majority of ordinary people's responses are as predicted by moral judgment internalism.
4. Hence, (bold) moral judgment internalism is true / (modest) there is empirical evidence in favor of moral judgment internalism.

For the purposes of assessing the methodology, it does not matter whether Premise 3 is true (the actual survey results conflict with each other). Let us assume, for the sake of argument, that it is the case. Some philosophers would reject Premise 1, and insist that the philosophical debate is about the *nature* or *essence* of moral judgment, which has nothing to do with our everyday *concept* of moral judgment. Perhaps nothing worth calling an intuition plays a role in philosophical methodology (Williamson 2007; Cappelen 2012). Others argue that even if *intuitions* are crucial to philosophical methodology because, for example, they are a source of evidence about modal facts, an intuition isn't the same thing as a response to a survey. Rather, an intuition is perhaps something like an intellectual appearance or seeming (Bealer 2000; Huemer 2001), or a belief or at least an attraction to assent to a proposition that results from mere adequate understanding (Audi 2004; Sosa 2007). Perhaps, as classical rationalists argued, we can have rational insight into the real essences of things. Surveys plausibly do not tap into intuitions in this sense—there is no telling if people's answers are based on intellectual appearances rather than something else altogether (Bengson 2013).

It may well be that these lines of response are more plausible in some philosophical debates than in others. In any case, I will grant that conceptual analysis does have at least some important role in philosophical theorizing, including in metaethical discussion. This means that Premise 2 is crucial for assessing experimental philosophy. At first sight, it seems obvious that ordinary people's responses to scenarios—for example, confident labeling of a subject's mental state as a moral judgment—is evidence about their concept, given that our grasp of the concept MORAL JUDGMENT to some extent guides the way we categorize things. So should philosophers set fire on their armchairs and run out to check whether people think a person who says that stealing is wrong but is not even slightly motivated to refrain from stealing really makes a moral judgment?

Not so fast. To begin with, consider that different people respond differently, yet seem to share the *same* concept, since they apparently *disagree* about its application. If that is the case, there must be a gap between what the shared folk concept (which is the object of philosophical interest) applies to and the way individual users of the concept classify things (which may in itself be of psychological or sociological interest). There are many mutually compatible explanations for the

existence of the gap between the folk concept and the folk's actual classifications. First, as Kripke (1981) emphasized, concepts are *normative*, not descriptive: our concept of addition tells us how we should respond to a calculation task, not how we actually do or are disposed to respond. Sometimes we make mistakes by our own lights—fail to be guided correctly by our own concepts. This may be systematic in some cases, with the result that a majority of people classify things incorrectly. Such tricky or borderline cases are often the most interesting philosophically. Second and related, different people have different levels of *competence* with a concept—some might apply it correctly to paradigm cases, but fare poorly when it comes to the harder ones. The result is that some people's responses may reflect their own shortcomings rather than the folk concept, while others will be more reliable judges.

Third, people's responses might be guided by *non-semantic* considerations. Take the 1868 *Desmond* case discussed by Nadelhoffer (2006). A group of Fenian activists tried to blow up a prison wall to free some comrades, but only succeeded in killing civilians nearby. Clearly, this latter effect was not intended, and probably not even foreseen by the Fenians. Yet when caught, the jury convicted them of murder, which implies intentionality. Nadelhoffer's plausible explanation, supported by his own survey results, is that the jury's willingness to blame the terrorists biased their judgment, leading them to attribute intentionality where none was present—where the folk concept of intentional action doesn't apply.

Fourth, *loose talk* is ubiquitous in non-philosophical contexts. In loose talk, people apply a concept to referents that may fulfill some of the criteria of application but lack some necessary features. For example, people may say "I knew it!" when they've made a lucky guess that turns out to have been correct. Here their belief meets one of the criteria for the application of "knows" (truth) but lacks a necessary condition (non-accidental justification). The same goes arguably for people's willingness to classify a robot that can respond differentially to colors as "seeing" a color (Sytsma and Machery 2010). The robot's circuitry is sensitive to light reflectance (a criterion of seeing) even though it lacks experience with a phenomenal character or the ability to know something (other potentially necessary conditions of seeing), so when people are not particularly interested in speaking literally, and when others can be expected to grasp this, they may well loosely describe the robot as 'seeing red' to convey that it can respond differentially to redness. This is no different from saying that a baby alarm *hears* the baby cry or that the iPad *knows* when its battery is low. We cannot in any of these cases draw conclusions about the concept of seeing or hearing or knowing.[2]

All these caveats mean that ordinary people's responses to cases provide weak evidence about their concepts. The evidence provided by dispassionate armchair reflection or open-minded dialogue will often be stronger (Kauppinen 2007).

[2] Note that I do not claim that the term 'seeing' is ambiguous between an informational and a phenomenal reading. Sytsma and Machery (2010) consider the ambiguity hypothesis, which they regard as *ad hoc* in the absence of an explanation of why the folk would use a different sense than philosophers do, and reject on the basis of their data. The hypothesis that the folk speak more loosely than philosophers do has a high prior probability, so it isn't *ad hoc*.

Whether there is any point in running a survey will depend on the comparative odds of mistakes being made in the armchair or on the streets, which may vary case by case. Here, of course, experimental study can provide evidence one way or another—not about people's concepts, but about how they come to make judgments about certain issues.[3] This type of psychological study will not itself either answer or provide evidence for philosophical questions, but may in principle help identify which responses are good sources of evidence about concepts. As such, it can play a potentially useful auxiliary role in explaining away discrepancies between the folk and philosophers, for example, or even in aetiological debunking of intuitions (see below, Sect. 3.2)—although when it comes to verdicts that have gained broad acceptance among philosophers, a psychologist has a heavy burden of proof to show that they do not reflect conceptual competence.

On the whole, the likelihood that surveys provide useful evidence of folk concepts is low. The odds are that either the outcome is easily anticipated from the armchair, or one or another distorting factor intervenes to produce results that merit no weight in conceptual analysis. Thus, even if this kind of experimental method has some place in the philosophical toolkit, it will be marginal.

16.2.2 From Best Explanation of Data to the Nature of Moral Judgment

A very different experimental approach to metaethics takes its departure from the thought that moral judgment is a natural kind—the sort of thing whose nature or essence can be discovered *a posteriori* by looking at what actually happens in people's minds when they make moral judgments. I will focus on Jesse Prinz's (2007a, b) version of this kind of argument. As I construe it, it involves an inference to the best explanation of observations:

1. Moral judgment is a natural kind whose nature can be found by examining what happens in actual paradigm cases.
2. Psychological and neuroimaging data show, among other things, that manipulating emotions changes moral judgment, emotional activation coincides with moral judgment, and emotional deficits lead to deficits in moral judgment.
3. The best explanation of the data is that moral judgments consist in emotions, which are the best fit for the natural kind that constitutes moral judgment.
4. Hence, moral judgments consist in emotions.

Prinz is clearly committed to something like the first premise, given that he says that the way to avoid the 'impasse' resulting from conflicting intuitions is "turning to psychology and neuroscience, which give us techniques for investigating what goes on in the mind when people are actually engaged in moral evaluation"

[3] This more modest goal is sometimes emphasized by Joshua Knobe (e.g. Knobe 2007).

(Prinz 2009, 702). Is this true? That depends in part on what natural kinds are. There are many ways to think about them. According to one prominent view, deriving from Kripke (1980) and Putnam's (1975) work in the philosophy of language, natural kinds are roughly speaking *a posteriori discoverable microstructural essences*. Water is a paradigm case here: since it turns out, *a posteriori*, that the *actual* watery stuff around us is H_2O, water is *necessarily* H_2O. Roughly, water is *that* stuff there in the rivers and lakes and rain (the term 'water' is a rigid designator); anything that is not that very substance, however similar in superficial properties, isn't water. Hence, the XYZ on Putnam's Twin Earth isn't water.

Another well-known contender is Richard Boyd's view. According to Boyd, when we look for a definition of a natural kind K, we're looking for those commonalities in the causal profiles of the things we classify as Ks that explain our explanatory and inductive success with respect to our term for K (Boyd 2010, 215). Such projectable patterns are *homeostatic property clusters*—sets of properties that reliably co-occur in virtue of some law-like connection, either because the presence of some properties favors the presences of others or because some underlying mechanism favors co-presence (Boyd 1999). According to Boyd's 'accommodationist' semantics, natural kind terms refer to the property clusters that causally regulate their use, even if people have false beliefs about their nature (so that alchemists, for example, succeed in talking about mercury).

There is good reason to think that moral judgments do not form a natural kind in the microstructural sense. We just do not think of moral judgments as psychological states like *that* (pointing to some paradigmatic case of moral judgment), so that nothing that is constituted by the same pattern of brain activation or mental states is a moral judgment. Rather, moral judgment seems to be a *functional* kind: any psychological state that plays a certain functional role is a moral judgment, however it is realized in the mind and brain. In this respect, moral judgments are more like chairs than like water: even if all actual chairs happened to be made of plastic, being made of plastic would be an accidental property of chairs. What makes something a chair is that it's an artifact with a certain practical function. Similarly, even if Twin Earthers have a very different kind of brain and mind from ours, as long as they make judgments that are categorical (apply to agents regardless of their desires or interests), presumptively universalizable (apply to all non-morally similar cases), have felt intersubjective authority, and are somehow linked to non-self-interested sanctioning behavior, to take a few relatively uncontroversial marks of moral judgment, they do make moral judgments.[4] An indication of this is that it is possible for us to *disagree* with them about moral matters, which would not be the case if they were incapable of moral thoughts.

Here is another way to make the case that the concept of moral judgment is not a natural kind concept. This line of argument does not assume that essence must be microstructural, or that MORAL JUDGMENT is necessarily a functional concept (I will use small caps to indicate I'm talking about a concept). Supposed it turned out that

[4] There is now some controversy about this; see Sinnott-Armstrong (2008b), Sinnott-Armstrong & Wheatley (2012) for an argument in favour of disunity of moral judgment.

the psychological states we actually identify as moral judgments only motivate people by way of a desire to look good in the eyes of others. Gunnar Björnsson and Ragnar Francén Olinder (2013), on whose work I draw here, dub this the Cynical Hypothesis. Would its truth mean that internalism is false—or that what we thought were moral judgments were not moral judgments after all? That depends on what kind of concept MORAL JUDGMENT is. Björnsson and Francén suggest it is parallel to TIGER. Take Kripke's (1980) example of the putative conceptual truth that TIGERS ARE MAMMALS. What if it turned out that all animals we actually identify as tigers, or at least the paradigmatic 'tigers', are reptiles? According to Kripke, it would not follow that there are no tigers. Rather, it would turn out that we were wrong about the nature of tigers. Our concept of a tiger is a concept of an animal like *those* (demonstrating paradigm examples of the animal we actually identify as a tiger), whatever kind of animal it turns out to be. (Perhaps tigers need not even be animals.) Björnsson and Francén claim the same goes for moral judgments. If it turns out the Cynical Hypothesis is true, it is not that we don't make any moral judgments, but that we mistook a common correlation between judgment and motivation as a conceptual truth. As they say:

> The cynical hypothesis concerns the actual states of mind that we paradigmatically think of as moral opinions, and it allows that they have almost all the characteristics we normally ascribe to them. They are still categorical, based on familiar moral considerations (e.g. wellbeing, autonomy and respect for rights), often in competition with our prudential considerations, invoked to settle practical issues, and expressed to condemn behaviour near and far. Moreover, people are still affected by moral considerations, some more than others. What is different is just that moral opinions affect action less directly than most of us think. (Björnsson and Francén Olinder 2013, 8)

The other option is that if the Cynical Hypothesis is true, no one makes moral judgments. This parallels the case of WITCH. It is evidently possible for paradigmatic 'witches' or all people we identify as witches to fail to be witches. Why so? Because having supernatural powers as a result of an alliance with an evil it is part of our concept of a witch, and no one has such powers.[5] Why is it part of WITCH? The appealing answer Björnsson and Francén suggest is roughly that there is a certain interest of ours that the concept serves (or served). This is plausibly not just the purpose for which the concept was introduced, but, let us say, the purpose that sustains its use. Having supernatural powers is essential to being a witch, because the point of talking about witches is to identify those with supernatural powers as a result of an alliance with evil. If it turns out no one actually identified as a witch has magical powers, it is not that we were wrong about witches, but that there are no witches at all.

The key question, then, is what interest our concept of moral judgment serves. Would there be a point in attributing people moral judgments if the Cynical

[5] An anonymous referee pointed out that people who self-identify as witches do not think being a witch involves having supernatural powers. Alas, I do not think that believing that one is a witch gives one any special conceptual insight. Indeed, thinking that you are a witch without thinking that you have supernatural powers shows a rather poor grasp of the concept of a witch.

Hypothesis turned out to be true? The internalist will respond: no, it *is* an essential part of the point of talking about moral judgment to distinguish between people who are motivated by what they think is right, as opposed to people who are motivated only by what others think about them. Consider this: why would we introduce in our language an expression for "Martina thinks that X is morally wrong?". Maybe Martina engages in punishing behavior for X, where X involves harming a third party, for example. But why—only because she would be otherwise punished or thought badly of by third parties, or because she thinks X is wrong? The internalist may note that we talk about social norms in the former case. Social norms, after all, overlap with moral norms, and can play the roles that Björnsson and Francén list in the quotation above. They can be categorical (as Philippa Foot (1972) noted, even the norms of etiquette are), promote autonomy, compete with prudential considerations, and so on. For the internalist, the *crucial* difference between moral judgment and socially normative judgment is precisely that the former motivates without regard for and sometimes against what others think. Externalists, too, think that moral judgments motivate by way of something like a desire to do the right thing, and not the (cynical) desire to look good in the eyes of others. So if the Cynical Hypothesis is true and it turns out that states of mind actually identified as a moral judgment only motivate by way of desire to please others, it is not that we were wrong about the nature of moral judgment, but that there are no moral judgments.[6] This, of course, would be a startling discovery, but about human beings rather than about moral judgment.

I do not think this issue can be definitively settled here. All I want to say is that the internalist rejoinder is plausible, and it if is true, it is not an empirical possibility that moral judgments fail to motivate—the empirical possibility is merely that what we actually identify as paradigm cases of moral judgment are not such. We cannot get at the nature of moral judgments by looking at states actually *believed* to be moral judgments, since it may turn out that they are not moral judgments after all. What settles this is an *a priori* investigation into the point of using the relevant concepts. What Björnsson and Francén successfully establish is that *if* that inquiry goes one way, MORAL JUDGMENT is a natural kind concept, and the truth of internalism turns out to be an empirical question. However, I believe that reflection on the point of using the concept supports the opposite conclusion in this case.[7]

What about the Boydian conception of natural kinds? It does appear to be the case that moral thoughts can play a role in explanation and prediction—for

[6] Consider also a Supercynical Hypothesis: not only the states of mind we actually identify as moral judgments not intrinsically motivating, but they are also not in fact based on considerations like rights and well-being, but only what agents unconsciously take to be in their self-interest. Would we still feel the pressure to say that there are moral judgments, but we are wrong about their nature? Why not, if moral judgment is a natural kind whose nature we can identify *a posteriori*?

[7] Mark Alfano pointed out that there is a further possibility I do not consider in the text: reforming our concept as a result of an empirical discovery. I agree that this is a significant option. It might make more sense to modify our concept rather than stop using it, if the world does not cooperate, especially if there is another natural kind in the Boydian sense in the vicinity. Whether this is the case for philosophically interesting concepts remains to be seen.

example, people tend to do what they genuinely think they ought to do, and people tend to think an action is wrong when it involves hurting people they care about. If that's all it takes to form a natural kind, then surely moral judgment is one. One way to see Prinz's argument is as making the case that this natural kind is *constituted* by another natural kind, namely sentiments of approbation and disapprobation. This would explain the empirical observations about emotion, as well as at least many of the other regularities we observe anyway, such as a defeasible link to motivation and tendency for negative judgment when innocent people are harmed, given that both are features of emotional responses. So there is some support for Prinz's constitution claim.

This argument relies crucially on the assumption that the empirical observations (and conceptual platitudes) are *best explained* by taking emotions of approbation and disapprobation to constitute moral judgment. It is thus open to challenge that there is an even better explanation available. I have elsewhere proposed that there is a better candidate: moral *intuition* (Kauppinen forthcoming). As noted above, there is controversy about the nature of intuitions in general, but there is much to be said in favor of thinking of intuitions as *intellectual appearances*: spontaneous and compelling non-doxastic seemings that result from merely thinking about (as opposed to perceiving or remembering) something (see e.g. Huemer 2001). What I have argued is that emotional manifestations of moral sentiments can also constitute intellectual appearances in this sense: when we merely think about taking advantage of someone's disability or disrespecting a national hero, we may have a spontaneous and compelling emotional experience that manifests our disapprobation and presents the action as morally wrong. Such sentimental intuitions can both cause and justify belief (just in the same defeasible way as other intellectual or perceptual appearances do) and motivate us to act. I emphasize that not all moral judgments are based on intuitions: we may also engage in reasoning or simply be disposed to apply rules. This is important, because on my picture, unlike on Prinz's, it is possible (and indeed common) for people to make moral judgments without having emotional responses.

If it is indeed possible to judge without emotion, radical sentimentalist views of Prinz's type are wrong. The crucial test cases here are people with emotional deficits. The most discussed case is that of *psychopaths*. Prinz argues that they can have moral thoughts only *deferentially*, by reference to what other, emotionally typical people regard as right or wrong (e.g. this volume, Chap. 6, p. 101). Yet it is easy enough to imagine a psychopath, or some other emotionally deficient character, making a non-deferential moral judgment and thinking, for example, that everyone else is making a moral mistake. And we ourselves seem to make entirely unsentimental judgments much of the time—although we should take this data point with a grain of salt, given the limits of introspection. Further, *a priori* support comes from considering the conceptual possibility of amoralists, subjects who make moral judgments without any motivation. Insofar as amoralists are possible, there is little reason to think that judgments are constituted by inherently motivating states like the emotions. So, once we distinguish between moral appearances (intuitions) and beliefs (judgments), the best explanation of both the empirical data and conceptual

platitudes is that moral intuitions rather than judgments are sentimental in nature. Premise 3 of the Prinz-style argument is thus false.

So, in short, given that moral judgments do not form a natural kind in the Kripkean sense (MORAL JUDGMENT isn't a natural kind concept), we cannot investigate their nature by observing 'what happens in the head' in the actual paradigm cases. Even if there are natural kinds in the property cluster sense associated with moral judgment, we need to engage in *a priori* reflection to figure out whether they constitute moral judgment or some other associated state. In this kind of reflection we draw on conceptual connections that are *not* discovered *a posteriori*, for example on views about the connection between moral judgment and motivation. Since such key features of moral thoughts are assumed rather than discovered in this empirically informed inquiry, its metaethical scope and significance are limited.

16.3 Empirically Informed Normative Ethics?

As a reminder, these are the Ethical Empiricist theses about normative ethics I want to look at next:

Bold version: normative ethical questions are empirical questions.
Modest version: empirical results are an important source of evidence about non-derivative moral truths and/or the empirical presuppositions of normative theories.

Whatever the status of metaethics, both bold and modest ethical empiricists face the challenge of justifying the move from an 'is' to an 'ought.' This is something that has been attempted in a number of ways. In this section, I will examine one bold and three modest attempts to make use of psychological evidence in normative ethics.

16.3.1 Via Reduction to Normative Conclusions

A radical way of closing the is-ought gap is proposed by Prinz (2007b). As a radical naturalist, he believes that all facts are natural, so "moral facts are natural facts, if they are facts at all" (Prinz 2007b, 3). We can derive moral conclusions from facts whose truth can (at least in principle) be empirically established. To his credit, Prinz lays his cards on the table and gives a very clear account of how he believes this can be done. His example features a character called Smith, whose obligation to give to charity, Prinz claims, is entailed by a set of non-moral premises. Here is his argument (Prinz 2007b, 5):

1. Smith has an obligation to give to charity if 'Smith ought to give to charity' is true.
2. 'Smith ought to give to charity' is true, if the word 'ought' expresses a concept that applies to Smith's relationship to giving to charity.

3. The word 'ought' expresses a prescriptive sentiment.
4. Smith has a prescriptive sentiment towards giving to charity.
5. Thus, the sentence 'Smith ought to give to charity' is true. (2, 3, 4)
6. Thus, Smith has an obligation to give to charity. (1, 5)

The first two premises are surely uncontroversial (provided 2 is read charitably), regardless of what theory of truth is correct, and so is the step from 5 to 6. Premise 4 is a factual stipulation. That leaves Premise 3. Whether it is true is a metaethical question, which I've already argued cannot be settled by empirical study. If that is the case, it's already sufficient to render the derivation non-empirical (while still preserving its status as an inference from an is to an ought). But suppose Premise 3 is true. Does the conclusion then follow? No, because 5 does not follow from 3 and 4.

Why is this the case? Well, if the word 'ought' expresses a prescriptive sentiment, it is surely the *speaker's* prescriptive sentiment. If I say you ought to clean your room, I am expressing, at most, my own sentiment in favor of your cleaning the room. Maybe you do not share that sentiment. No matter. On Prinz's semantics, according to which concepts are psychological entities such as sentiments, my utterance of "You ought to clean your room" still expresses a concept that applies to your cleaning your room. By parallel reasoning, in Prinz's example, it does not matter to the truth of "Smith ought to give to charity" whether *Smith* has a prescriptive sentiment towards giving to charity. Premise 4 is irrelevant.

But whose prescriptive sentiment, then, makes Premise 5 true, if we grant Prinz the rest of his premises? That is a tricky question. Consider first a semantic relativist variant, 5′:

5′. Thus, the sentence 'Smith ought to give to charity' is *true-for-S*.

To reach *that* conclusion, premise 4 would have to be

4′. S has a prescriptive sentiment towards giving to charity.

(Here S may or may not be identical with Smith.) As a relativist, Prinz might be sympathetic to this move. To be sure, it is not clear whether we can make sense of relative truth, though valiant efforts have been made (e.g. MacFarlane 2005). But let us suppose we can. Have we then accomplished the goal of deriving an ought from an is? No, because 6 doesn't follow from 5′ together with 1. 1, the uncontroversial disquotational principle, appeals to *unrelativized* truth. But it does not follow from the *truth-for-S* of "Smith ought to give to charity" that Smith ought to give to charity. After all, whether Smith ought to give to charity is not a perspective-relative fact. Also, given different sentiments on part of some S_2, "It is not the case the Smith ought to give to charity" could be true-for-S_2, so that applying disquotation would give rise to (ontological) contradiction—it being the case both that Smith ought and ought not give to charity.

Premise 6 would, to be sure, follow from the original 5 and 1. But what would make the original 5 non-relatively true, assuming for the sake of argument that 'ought' expresses a prescriptive sentiment? The only plausible candidate is that it is

correct or *appropriate* to have a prescriptive sentiment towards giving to charity. But that is not an empirical fact (to assume otherwise would be to beg the question—the argument is precisely meant to establish that normative facts are empirical). Instead, it is itself a *normative* fact. So, in short, Prinz's argument is either invalid (because Premise 5 doesn't follow from 3 to 4, and if 4 and 5 are replaced by 4′ and 5′, the conclusion does not follow), or involves an 'ought' premise. I do not think there is any way to fix the argument. Bold versions of normative ethical empiricism have little hope of success. But that leaves a number of modest theses that might be viable.

16.3.2 Via Aetiological Debunking to Normative Conclusions

A very different kind of normative ethical empiricist argument has received a lot of attention in recent years. It aims to show that key non-consequentialist beliefs are best explained as the result of emotional reactions, and that their aetiology renders them untrustworthy. Given that we should not base our normative theories on or accommodate untrustworthy beliefs, this shows that we should reject nonconsequentialist ethics. The general form of the argument is the following:

Aetiological Debunking Argument

1. Empirical investigation shows that belief that p results from process X.
2. Process X does not confer justification to/undermines the justification of beliefs it gives rise to.
3. Hence, empirical investigation undermines the justification for belief that p.

As a starting point, everyone but the most hardcore skeptic agrees that some causal processes that result in beliefs are justification-conferring or transmitting. For example, competent logical deduction transmits justification from belief in premises to belief in conclusion. But many of our beliefs do not result from any kind of reasoning. Perceptual beliefs are one paradigm case of such *non-inferential beliefs*. Some say that their justification is exclusively a matter of *coherence*, their fit together with the rest of our beliefs. But pure coherentism seems to sell perceptual beliefs short. Surely their justification has something to do with their causal history as well. Indeed, it seems that perceptual beliefs can be justified in spite of clashing with our prior beliefs. In the absence of a reason to doubt, if I see my Head of Department peel off his skin and reveal the shiny robotic machinery underneath, I should revise a lot of my beliefs rather than reject the poorly cohering perception.[8]

[8] Granted, in extreme cases like this there generally is a reason to doubt and check the initial appearance, as Markus Christen pointed out to me. Nevertheless, perceptions do start out with initial credibility independent of coherence.

To stick with the case of perception, why are (some) non-inferential perceptual beliefs justified? I will focus on just two influential schools of thought. According to one *externalist* view, non-inferential beliefs are justified when they result from a causal process that *reliably tracks the truth*, even if the believer is unaware of this (Goldman 1979; Nozick 1981). According to a recently popular *internalist* view I will call epistemic liberalism, non-inferential beliefs are justified when they are based on *appearances there is no sufficient reason to doubt* (Pryor 2000; Bengson 2010). Internalists often hold that justification has to do with epistemic praise- or blameworthiness, and that there is no reason to blame someone who believes things to be the way they seem to be, if he or she has no reason to doubt the appearances. These two views of justification give rise to different criteria for evaluating processes that result in non-inferential beliefs: they fail to confer justification if they do not reliably track the truth or if they do not involve appearances beyond reasonable doubt.

The specific aetiological debunking argument made by Joshua Greene (2008) and Peter Singer (2005) has this form:

The A Posteriori Argument for Consequentialism

1. Empirical investigation shows that nonconsequentialist moral intuitions* are proximately caused by emotional reactions.
2. Emotional reactions do not confer justification to the beliefs they give rise to.
3. So, empirical investigation undermines the justification of nonconsequentialist moral intuitions*.
4. Nonconsequentialist moral theory rests crucially on nonconsequentialist intuitions*.
5. So, nonconsequentialist moral theory is unsupported by evidence.

(In this argument, 'intuitions' are taken to be spontaneous, non-inferential beliefs rather than intellectual appearances. Since precision is important here, I'll use 'intuition*' to refer to such beliefs to distinguish them from intuitions proper.) To begin with Premise 1, in the background of Greene and his colleagues' argument is a general *Dual Process Model* of the mind. Roughly speaking, the model distinguishes between System 1—automatic, uncontrolled, fast, associative, and often affective processes functioning below the level of consciousness—and System 2, which is conscious, slow, effortful, and capable of reasoning (for a general picture, see Sloman 1996; Kahneman 2011; see also Chap. 7 by Tanner and Christen, this volume). The key empirical data suggest that nonconsequentialist judgments selectively involve the activation of areas of the brain associated with emotion, involve faster reaction times, and go missing in subjects who suffer from emotional defects (Greene et al. 2001, 2009). Consequentialist judgments, in contrast, appear to engage System 2 reasoning. These results and interpretations have been challenged. For example, McGuire et al. (2009) argue that there is no difference between consequentialist and nonconsequentialist responses in reaction times, and Klein (2011) argues that the fMRI evidence does not in fact suggest selective emotional activation in nonconsequentialist responses. And finally, perhaps most decisively, Kahane et al. (2012) find that in cases in which the nonconsequentialist response is

counterintuitive (for example, it calls for speaking the truth to the murderer at the door), it is nonconsequentialist responses that take more conscious effort, suggesting that what engages System 2 is overriding intuitions, not consequentialist rationality.

There is thus plenty of reason to doubt the empirical premise of the A Posteriori Argument for Consequentialism. But suppose there is some truth in it—that emotional responses play a different role in accounting for nonconsequentialist beliefs than consequentialist ones, at least in the Trolley Cases. Premise 2 then becomes crucial. Why does not the fact that thinking about being pushed off a bridge or thinking about pushing someone off a bridge in order to save more people feels bad provide some justification for believing that it is morally wrong? Although some of the things that Greene says suggest that the problem is that it is the mere fact that emotions are involved undermines justification, his considered position is that emotions are *responsive to morally irrelevant factors* (and therefore, presumably, fail to track moral truth). This, of course, breaks down to two claims: emotions are responsive to factors x, y, and z, say, and x, y, and z are morally irrelevant. The first claim is clearly empirical. The second claim, however, is not empirical, as critics like Selim Berker (2009), have pointed out. Its truth must be established the same way as the truth of any other moral claim, perhaps involving appeal to substantive (and controversial) moral intuitions.

But Greene is surely right in responding that while this is true, the scientific data still does important work in the normative argument (Greene manuscript, 9). It may, after all, be a surprising discovery that our beliefs track features x, y, and z. We may, on reflection, agree that x, y, and z are morally irrelevant. In Greene's case, the factor he sees as crucial to explaining people's responses is the *use of personal force*. As he notes, it is not question-begging for a consequentialist to take this to be morally irrelevant: "Whether your normative proclivities are consequentialist, deontological, or otherwise, it's hard for you to argue that personal force is morally relevant." (Greene manuscript, 17) It is thus very plausible that psychological processes that track the use of personal force do not track moral truth, and the beliefs that are their outputs lack justification in the externalist sense. (Insofar as a subject is *aware* of what underlies her responses, she presumably lacks justification in the internalist sense as well.)

In support of Premise 2, Greene (manuscript) further argues that it is likely that emotions will be responsive to irrelevant factors, especially in novel situations. The distal explanation of why we have particular affective responses is that they have been, on the whole, fitness-enhancing in the course of human evaluation. It pays off, as a rule, for us to be afraid of big things moving fast toward us, since most such things were (and are) dangerous. But this response will sometimes misfire, especially in evolutionarily novel situations (the subway train will not leap off its track to pounce on us). Similarly, the Greene/Singer hypothesis is that evolution has favored the development of negative emotions to using up close and personal violence. Such innate aversion is fitness-enhancing for some reason (presumably it reduces interpersonal conflict). Violence (or assistance) at a distance, however, was not an issue during the era of human evolutionary adaptation. Consequently, our

automatic, 'point-and-shoot' moral emotions are likely to misfire in modern, complex, or unusual situations—to fail to respond to morally relevant factors.

This is an impressive line of argument. If the aetiology of beliefs is relevant to their justificatory status, then surely empirical study of the aetiology can in principle reveal that they lack justification. But I do want to raise three concerns with Greene's case: not all emotions are created equal; intuitions aren't so easily done away with; and what counts at the end of the day is not whether particular individuals are justified but whether justification is available for nonconsequentialist beliefs. Before I go into these, however, I want to register some doubts about an approach that has gained popularity recently. According to this type of response, emotional intuitions can be reliably truth-tracking in just the same way as *expert intuitions** in general (see Chap. 7 by Tanner and Christen, this volume; Chap. 11, by Musschenga, this volume; Chap. 13 by Narvaez and Lapsley, this volume; Allman and Woodward 2008). Expert intuitions* are, roughly, spontaneous judgments that result from automatic, System 1 processes that respond to environmental cues that the subject is not consciously aware of, but are nevertheless reliable. Paradigmatic examples are quick situational assessments by chess masters and experienced nurses or firemen: without knowing just why, the fireman feels that the building is about to collapse and reacts to save himself at just the right time. If moral intuitions* of at least some people were of this type, there would be no reason to suspect them.

Alas, contrary to optimists, they cannot be. As an authoritative recent overview (Kahneman and Klein 2009) argues, there are two conditions for the development of intuitive expertise or implicit learning. First, the environment must exhibit regularities that the associative System 1 can latch onto. This may or may not be the case for morality in general, but surely will not be for outlandish philosophical thought experiments. Most importantly, however, training System 1 requires "prolonged practice and feedback that is both rapid and unequivocal" (Kahneman and Klein 2009: 524). A nurse who diagnoses and treats a baby will typically be able to check whether the baby's condition is improving (temperature returning to normal etc.), and thus gets feedback on the correctness of the diagnosis. There is nothing analogous to this in the case of moral judgment. Even if there is a recurring type of moral problem, there's no rapid and unequivocal indication that a subject is judgment is on the right track. If you judge that abortion is wrong even if it is not and act on your belief, there is no negative feedback that results simply from your having made a moral mistake. (The only reliable negative feedback you will get for acting on a moral judgment is from people who disagree with you, but that is not an indication that you are wrong.) So we cannot train our intuitive system to respond to moral truths in the same way we can train it to respond to truths about good chess moves or ill infants. The expertise defense of moral intuitions* is unsuccessful.[9]

[9] To be sure, I do not mean to deny that there can be moral expertise in some meaningful sense—some people are better at articulating principles, more consistent, better informed about pertinent non-moral facts, and so on. Perhaps it is even the case that their judgments should be privileged in reflective equilibrium, as Musschenga argues (Chap. 11 this volume). But nonconsequentialists cannot defend intuitions* on these grounds.

If the expertise defense will not work, how can nonconsequentialists respond to Greene's challenge? To begin with the first option I mentioned, Greene stakes a bold claim about nonconsequentialist intuitions*: "All of the factors that push us away from consequentialism will, once brought into the light, turn out to be things that we will all regard as morally irrelevant." (manuscript, 21) So when we trace down the aetiology of any nonconsequentialist intuition*, we always hit an affective reaction that is caused by a factor that is, on reflection, morally irrelevant. However, it is one thing to say that some morally relevant emotions are triggered by simulating the use of personal force or some other morally irrelevant factor, and another to say that *all* are. For example, it is extremely plausible that we have a negative emotional response, such as resentment, to being used as a mere means by someone else, as well as a weaker sympathetic response to imagining ourselves in such a position. Such reactions are also almost certainly fitness-enhancing, at least in the personal case—they motivate retaliation and decrease the likelihood of being exploited in the future. Being used as a mere means, in turn, is not uncontroversially a morally irrelevant factor—to claim otherwise is to beg the question in favor of consequentialism. This means that at least some emotions are responses to factors that are plausibly morally relevant. Note also that there is a long tradition of sentimentalist ethics arguing that such reactions need not be rooted in an egocentric perspective, but can also be felt from what Hume called the 'Common Point of View' and Adam Smith called the impartial spectator's perspective. I argue elsewhere that precisely such impartially empathetic emotional responses constitute canonical moral appearances or intuitions (Kauppinen forthcoming).

Sentiments felt from the Common Point of View are far from the kind of automatic gut reactions that Greene discusses. They are not or need not be quick and unreflective, evolutionary fitness-enhancing, or responsive to features that are uncontroversially morally irrelevant. So insofar as nonconsequentialist moral judgments are based on *that* kind of emotional intuition, there is no obvious reason to think they lack justification. From this perspective, Greene's problem is that he works with a palette that is too narrow: it is either reasoning or gut reaction, and nothing in between.

Of course, it remains to be shown that at least some nonconsequentialist judgments result from the better kind of emotional response. The current data does not settle the issue even concerning the Trolley Cases. Although people are more likely to condemn the agent who pushes a fat man down (where there is both personal force and use as a means) than an agent who drops the fat man through a trapdoor (where there is use as a means but no personal force), they are nevertheless more likely to condemn the latter than an agent in the standard Switch cases (where there is neither personal force nor use as a means) (see Greene et al. 2009). So *use as a means* has an effect independently of personal force. Indeed, one possible explanation for why the use of personal force plays a role may be that it raises the *salience* of the use as mere means (cf. Chap. 6 by Prinz, this volume, p. 106). Moreover, many philosophers report the intuition that the trapdoor drop is wrong, as well as intuitions about other more fine-grained scenarios. These are unlikely to be mere gut reactions, since they are reflectively stable. But they may well be the good kind

of sentimental intuitions I talk about. We may not be able to assess their *reliability* in a non-circular fashion (we will have to assume that using someone as a mere means is wrong, for example), but we can at least say that they're *moral appearances* we have been given no reason to doubt.[10]

The second problem is that reliance on intuitions may be unavoidable. Greene insists that in the psychological sense of "intuition" (by which he means judgment resulting from unconscious, automatic process), "Consequentialism can do just fine without intuitions" (manuscript, 20). But this seems inconsistent with Greene's own acknowledgement that the source of evidence for the moral irrelevance of the use of personal force is "substantive moral intuitions" (manuscript, 7), unless of course the substantive moral intuitions* are not intuitions in the psychological sense. But consequentialism does seem to rely on precisely the same sort of intuitions* (in the psychological sense) as nonconsequentialism. For example, we judge that in Trolley Cases, the "body count" is not morally irrelevant (for consequentialists, it is the only relevant feature). But why? Is it not also an evolved emotional reaction to prefer fewer deaths to more deaths? Surely it is. But if point-and-shoot emotions are unreliable for principled reasons, then so is the core utilitarian intuition*. If the positive response to maximizing is what I have called the good kind of emotional intuition—which I think is likely—then it does have justificatory force, but so do at least some nonconsequentialist intuitions. There is no dialectical advantage here for consequentialism.

The third and final point is that for some purposes, crucially including the choice of which normative theory to accept, the justificatory status of particular individual beliefs does not matter. Those who accept Premise 4 of the A Posteriori Argument for Consequentialism may grant that most people's nonconsequentialist beliefs are based on knee-jerk reactions that undermine their justification, while insisting that genuine intuitive propositional justification is *available* for nonconsequentialist beliefs. That is all that is needed to justify nonconsequentialist theory. Some Kantian nonconsequentialists reject the premise altogether (e.g. Wood 2011). If there is rational justification available for nonconsequentialist beliefs, it again does not matter if *most people* believe the right thing for the wrong reasons. Suppose, for a parallel, that most people believed the Earth is round because a holy book written thousands of years ago happened to say so, without any scientific evidence. That would hardly be relevant to whether *I* or the scientific community in general should accept or reject that the Earth is round. Similarly, Premise 5 does not follow even if people in general lack justification for nonconsequentialist beliefs.

In short, although it is in principle possible that empirical evidence concerning aetiology would undermine the justification of some moral beliefs, the path is far from straightforward. Merely showing that some judgments are intuitive does not

[10] I argue elsewhere that we do have a non-question-begging way of evaluating whether certain kinds of intuitions are trustworthy. This involves appealing to the practical function of making moral judgments, roughly making peaceful social relations possible without a Hobbesian sovereign ruling by force, and noting that being guided by intuitions felt from the Common Point of View is reliably conducive to that goal.

suffice, and for some crucial purposes, such as choice between moral theories, it does not even matter whether most people are justified in believing one way or another.

16.3.3 Via Ethical Conservatism to Normative Conclusions

Shaun Nichols, Mark Timmons, and Theresa Lopez develop a novel modest ethical empiricist argument developed in their contribution to this volume. They argue, first, that many of our central ethical commitments cannot be rationally justified, but result from "a-rational and a-reliable emotional processes" (this volume, Chap. 9, p. 160). But some of such commitments nevertheless have normative authority, which presumably entails that the subjects are justified in believing in their contents. This seems to be the structure of their argument:

1. Entrenched ethical commitments have normative authority in spite of resulting from non-rational and non-truth-tracking emotional processes (Ethical Conservatism)
2. Empirical study can identify which commitments are entrenched.
3. Hence, empirical study can identify which ethical commitments have normative authority.

If empirical study can establish which commitments have normative authority, it surely has more the marginal significance for ethics. So this is an interesting new line of argument.

For a commitment to be *entrenched* is for it to be non-inferential and the result of natural human emotional reactions (or at least resonate with such reactions). It seems plausible that empirical study can indeed establish which commitments are entrenched in this sense, as Premise 2 says, and do so better than armchair reflection. Nichols, Timmons, and Lopez provide an example of how to do it with their studies of outcome-dependent blame, which suggest that even if intention and reasons for action are held fixed, people regard an agent as more blameworthy if the outcome is bad, as long as the agent has been negligent. For my purposes, the details and the soundness of this argument do not matter.

The definition of an entrenched commitment appeals to natural human emotional reactions. I take it that 'natural' here means being part of the normal human biological makeup. A number of contributors to this volume argue, in line with much recent biological research (e.g. de Waal 1996), that some morally relevant emotions are indeed natural in this sense. For example, Van Schaik et al. (Chap. 4, this volume) note that humans, unlike other primates, engage in prosocial behaviors not only reactively—in response to need, proximity, or the presence of an audience—but also proactively, as seen in our tendency to cooperate and share in economic games. Why? Crudely, as the kind of foragers we are, we have to cooperate with each other to survive. As cooperative breeders, we have a tendency to respond to need and conform to expectations; as cooperative hunters, we also have a tendency to match

rewards with contributions and build a reputation as reliable reciprocators. Van Schaik et al. hypothesize that these four psychological elements—sympathy, wish to conform, sense of fairness, and concern with reputation—are "the major components of human moral psychology, upon which our reflective morality is built" (this volume, Chap. 4, p. 77). They suggest that moral emotions are "the subjective side of the evolved proximate regulators of human cooperation" (p. 77); see also Naves de Brito (Chap. 3, this volume). They are likely to emerge early and cross-culturally, and will be to an extent independent of conscious control.

As Jesse Prinz (Chap. 6, this volume) points out, even if morally relevant emotions are natural in this sense, it does not mean that our capacity to make moral judgments or tendency to adopt certain moral rules is an evolutionary adaptation. After all, other species that have similar responses and behaviors (see Chap. 5 by Brosnan, this volume) plausibly do not make moral judgments. Prinz's suggestion is that the human capacity to make moral judgments is an evolutionary byproduct of putting together capacities that are adaptations for other purposes, including imitation and capacity for abstract thought in addition to prosocial and reactive emotions. Support for this hypothesis can also be found in neuroscience, if, as Prehn and Heekeren (Chap. 8, this volume) argue, "the "moral brain" can be broken up into several modules whose functions originally have nothing to do with morality (emotion, social cognition, cognitive control, etc.)." (p. 156)

Biological considerations thus support the hypothesis that some moral commitments are entrenched, and indeed provide clues about which commitments are likely to be such. I am not going to take issue with the psychological part of Nichols, Timmons, and Lopez's Chap. 9 regarding which commitments are entrenched. The important question concerns the *epistemic standing* of entrenched commitments. Precisely what does normative authority mean in this context, and why should entrenched commitments have it? To begin with the former, the parallels that Nichols, Timmons, and Lopez draw between entrenched commitments and other beliefs suggest that they think there is *no reason to suspend* beliefs that have normative authority. This may or may not mean that the beliefs are *justified*—perhaps there are reasons not to suspend beliefs that are independent of their justification. Unfortunately, the epistemic part of the paper is extremely sketchy, so it is not possible to determine what the exact view is. In any case, at the end of the paper, Nichols, Timmons, and Lopez offer a further suggestion: some commitments may be entrenched yet biased, in which case they lack normative authority. They argue that bias can be exposed by seeing "whether people withdraw their judgments under full information" (p. 173).

Why should we not suspend entrenched commitments, even if they are not truth-tracking, and even if we know this? Nichols, Timmons, and Lopez offer two suggestions. The first appeals to the *undesirable consequences* of suspending entrenched commitments: "If we give up all of the ethical judgments that critically depend on our a-rational and a-reliable processes, then we might well be left with an ethical world view more barren than almost anyone is willing to accept." (p. 160). This appears to suggest a *pragmatic and non-epistemic* reason for maintaining entrenched commitments: they are not epistemically justified, but if we give them up, we are

left with a barely recognizable ethical outlook, which is a bad thing (at least from our current perspective). How dramatic the change would be depends on how important entrenched commitments are to our actual ethical outlook. In any case, from the perspective of an ethical theorist, the pragmatic argument is extremely weak. If the truth is that most or all of our current ethical beliefs are unjustified, then that is the truth, however unpleasant and hard to accept it is. Error theorists in metaethics are in fact quite happy to accept this, and have argued that evolutionary influences on our moral judgments do warrant such global moral skepticism (Joyce 2006).

The second suggestion that Nichols, Timmons, and Lopez make draws on an analogy with aesthetics. They claim that "Finding out that one's aesthetic tastes (and related judgments) in music are grounded in a-rational and a-reliable mechanisms is not itself a good reason for rejecting those tastes and related judgments" (footnote 2, p. 161). If ethical judgments are relevantly similar, the same goes for them. But there is much reason to doubt this. The reason why ungrounded judgments about music, for example, are relatively immune to rejection is either that there is no fact of the matter or that the facts are relative to individual subjects' tastes (in which case taste-based judgments are automatically truth-tracking and hence justified). I will not rehearse familiar arguments against moral nihilism or relativism here (see e.g. Shafer-Landau 2003). Suffice it to say that there is not much point in normative inquiry of any sort, empirically informed or not, if there are no objective facts of the matter. And why would a commitment have normative authority if any contrary judgment would be just as justified? Normative authority is precisely what ungrounded aesthetic judgments lack—for example, I have no reason to resist acquiring a new taste in ice creams, since liking pistachio would be just as unproblematic as liking chocolate.

So far there is little reason to regard entrenched commitments as prima facie justified or authoritative. Indeed, there is some positive reason to doubt this. Suppose it turns out to be an entrenched commitment, at least for some people, that homosexuality is morally wrong. This is not implausible, and certainly not impossible. Should we then regard belief in the wrongness of homosexuality as prima facie justified, or authoritative for those who hold it? I do not think so. Ethical conservatism threatens to become conservative ethics. Further, the natural emotional reactions underlying entrenched commitments can *conflict*. As Van Schaik et al. point out, sympathy for someone's suffering can conflict with the sense of fairness. Perhaps the person is starving because he did not bother to go on a hunt when everyone else did. If caring and justice are equally entrenched, which side has normative authority in the case of conflict? If it is both, how do we decide between the claims?

So my first problem with ethical conservatism is that entrenched commitments do not, as such, seem to merit normative authority. The second issue is that it is not clear why we should think of entrenched or other emotionally driven commitments as *unreliable* or *non-truth-tracking* in the first place. (Nichols, Timmons, and Lopez use the word 'a-reliable', but there's nothing else it could mean.) As I argued in the earlier sections, there are other ways of privileging certain emotional responses in ethics. It may well be that informed, impartially sympathetic emotions track moral

truth, either because moral truths simply are truths about how we would respond, were we to be impartially sympathetic and informed, or because they just happen to tap into mind-independent moral facts. It is, for example, morally wrong to rape a child or knowingly sell a faulty product. Most of us have a non-accidental negative emotional response to raping a child or knowingly selling a faulty product. These responses, then, appear to track at least some moral truths. To establish that they are reliable, we would naturally need to tell much more of a story of how they not only accidentally coincide with moral facts. I will not attempt to do so here. In any case, my bet is that when we have fuller story of which ethical emotions are trustworthy, their being *entrenched* will turn out to play no role in it. Thus, even if empirical research can establish which commitments are entrenched, that discovery will not provide evidence for or against normative views.

16.3.4 Via Psychological Unfeasibility to Normative Conclusions

The final kind of normative argument based on empirical psychology that I want to consider is relatively old. It takes its point of departure from the thought that ethics is for human beings, and thus has to take into account human cognitive and motivational limitations. Moral ideals and demands have to be *psychologically feasible* for the kind of beings we are. This constraint on moral theories is closely related to the old thesis that 'ought implies can'—it cannot be the case that morality requires people to do things they are unable to do, because it would be wrong to blame them for failing to do the impossible. There are deep questions concerning these constraints—What exactly does it mean that someone is psychologically unable to do something? Are there normative demands that do not imply an ought or blame for failure? —but I will assume here that they are along the right lines. This opens up a different kind of potential role for empirical psychology. Since it is an empirical question what human abilities are like, scientific psychology can in principle lead to new normative insights.

The best-known recent argument along these lines is the situationist attack on virtue ethics, in particular its focus on becoming a certain kind of person with certain character traits. Its structure is basically as follows:

1. Virtue ethics tells people to cultivate robust character traits.
2. Most people's behavior varies in response to contextual factors, including very minor ones.
3. Behavioral variance in response to minor contextual factors is inconsistent with the common existence of robust character traits.
4. So, empirical evidence shows robust character traits are, at best, rare/the existence of robust character traits is not empirically supported. (2, 3)
5. An ideal that most people cannot live up to is not psychologically feasible.
6. So, the virtue ethical ideal is not psychologically feasible. (1, 4, 5)

7. A moral theory whose ideal is not psychologically feasible should be rejected. (The Feasibility Constraint)
8. Hence, virtue ethics should be rejected. (6, 7)

In premise 1, robust character traits are "dispositions that lead to trait-relevant behavior across a wide variety of trait-relevant situations" (Doris and Stich 2005: 119) or "relatively long-term stable disposition[s] to act in distinctive ways" (Harman 1999: 317). For example, honesty is a disposition to be truthful and forthcoming in a wide variety of situations in which there might be something to be gained by deception. The perhaps counterintuitive Premise 2 is supported by a large number of social psychological studies that have found, among other things, that people's helping behavior systematically varies due to contextual factors like mood, hurry, and the presence of others, and that a large majority of subjects are willing to hurt others under minor social pressure (for thorough overviews, see Doris 2002; Alfano 2013). Premise 3 draws on the idea that if people had robust character traits, their behavior, especially in such morally relevant cases, would vary from person to person, depending on how virtuous they were. But in fact it seems that it is the situational features rather than people's dispositions that seem to account for manifest behaviors. The remaining steps draw out the conclusions: at most few people seem to have robust character traits. There are, at best, fragmentary character traits like "office-party-temperance" (Doris 2002) that are nothing like virtues. So the virtue ethical ideal is unfeasible and should not be adopted.

In response, virtue ethicists have typically attacked Premises 3 and 5 instead of rejecting the Feasibility Constraint. The first line of defense begins with the rejection of the understanding of character traits that underlies the situationist attack. Character traits are not dispositions to *act*, it says. Rather, they are in the first instance dispositions to perceive, feel, and reason in certain ways, and consequently, perhaps, to act. There is a gap between manifest behavior and character traits (see e.g. Sreenivasan 2002). Perhaps, as Julia Annas (2011) maintains, they are akin to *skills*. This complicates the task of showing the non-existence of traits, since mere behavioral evidence is not sufficient. So, for all the current evidence shows, people may after all have robust character traits. A weakness of this response is that if people's perceptions of reasons do not make a difference to how they act, there is not much reason to focus on them in ethical theorizing. Nor do those who take this line of response typically provide positive empirical evidence for the existence of character traits (although see Russell 2009).

The other main line of response is to grant that virtue is rare, but nevertheless an attainable or at least practically useful ideal (e.g. Appiah 2008, 47)—in my terms, to deny Premise 5, the notion that a psychological ideal few can live up to is psychologically unfeasible in the relevant sense. In their rejoinder, Doris and Stich say that "if virtue is expected to be rare, it is not obvious what role virtue theory could have in a (generally applicable) programme of moral education." (Doris and Stich 2005, 120) This is a weak objection for many reasons. First, it assumes that it is an important standard for assessing normative theories is whether they serve practical didactic aims. This surely need not be the key aspiration of any normative theorist. Second,

rarity and difficulty of attaining an ideal do not in any obvious way render it didactically obsolete. Suppose it's very rare to anyone to play guitar as well as Mark Knopfler (as it is). Does that mean it is a bad idea to try to play like Knopfler, when you are practicing to become a better player? Hardly.

Mark Alfano (2013) has a different objection to the virtue-as-an-ideal response. He notes that among the hard core of virtue ethics are claims about the explanatory and predictive power of character traits, as well as what he calls egalitarianism (almost anyone can reliably act in accordance with virtue) and cross-situational consistency in response to reasons. If virtue is hard and rare, Alfano says, "the virtues are loose cogs in our motivational machinery, reliably licensing neither the explanation nor the prediction of behavior" (Alfano 2013: 63). This rejoinder illustrates the common mistake of treating virtue as an all-or-nothing property. It is, however, much more natural to think of virtue as a matter of degree. We can be more or less honest or chaste—that is to say, roughly, we may be more or less sensitive to reasons for truth-telling or abstinence.[11] The truth of positive virtue attribution will depend on the context (a chaste French politician does not cheat on his mistress), much as the truth of other utterances containing scalar adjectives (such as 'tall') does.

The empirical evidence certainly suggests that we may possess such traits to a lower degree than we like to think, so that most of us perhaps cannot, in most contexts, truthfully be described as brave or just, period. But that is to say we are *to some degree* brave or just, so that our behavior may be to some extent be explained and predicted by reference to bravery or justice. Almost everyone can become *more* virtuous, and the more they approach the ideal of the *phronimos*, the more the attribution of virtue traits will explain and predict their behavior. That is how thinking of virtue as gradable reconciles the virtue-as-an-ideal line with explanatory/predictive power and egalitarianism.

Edouard Machery (2010) has recently developed the situationist critique further. As he sees it, the real problem is that virtue ethical ideals presuppose *unified agency*. By this he means that…

> …the psychological causes that are meant to constitute our character and the kind of person we are (our values, desires, norms, emotions, etc.) have a specific causal structure: They (or at least many of them) are unified. That is, they are causally influenced by a common cause or they causally influence one another. (Machery 2010, 225)

[11] Following a lead from Robert Adams (2006), who in turn draws on the old distinction between imperfect and perfect duties, Alfano notes that some 'low-fidelity' virtues, such as generosity, require one to be responsive to some occasions in which giving is called for, while other 'high-fidelity' virtues, such as chastity or justice, require a high degree of consistency—to possess them one has to respond suitably nearly every time. I do not think this is the same dimension I am talking about. The degree of virtuousness is not identical with frequency of acting on a certain kind of reason. You do not have to be very chaste to refrain from sleeping with someone other than your partner 100 % of the time, because the reason to do so is strong. (Insofar as chastity is a virtue, the degree to which it is possessed is manifest in the subtle ways one interacts with attractive non-partners.) Hence, even a low degree of chastity explains and predicts full faithfulness in deed. At the other end, even the most perfectly generous person will not give on every occasion, as the contrary demands of justice, friendship, and other virtues intervene, and the strength of her reasons to give diminishes the less she has to give or more she deprivation she herself suffers.

Machery then argues that human agency is not unified in this sense. He draws on Dual Process Models and research on implicit biases, which suggests that people's conscious values often come apart from their automatic responses. But why is the potential, and indeed frequent disunity between System 1 and System 2 processes a problem for virtue ethics? The reason Machery gives is that "we have no direct control over some psychological causes—namely over the automatic systems—suggesting that it might be difficult to bring them in step with the other states and dispositions that are meant to constitute character." (Machery 2010, 227) But this lack of control, surely, does not come as a surprise to the virtue ethicist. Aristotle, after all, is explicit that acquiring virtue is slow work and significantly subject to moral luck when it comes to having the right sort of temperament, teachers, and environment. What is more, this still looks like a version of the difficulty challenge. So even if Machery is right about the disunity of agency, that does not seem to pose a new problem for the virtue ethicist.

This substantial response leaves Machery's methodological challenge intact, however. He argues that "the proper response to the situationist threat involves examining the empirical literature on agency in detail. There is no easy way for moral philosophers out of a laborious study of human behavior." (Machery 2010, 227) So any defense of virtue ethics must be empirically informed to be credible. To be sure, insofar as we accept Ought Implies Can or Feasibility Constraint, it is hard to deny that empirical facts about human agency potentially undermine character-based ethics. But I still want to reject Machery's methodological thesis. I believe the burden of proof here is on the critic who denies the commonsense view of character that virtue ethics relies on. That is, it is not that the virtue ethicist has to dig through empirical literature to show that courage or kindness is possible (even if rare). Recall the point I made above: the core empirical assumption is not that some or many people are perfectly courageous or kind, but that *people are more or less courageous or kind*, and that most of us can improve in these respects. For all the evidence situationists have presented, we still have no good reason to believe this is false.

16.4 Conclusion: Building a Better Armchair

I have charted various ways in which empirical psychological results might be or have been claimed to be important to metaethics and normative ethics in ways that go beyond Armchair Traditionalism. I believe that we have not been given any good reason to believe in bold versions of Ethical Empiricism. Neither metaethical nor normative questions are empirical questions, or questions that could be settled by empirical findings. I have also found various Modest Ethical Empiricist arguments wanting. Generally, the empirical evidence does not do the work it is alleged to do, or provides weak support for one view or another only under strong non-empirical assumptions. Too often, empirical information is noise that distracts from the core issues.

Nevertheless, I cannot claim to have vindicated Armchair Traditionalism either. I have left the door open for the possibility that empirical discoveries may help in conceptual analysis (although only indirectly) and that they may help identify what natural kinds constitute moral thoughts (although the actual identification draws crucially on armchair reflection). I have also allowed that normative ethics may yet benefit from understanding the roots of our intuitions and the feasibility of ethical ideals, even if the existing claims are exaggerated. Perhaps the best overall conclusion to draw is that while armchair reflection will and ought to continue to be central to ethical inquiry, findings about what, why, and how we judge may stimulate and even challenge its results at several important junctures.

References

Adams, R. 2006. *A theory of virtue*. Oxford: Oxford University Press.

Adler, J.E. 1996. An overlooked argument for epistemic conservatism. *Analysis* 56: 80–84.

Alexander, R.D. 1987. *The biology of moral systems*. New York: Aldine de Gruyter.

Alfano, M. 2013. *Character as moral virtue*. Cambridge: Cambridge University Press.

Allman, J., and J. Woodward. 2008. What are moral intuitions and why should we care about them? A neurobiological perspective. *Philosophical Issues* 18: 164–185.

Ammann, C. 2007. *Emotionen – Seismographen der Bedeutung. Ihre Relevanz für eine christliche Ethik*. Stuttgart: Kohlhammer.

Amodio, D.M., and C.D. Frith. 2006. Meeting of minds: The medial frontal cortex and social cognition. *Nature Reviews Neuroscience* 7(4): 268–277.

Anderson, J.R. 1981. *Cognitive skills and their acquisition*. Hillsdale: Erlbaum.

Anderson, J.R. 1990. *Cognitive psychology and its implications*, 3rd ed. New York: W H Freeman and Company.

Anderson, S.W., A. Bechara, H. Damasio, D. Tranel, and A.R. Damasio. 1999. Impairment of social and moral behavior related to early damage in human prefrontal cortex. *Nature Neuroscience* 2(11): 1032–1037.

Annas, J. 2011. *Intelligent virtue*. Oxford: Oxford University Press.

Apel, K.-O. 1973–1976. *Transformation der Philosophie*. Frankfurt am Main: Suhrkamp.

Appiah, K.A. 2008. *Experiments in ethics*. Cambridge: Harvard University Press.

Appiah, K.A. 2010. *The honor code: How moral revolutions happen*. New York: W.W. Norton & Company.

Aquino, K., and A. Reed II. 2002. The self-importance of moral identity. *Journal of Personality and Social Psychology* 83(6): 1423–1440.

Aquino, K., A. Reed 2nd, S. Thau, and D. Freeman. 2007. A grotesque and dark beauty: How the self-importance of moral identity and mechanisms of moral disengagement influence cognitive and emotional reactions to war. *Journal of Experimental Psychology* 43: 385–392.

Aristotle. 1999. *Nicomachean ethics*. Trans. and introd. T. Irwin. Indianapolis: Hackett Publishing Co.

Arn, C. 2009. Methoden – Ethik als Instrument im Gesundheitswesen. In *Ethikwissen für Fachpersonen*, Handbuch Ethik im Gesundheitswesen, Band 2, ed. C. Arn and T. Weidmann-Hügle, 125–150. Basel: Schwabe/EMH.

Arras, J.D. 2007. The way we reason now: Reflective equilibrium in bioethics. In *The Oxford handbook of bioethics*, ed. B. Steinbeck, 51. New York: Oxford University Press.

Asch, S.E. 1952. *Social psychology*. New York: Prentice- Hall.

Asch, S.E. 1955. Opinions and social pressure. *Scientific American* 193: 31–35.

M. Christen et al. (eds.), *Empirically Informed Ethics: Morality between Facts and Norms*, Library of Ethics and Applied Philosophy 32, DOI 10.1007/978-3-319-01369-5,

Asch, A. 1988. Reproductive technologies and disability. In *Reproductive laws for the 1990s*, ed. S. Cohen and N. Taub, 69–124. New York: Humana Press.

Ashman, K.M., and P.S. Bahringer (eds.). 2001. *After the science wars*. London: Routledge.

Atran, S., R. Axelrod, and R. Davis. 2007. Sacred barriers to conflict resolution. *Science* 317: 1039–1040.

Audi, R. 2004. *The good in the right*. Princeton: Princeton University Press.

Audi, R. 2010. The place of ethical theorie in business ethics. In *The Oxford handbook of business ethics*, ed. G.G. Brenkert and T.L. Beauchamp, 46–69. Oxford: Oxford University Press.

Aureli, F., and C.M. Schaffner. 2002. Relationship assessment through emotional mediation. *Behaviour* 139: 393–420.

Ayer, A.J. 1952. *Language, truth, and logic*. New York: Dover.

Bahnemann, M., I. Dziobek, K. Prehn, I. Wolf, and H.R. Heekeren. 2010. Sociotopy in the temporoparietal cortex: Common versus distinct processes. *Social Cognitive and Affective Neuroscience* 5(1): 48–58.

Bandura, A. 1965. Influence of models' reinforcement contingencies on the acquisition of imitative responses. *Journal of Personality and Social Psychology* 1: 589–595.

Bandura, A. 1986. *Social foundations of thought and action: A social cognitive theory*. Englewood Cliffs: Prentice Hall.

Bandura, A. 1991. Social cognitive theory of moral thought and action. In *Handbook of moral behavior and development*, vol. 1, ed. K.W. Murtines and J.L. Gewirtz, 45–103. Hillsdale: Erlbaum.

Bandura, A. 2001. Social cognitive theory: An agentic perspective. *Annual Review of Psychology* 52: 1–26.

Bandura, A., G.V. Caprara, C. Barbaranelli, C. Pastorelli, and C. Regalia. 2001. Sociocognitive self regulatory mechanisms governing transgressive behavior. *Journal of Personality and Social Psychology* 80(1): 125–135.

Bargh, J.A. 1989. Conditional automaticity: Varieties of automatic influence in social perception and cognition. In *Unintended thought*, ed. J.S. Uleman and J.A. Bargh, 3–51. New York: Guilford.

Bargh, J.A. 1996. Automaticity in social psychology. In *Social psychology: Handbook of basic principles*, ed. E.T. Higgins and A.W. Krugalski, 169–183. New York: Guilford.

Bargh, J.A. 1997. The automaticity of everyday life. In *The automaticity of everyday life: Advances in social cognition*, vol. 10, ed. R.S. Wyer Jr., 1–61. Mahwah: Lawrence Erlbaum.

Barilan, Y.M., and M. Brusa. 2011. Triangular reflective equilibrium: A conscience-based method for reflective deliberation. *Bioethics* 25: 303–319.

Barker, A.T., R. Jalinous, and I.L. Freeston. 1985. Non-invasive magnetic stimulation of human motor cortex. *Lancet* 1(8437): 1106–1107.

Baron, J. 1998. *Judgment misguided: Intuition and error in public decision making*. New York: Oxford University Press.

Baron, J., and J.C. Hershey. 1988. Outcome bias in decision evaluation. *Journal of Personality and Social Psychology* 54: 569–579.

Baron, J., and M. Spranca. 1997. Protected values. *Organizational Behavior and Human Decision Processes* 70: 1–16.

Barrash, J., D. Tranel, and S.W. Anderson. 2000. Acquired personality disturbances associated with bilateral damage to the ventromedial prefrontal region. *Developmental Neuropsychology* 18(3): 355–381.

Bartal, I.B.-A., J. Decety, and P. Mason. 2011. Empathy and pro-social behavior in rats. *Science* 334: 1427–1430.

Bartsch, K., and J.C. Wright. 2005. Towards an intuitionist account of moral development. *Behavioral and Brain Sciences* 28: 546–547.

Bass, B.M., and P. Steidlmeier. 1999. Ethics, character, and authentic transformational leadership behavior. *The Leadership Quarterly* 10: 181–217.

Bateson, M., D. Nettle, and G. Roberts. 2006. Cues of being watched enhance cooperation in a real-world setting. *Biology Letters* 2: 412–414.

Batson, C.D. 1991. *The altruism question: Toward a social psychological answer*. Hillsdale: L Erlbaum.

Bauer, A.W. 2005. Wissenschaftliche Ethik als Demoskopie der Alltagsmoral? Kritische Anmerkungen zur Begründungsfrage in der Medizinischen Ethik. In *Wie viel Ethik verträgt die Medizin?* ed. M. Düwell and J.N. Neumann, 135–144. Münster: Mentis.

Baumeister, R.F. 1998. The self. In *Handbook of social psychology*, 4th ed, ed. D.T. Gilbert, S.T. Fiske, and G. Lindzey, 680–740. New York: McGraw-Hill.

Baumeister, R.F., and J.J. Exline. 1999. Virtue, personality, and social relations: Self-control as the moral muscle. *Journal of Personality* 67: 1165–1194.

Baumeister, R.F., M. Gailliot, C.N. DeWall, and M. Oaten. 2006. Self-regulation and personality: How interventions increase regulatory success, and how depletion moderates the effects of traits on behavior. *Journal of Personality* 74: 1773–1801.

Baumeister, R.F., K.D. Vohs, C.N. DeWall, and L. Zhang. 2007. How emotion shapes behavior: Feedback, anticipation, and reflection, rather than direct causation. *Personality and Social Psychology Review* 11(2): 167–203.

Bealer, G. 2000. A theory of the a priori. *Pacific Philosophical Quarterly* 81: 1–30.

Beauchamp, T.L. 2010. Relativism, multiculturalism, and universal norms: Their role in business ethics. In *The Oxford handbook of business ethics*, ed. G.G. Brenkert and T.L. Beauchamp, 235–266. Oxford: Oxford University Press.

Beauchamp, T.L., and J.F. Childress. 1979/2001/2008. *Principles of biomedical ethics*, 5th/6th ed. New York/Oxford: Oxford University Press.

Beccaria, C. 1764/1986. *On crimes and punishments*. Trans. David Young. Indianapolis: Hackett.

Bechara, A. 2005. Decision making, impulse control and loss of willpower to resist drugs: A neurocognitive perspective. *Nature Neuroscience* 8: 1458–1463.

Been, G., T.T. Ngo, S.M. Miller, and P.B. Fitzgerald. 2007. The use of tDCS and CVS as methods of non-invasive brain stimulation. *Brain Research Reviews* 56(2): 346–361.

Beilcock, S.L., and T.H. Carr. 2001. On the fragility of skilled performance: What governs choking under pressure? *Journal of Experimental Psychology. General* 130(4): 701–725.

Bekoff, M. 2000. *The smile of a Dolphin: Remarkable accounts of animal emotions*. New York: Discovery Books/Random House.

Bekoff, M. 2001. Social play behavior: Cooperation, fairness, trust, and the evolution of morality. *Journal of Consciousness Studies* 8: 81–90.

Belliveau, J.W., D.N. Kennedy, R.C. McKinstry, B.R. Buchbinder, R.M. Weisskoff, M.S. Cohen, et al. 1991. Functional mapping of the human visual cortex by magnetic resonance imaging. *Science* 254(5032): 716–719.

Bengson, J. 2010. *The intellectual given*. Dissertation, University of Texas at Austin.

Bengson, J. 2013. Experimental attacks on intuitions and answers. *Philosophy and Phenomenological Research* 86(3): 495–532.

Benhabib, S. 1992. *Situating the self. Gender, community and postmodernism in contemporary ethics*. London/New York: Routledge.

Berger, P.L., and T. Luckmann. 1969. *The social construction of reality*. New York: Doubleday.

Berker, S. 2009. The normative insignificance of neuroscience. *Philosophy and Public Affairs* 37(4): 293–329.

Berman, J.J., V.A. Murphy-Berman, and P. Singh. 1985. Cross-cultural similarities and differences in perceptions of fairness. *Journal of Cross-Cultural Psychology* 16: 55–67.

Berndt, T.J. 1979. Developmental changes in conformity to peers and parents. *Developmental Psychology* 15: 608–616.

Bernhard, H., U. Fischbacher, and E. Fehr. 2006. Parochial altruism in humans. *Nature* 442: 912–915.

Bernstein, R. 1985. *Beyond objectivism and relativism: Science, hermeneutics, and praxis*. Philadelphia: University of Pennsylvania Press.

Berthoz, S., J.L. Armony, R.J.R. Blair, and R.J. Dolan. 2002. An fMRI study of intentional and unintentional (embarrassing) violations of social norms. *Brain* 125: 1696–1708.

Berthoz, S., J. Grèzes, J.L. Armony, R.E. Passingham, and R.J. Dolan. 2006. Affective response to one's own moral violations. *NeuroImage* 31(2): 945–950.

Bethell, E., A. Whiten, G. Muhumaza, and J. Kakura. 2000. Active plant food division and sharing by wild Chimpanzees. *Primate Report* 56: 67–70.

Biller-Andorno, N. 2001. *Gerechtigkeit und Fürsorge. Zur Möglichkeit einer integrativen Medizinethik*. Frankfurt/New York: Campus.
Birnbacher, D. 1991. Sind wir für die Natur verantwortlich? In *Ökologie und Ethik*, ed. ibid., 103–139. Stuttgart: Reclam.
Birnbacher, D. 1995. *Verantwortung für zukünftige Generationen*. Stuttgart: Reclam.
Birnbacher, D. 1999. Ethics and the social science: Which kind of co-operation? *Ethical Theory and Moral Practice* 2: 319–336.
Birnbacher, D. 2003. *Analytische Einführung in die Ethik*. Berlin: De Gruyter.
Birnbacher, D. 2006. *Bioethik zwischen Natur und Interesse*. Frankfurt: Suhrkamp.
Bittles, A.H. 1990. Consanguineous marriage: Current global incidence and its relevance to demographic research. Research report no 90–186. Population Studies Center, University of Michigan, Ann Arbor.
Björnsson, G., and R. Francén Olinder. 2013. Internalists beware – We might all be amoralists! *Australasian Journal of Philosophy* 91(1): 1–14.
Blackburn, S. 1984. *Spreading the word: Groundings in the philosophy of language*. Oxford: Oxford University Press.
Blair, R.J.R. 1995. A cognitive developmental approach to morality: Investigating the psychopath. *Cognition* 57: 1–29.
Blair, R.J.R. 2008. The amygdala and ventromedial prefrontal cortex: Functional contributions and dysfunction in psychopathy. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 363(1503): 2557–2565.
Blasi, A. 1980. Bridging moral cognition and moral action: A critical review of the literature. *Psychological Bulletin* 8: 1–45.
Blasi, A. 1983. Moral cognition and moral action: A theoretical perspective. *Developmental Review* 3: 178–210.
Blasi, A. 2005. Moral character: A psychological approach. In *Character psychology and character education*, ed. D.K. Lapsley and F.C. Power, 67–100. Notre Dame: University of Notre Dame Press.
Bleisch, B., and M. Huppenbauer. 2011. *Ethische Entscheidungsfindung. Ein Handbuch für die Praxis*. Zürich: Versus Verlag.
Bleisch, B., and P. Schaber (eds.). 2007. *Weltarmut und Ethik*. Paderborn: Mentis-Verlag.
Boehm, C. 1999. *Hierarchy in the forest: The evolution of Egalitarian behavior*. Cambridge, MA: Harvard University Press.
Boesch, C. 1994. Cooperative hunting in wild Chimpanzees. *Animal Behavior* 48: 653–667.
Bogner, A. 2005. *Grenzpolitik der Experten*. Weilerwist: Velbrück Wissenschaft.
Bonabeau, E. 2002. Agent-based modeling: Methods and techniques for simulating human systems. *Proceedings of the National Academy of Sciences of the United States of America* 99(Suppl 3): 7280–7287.
Bonß, W., and H. Hartmann (eds.). 1985. *Entzauberte Wissenschaft*, Soziale Welt, Special issue 3. Göttingen: O. Schwartz.
BonJour, L. 1985. *The structure of empirical knowledge*, 93–95. Cambridge/London: Harvard University Press.
Bonnie, K., and F.B.M. de Waal. 2004. Primate social reciprocity and the origin of gratitude. In *The psychology of gratitude*, ed. R. Emmons and M. McCullough, 213–229. Oxford: Oxford University Press.
Bourdieu, P. 1977. *Outline of a theory of practice*. Cambridge: Cambridge University Press.
Bourdieu, P. 1984/1988. *Homo academicus*. Stanford: Stanford University Press.
Bowie, N. 1999. *Business ethics. A Kantian perspective*. Malden: Blackwell Publishers.
Bowlby, J. 1988. *A secure base: Parent–child attachment and healthy human development*, Tavistock professional book. London: Routledge.
Boyd, R. 1999. Homeostasis, species, and higher taxa. In *Species. New interdisciplinary essays*, ed. R.A. Wilson. Cambridge, MA: The MIT Press.
Boyd, R. 2010. Realism, natural kinds, and philosophical methods. In *The semantics and metaphysics of natural kinds*, ed. H. Beebee and N. Sabbarton-Leary. London: Routledge.

Boyd, R., and P.J. Richerson. 1985. *Culture and the evolutionary process*. Chicago: University Of Chicago Press.
Boyd, R., and J.B. Silk. 2009. *How humans evolved*, 5th ed. New York: WW Norton & Company.
Brandt, R.B. 1979. *A theory of the good and the right*. Amherst: Prometheus Books.
Brandt, R.B. 1990. The science of man and wide reflective equilibrium. *Ethics* 100: 259–278.
Bransford, J.D., A.L. Brown, and R.R. Cocking (eds.). 1999. *How people learn: Brain, mind, experience, and school*. Washington, DC: National Academy Press.
Brehm, J.W.A. 1966. *A theory of psychological reactance*. New York: Academic.
Brink, D. 1997. Moral motivation. *Ethics* 108: 4–32.
Brito, A.N. de. 2008a. The role of reasons and sentiments of Tugendhat's moral philosophy. *Crítica, Mexico-DF* 40(119): 29–43.
Brito, A.N. de. 2008b. Hume e o universalismo na Moral: por uma alternativa não kantiana. *Ethic@* 7(2): 123–136.
Bronfenbrenner, U. 1979. *The ecology of human development*. Cambridge, MA: Harvard University Press.
Brosnan, S.F. 2006. Nonhuman species' reactions to inequity and their implications for fairness. *Social Justice Research* 19: 153–185.
Brosnan, S.F. 2011. Property in nonhuman primates. *New Directions in Child and Adolescent Behavior* 132: 9–22.
Brosnan, S.F., and R. Bshary. 2010. Cooperation and deception: From evolution to mechanisms. *Philosophical Transactions of the Royal Society London B* 365(1553): 2593–2598.
Brosnan, S.F., and F.B.M. de Waal. 2002. A proximate perspective on reciprocal altruism. *Human Nature* 13(1): 129–152.
Brosnan, S.F., and F.B.M. de Waal. 2003. Monkeys reject unequal pay. *Nature* 425: 297–299.
Brosnan, S.F., and F.B.M. de Waal. 2012. Fairness in animals: Where to from here? *Social Justice Research* 25(3): 336–351. doi:10.1007/s11211-012-0165-8.
Brosnan, S.F., H.C. Schiff, and F.B.M. de Waal. 2005. Tolerance for inequity may increase with social closeness in Chimpanzees. *Proceedings of the Royal Society London Series B* 1560: 253–258.
Brosnan, S.F., C. Freeman, and F.B.M. de Waal. 2006. Partner's behavior, not reward distribution, determines success in an unequal cooperative task in capuchin monkeys. *American Journal of Primatology* 68: 713–724.
Brosnan, S.F., J. Henrich, M.C. Mareno, S. Lambeth, S. Schapiro, and J.B. Silk. 2009a. Chimpanzees (Pan troglodytes) do note develop contingent reciprocity in an experimental task. *Animal Cognition* 12(4): 587–597.
Brosnan, S.F., N.E. Newton-Fisher, and M. van Vugt. 2009b. A melding of the minds: When primatology meets social psychology. *Personality and Social Psychology Review* 13: 129–147.
Brosnan, S.F., L. Salwiczek, and R. Bshary. 2010a. The interplay of cognition and cooperation. *Philosophical Transactions of the Royal Society, Series B* 365: 2699–2710.
Brosnan, S.F., C. Talbot, M. Ahlgren, S.P. Lambeth, and S.J. Schapiro. 2010b. Mechanisms underlying the response to inequity in Chimpanzees, Pan troglodytes. *Animal Behavior* 79: 1229–1237.
Brosnan, S.F., D. Houser, K. Leimgruber, E. Xiao, T. Chen, and F.B.M. de Waal. 2010c. Competing demands of prosociality and equity in monkeys. *Evolution and Human Behavior* 31: 279–288.
Brown, M.E., and L.K. Treviño. 2006. Ethical leadership: A review and future directions. *The Leadership Quarterly* 17: 595–616.
Brown, M.E., L.K. Treviño, and D.A. Harrison. 2005. Ethical leadership. A social learning perspective for construct development and testing. *Organizational Behavior and Human Decision Processes* 97: 117–134.
Buckholtz, J.W., C.L. Asplund, P.E. Dux, D.H. Zald, J.C. Gore, O.D. Jones, et al. 2008. The neural correlates of third-party punishment. *Neuron* 60(5): 930–940.
Burkart, J.M., and C.P. van Schaik. 2010. Cognitive consequences of cooperative breeding in primates. *Animal Cognition* 13: 1–19.
Burkart, J., E. Fehr, C. Efferson, and C.P. van Schaik. 2007. Other-regarding preferences in a non-human primate: Common marmosets provision food altruistically. *Proceedings of the National Academy of Sciences* 104(50): 19762–19766.

Butler, J. 2003. *Giving an account of oneself. A critique of ethical violence* [First in German: Kritik der ethischen Gewalt]. Frankfurt: Suhrkamp.

Bystrova, K., V. Ivanova, M. Edhborg, A.S. Matthiesen, A.B. Ransjö-Arvidson, R. Mukhamedrakhimov, K. Uvnäs-Moberg, and A.M. Widström. 2009. Early contact versus separation: Effects on mother-infant interaction one year later. *Birth* 36(2): 97–109.

Cáceda, R., G.A. James, T.D. Ely, J.R. Snarey, and C.D. Kilts. 2011. Mode of effective connectivity within a putative neural network differentiates moral cognitions related to care and justice ethics. *PloS One* 6(2): e14730.

Cacioppo, J.T., R.E. Petty, J.A. Feinstein, and W. Jarvis. 1996. Dispositional differences in cognitive motivation: The life and times of individuals varying in need for cognition. *Psychological Bulletin* 119(2): 197–253.

Camerer, C.F. 2003. *Behavioral game theory: Experiments in strategic interaction*. Princeton: Princeton University Press.

Camille, N., G. Coricelli, J. Sallet, P. Pradat-Diehl, J.-R. Duhamel, and A. Sirigu. 2004. The involvement of the orbitofrontal cortex in the experience of regret. *Science* 304(5674): 1167–1170.

Campagna, A., and S. Harter. 1975. Moral judgment in sociopathic and normal children. *Journal of Personality and Social Psychology* 31: 199–205.

Caplan, A. (ed.). 1979. *The sociobiology debate*. New York: HarperCollins.

Cappelen, H. 2012. *Philosophy without intuitions*. Oxford: Oxford University Press.

Carson, T.L., and P.K. Moser. 2000. *Moral relativism: A reader*. New York: Oxford University Press.

Carver, C.S., and M.E. Scheier. 1981. *Attention and self-regulation: A control theory approach to human behavior*. New York: Springer.

Carver, C.S., and M.E. Scheier. 1990. Origins and functions of positive and negative affect: A control-process view. *Psychological Review* 97: 19–35.

Casebeer, W.D. 2003. Moral cognition and its neural constituents. *Nature Reviews Neuroscience* 4(10): 840–846.

Casti, J.L. 1995. *Complexification: Explaining a paradoxical world through the science of surprise*. New York: Harper Perennial.

Chadwick, R.F. (ed.). 1987. *Ethics, reproduction and genetic control*. New York: Croom Helm.

Chadwick, R., D. Wertz, F. Fletcher, and R. Zussmann. 1992. *Intensive care: Medical ethics and the medical profession*. Chicago: Chicago University Press.

Chaiken, S. 1980. Heuristic versus systematic information processing and the use of source versus message cues in persuasion. *Journal of Personal and Social Psychology* 39: 752–766.

Chaiken, S., and Y. Trope. 1999. *Dual-process theories in social psychology*. New York: Guilford.

Chambers, D.W. 2011. Developing a self-scoring comprehensive instrument to measure Rest's four-component model of moral behavior: The moral skills inventory. *Journal of Dental Education* 75(1): 23–35.

Chapais, B. 2008. *Primeval kinship: How pair-bonding gave birth to human society*. Cambridge, MA: Harvard University Press.

Chapman, L.J., J.P. Chapman, and M.L. Raulin. 1976. Scales for physical and social anhedonia. *Journal of Abnormal Psychology* 85: 374–382.

Charness, N. 1976. Memory for chess positions. *Journal of Experimental Psychology: Human Learning and Memory* 2: 641–653.

Chase, W., and H. Simon. 1973. Perception in chess. *Cognitive Psychology* 4: 55–81.

Chen, Q., J.B. Panksepp, and G.P. Lahvis. 2009. Empathy is moderated by genetic background in mice. *PloS One* 4(2): e4387.

Chi, M.T.H., R. Glaser, and M.J. Farr. 1988. *The nature of expertise*. Hillsdale: Erlbaum.

Chisholm, R. 1981. *The foundations of knowing*. Minneapolis: University of Minnesota Press.

Christen, M. 2009. Technisierte moral agents? Wechselwirkungen zwischen der neuroscience of ethics und dem therapeutischen Einsatz von Neurotechnologien. In *Das technisierte Gehirn*, ed. O. Müller, J. Clausen, and G. Maio, 253–272. Paderborn: Mentis.

Christen, M. 2010. Naturalisierung von Moral? Einschätzung des Beitrags der Neurowissenschaft zum Verständnis moralischer Orientierung. In *Die Strukture der moralischen Orientierung: Interdisziplinäre Perspektiven*, ed. J. Fischer and S. Gruden, 49–123. Berlin: LIT-Verlag.

Christen, M., and T. Ott. 2013. Quantified coherence of moral beliefs as a predictive factor for moral agency. In *What makes us moral? Library of ethics and applied philosophy*, ed. B. Musschenga. Berlin: Springer.

Christen, M., and M. Regard. 2012. Der "unmoralische Patient". Analyse der Nutzung hirnverletzter Menschen in der Moralforschung. *Nervenheilkunde* 31: 209–214.

Christen, M., F. Faller, U. Götz, and C. Müller. 2012. *Serious moral games. Erfassung und Vermittlung moralischer Werte durch Videospiele*. Zürich: Edition ZHdK.

Church, R.M. 1959. Emotional reactions of rats to the pain of others. *Journal of Comparative and Physiological Psychology* 52: 132–134.

Churchland, P.M. 1996. The neural representation of the social world. In *Minds and morals*, ed. L. May, M. Friedman, and A. Clark, 91–108. Cambridge, MA: MIT Press.

Churchland, P.M. 1998. Toward a cognitive neurobiology of the emotions. *Topoi* 17: 83–96.

Churchland, P.M. 2000. Rules, know-how, and the future of moral cognition. In *Moral epistemology naturalized*, ed. R. Campbell and B. Hunter. *Canadian Journal of Philosophy* (Suppl. 26): 291–306.

Churchland, P.S., and T. Sejnowski. 1992. *The computational brain*. Cambridge, MA: MIT Press.

Ciaramelli, E., M. Muccioli, E. Làdavas, and G. di Pellegrino. 2007. Selective deficit in personal moral judgment following damage to ventromedial prefrontal cortex. *Social Cognitive and Affective Neuroscience* 2(2): 84–92.

Ciulla, J.B. 1999. The importance of leadership in shaping business values. *Long Range Planning* 32: 166–172.

Ciulla, J.B. 2003. *The ethics of leadership*. Belmont: Wadsworth.

Ciulla, J.B. 2006. Ethics. The heart of leadership. In *Responsible leadership*, ed. T. Maak and N.M. Pless, 17–32. Oxon: Routledge.

Clark. A. 2000. Word and action. Reconciling rules and know-how in moral cognition. In *Moral epistemology naturalized*, ed. R. Campbell and B. Hunter B. *Canadian Journal of Philosophy* (Suppl. 26): 267–290.

Clarkeburn, H. 2002. A test for ethical sensitivity in science. *Journal of Moral Education* 31(4): 439–453.

Clutton-Brock, T.H. 2002. Breeding together: Kin selection and mutualism in cooperative vertebrates. *Science* 296: 69–72.

Cockerham, W.C. 1989. *Medical sociology*. Englewood Cliffs: Prentice Hall.

Colby, A., and W. Damon. 1992. *Some do care: Contemporary lives of moral commitment*. New York: Free Press.

Colby, A., L. Kohlberg, B. Speicher, A. Hewer, D. Candee, J.C. Gibbs, et al. 1987. *The measurement of moral judgment*. New York: Cambridge University Press.

Cole, P.M., M.K. Michel, and L.O. Teti. 1994. The development of emotion regulation and dysregulation: a clinical perspective. *Monographs of the Society for Research in Child Development* 59(2–3): 73–100.

Colman, A.D., K.E. Liebold, and J.J. Boren. 1969. A method for studying altruism in monkeys. *The Psychological Record* 19: 401–405.

Copp, D. 1995. *Morality, normativity and society*. Oxford: Oxford University Press.

Coricelli, G., H.D. Critchley, M. Joffily, J.P. O'Doherty, A. Sirigu, and R.J. Dolan. 2005. Regret and its avoidance: A neuroimaging study of choice behavior. *Nature Neuroscience* 8(9): 1255–1262.

Crane, A., and D. Matten. 2010. *Business ethics. Managing corporate citizenship and sustainability in the age of globalization*, 3rd ed. Oxford: Oxford University Press.

Cronin, K.A., K.K.E. Schroeder, E.S. Rothwell, J.B. Silk, and C. Snowdon. 2009. Cooperatively breeding cottontop tamarins (Saguinus oedipus) do not donate rewards to their long-term mates. *Journal of Comparative Psychology* 123: 231–241.

Cronin, K.A., K.K.E. Schroeder, and C. Snowdon. 2010. Prosocial behaviour emerges independent of reciprocity in cottontop tamarins. *Proceedings of the Royal Society London B* 277: 3845–3851.

Cushman, F. 2008. Crime and punishment: Distinguishing the roles of causal and intentional analysis in moral judgment. *Cognition* 108: 353–380.

Cushman, F., L. Young, and M. Hauser. 2006. The role of conscious reasoning and intuition in moral judgment: Testing three principles of harm. *Psychological Science* 17(12): 1082–1089.

Dahms, H.J. 1994. *Positivismusstreit*. Frankfurt: Suhrkamp.

Daly, M., and M. Wilson. 1999. *The truth about Cinderella*. New Haven: Yale University Press.

Damasio, A.R. 1994. *Descartes' error: Emotion, reason and the human brain*. New York: Harper Perennial.

Damasio, A.R. 1996. The somatic marker hypothesis and the possible functions of the prefrontal cortex. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 351(1346): 1413–1420.

Damasio, A.R. 1999. *The feeling of what happens*. New York: Harcourt and Brace.

Damasio, H., T. Grabowski, R. Frank, A.M. Galaburda, and A.R. Damasio. 1994. The return of Phineas Gage: Clues about the brain from the skull of a famous patient. *Science* 264(5162): 1102–1105.

Dane, E., and M.G. Pratt. 2007. Exploring intuition and its role in managerial decision making. *Academy of Management Review* 32(1): 33–54.

Daniels, N. 1979. Wide reflective equilibrium and theory acceptance in ethics. *Journal of Philosophy* 76: 256–282.

Daniels, N. 1996. *Justice and justification. Reflective equilibrium in theory and practice*. Cambridge: Cambridge University Press.

Daniels, N. 2000. Accountability for reasonableness: Establishing fair process for priority setting is easier than agreeing on principles. *BMJ* 321: 1300–1301.

Danielson, P. 1992. *Artificial morality: Virtuous robots for virtual games*. London: Routledge.

Darwin, C. 1871/1981. *The Descent of man, and selection in relation to sex*, vol. I & II. Princeton: Princeton University Press (original: London: Murray).

Darwin, C. 1872/1998. *The expression of the emotions in man and animals*, 3rd ed. London: HarperCollins.

Davidson, R.J. 2004. Well-being and affective style: Neural substrates and biobehavioural correlates. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 359(1449): 1395–1411.

Davidson, R.J., K.M. Putnam, and C.L. Larson. 2000. Dysfunction in the neural circuitry of emotion regulation – A possible prelude to violence. *Science* 289(5479): 591–594.

Davis, D.B. 1984. *Slavery and human progress*. New York: Oxford University Press.

Davis, E.P., and C.A. Sandman. 2010. The timing of prenatal exposure to maternal cortisol and psychosocial stress is associated with human infant cognitive development. *Child Development* 81(1): 131–148.

Dawkins, R. 1976. *The selfish gene*. Oxford: Oxford University Press.

De Groot, A.D. 1965. *Thought and choice in chess*. The Hague: Mouton.

De Heinzelin, J., J.D. Clark, T. White, W. Hart, P. Renne, G. WoldeGabriel, Y. Beyene, and E. Vrba. 1999. Environment and behavior of 2.5-Million-year-old Bouri hominids. *Science* 284: 625–629.

De Oliveira-Souza, R., R.D. Hare, I.E. Bramati, G.J. Garrido, F. Azevedo Ignácio, F. Tovar-Moll, et al. 2008. Psychopathy as a disorder of the moral brain: Fronto-temporo-limbic grey matter reductions demonstrated by voxel-based morphometry. *NeuroImage* 40(3): 1202–1213.

De Quervain, D.J.-F., U. Fischbacher, V. Treyler, M. Schellhammer, U. Schnyder, A. Buck, and E. Fehr. 2004. The neural basis of altruistic punishment. *Science* 305: 1254–1258.

De Waal, F.B.M. 1978. Exploitative and familiarity-dependent support strategies in a colony of semi-free living Chimpanzees. *Behaviour* 66: 268–312.

De Waal, F.B.M. 1992. Coalitions as part of reciprocal relations in the Arnhem Chimpanzee colony. In *Coalitions and alliances in humans and other animals*, ed. A.H. Harcourt and F.B.M. de Waal, 233–258. Oxford: Oxford University Press.

De Waal, F.B.E. 1996. *Good natured: The origins of right and wrong in humans and other animals*. Cambridge, MA: Harvard University Press.
De Waal, F.B.M. 1997. The Chimpanzee's service economy: Food for grooming. *Evolution and Human Behavior* 18: 375–386.
De Waal, F.B.M. 2006. *Primates and philosophers*. Princeton: Princeton University Press.
De Waal, F.B.M. 2009. *The age of empathy: Nature's Lessons for the kinder society*. New York: Harmony Books.
De Waal, F.B.M., and F. Aureli. 1996. Consolation, reconciliation, and a possible cognitive difference between macaque and Chimpanzee. In *Reaching in to thought: The minds of the great apes*, ed. A.E. Russon, K.A. Bard, and S.T. Parker, 80–110. Cambridge: Cambridge University Press.
De Waal, F.B.M., and L.M. Luttrell. 1988. Mechanisms of social reciprocity in three primate species: Symmetrical relationship characteristics or cognition? *Ethology and Sociobiology* 9: 101–118.
De Waal, F.B.M., and A. van Roosmalen. 1979. Reconciliation and consolation among Chimpanzees. *Behavioral Ecology and Sociobiology* 5: 55–66.
De Waal, F.B.M., J.A.R.A.M. van Hooff, and W.J. Netto. 1976. An ethological analysis of types of agonistic interaction in a captive group of Java-monkeys (Macaca fascicularis). *Primates* 17: 257–290.
De Waal, F.B.M., K. Leimgruber, and A. Greenberg. 2008. Giving is self-rewarding for monkeys. *Proceedings of the National Academy of Sciences* 105: 13685–13689.
DeGrazia, D. 2003. Common morality, coherence and the principles of biomedical ethics. *Kennedy Institute of Ethics Journal* 13: 219–230.
Dennett, D. 1984. *Elbow room: The varieties of Free Will Worth Wanting*. Cambridge, MA: MIT Press.
DePaul, M.R. 1993. *Balance and refinement. Beyond coherence methods of moral inquiry*. London/New York: Routledge.
DePaul, M.R. (ed.). 2001. *Resurrectiong old-fashioned foundationalism*. Oxford: Rowman & Littlefield.
Deutscher Bundestag. 2002. Schlussbericht der Enquetekommission Recht und Ethik der Modernen Medizin. Drucksache 14/9020.
Dewey, J. 1922/2000. *Human nature and conduct: An introduction to social psychology*. New York: Prometheus.
Dewey, J. 1929/1984. The quest for certainty. Reprinted in: *John Dewey, The later works*, vol. 4: 1925–1953, ed. J.A. Boydston. Carbondale and Edwardsville: Southern Illinois University Press
Dewey, J. 1930/1984. *Three independent factors in morals. The later works (1925–1953)*. Carbondale: Southern Illinois University Press.
Dewey, J. 1938/1986. *Logic: The theory of inquiry*. Carbondale: Southern University Press.
Dijksterhuis, A. 2004. Think different: The merits of unconscious thought in preference development and decision making. *Journal of Personality and Social Psychology* 87: 586–598.
Dijksterhuis, A., and L.F. Nordgren. 2006. A theory of unconscious thought. *Perspectives on Psychological Science* 1: 95–109.
Dijksterhuis, A., M.W. Bos, L.F. Nordgren, and R.B. von Baaren. 2006. On making the right choice: The deliberation-without-attention effect. *Science* 311: 1005–1007.
Domsky, D. 2004. There is no door: Finally solving the problem of moral luck. *The Journal of Philosophy* 101: 445–464.
Donders, F.C. 1969. On the speed of mental processes. *Acta Psychologica* 30: 412–431.
Doris, J.M. 2002. *Lack of character: Personality and moral behavior*. New York: Cambridge University Press.
Doris, J.M., and S.P. Stich. 2005. As a matter of fact: Empirical perspectives on ethics. In *The Oxford handbook of contemporary philosophy*, ed. F. Jackson and M. Smith. Oxford: Oxford University Press.
Draghi-Lorenz, R., V. Reddy, and A. Costall. 2001. Rethinking the development of "nonbasic" emotions: A critical review of existing theories. *Developmental Review* 21: 263–304.
Dreyfus, H.L., and S.E. Dreyfus. 1988. *Mind over machine: The power of human intuition and expertise in the era of the computer*. New York: Free Press.

Dreyfus, H.L., and S.E. Dreyfus. 1991. Towards a phenomenology of ethical expertise. *Human Studies* 14: 229–250.
Dreyfus, H.L., and S.E. Dreyfus. 2004. The ethical implications of the five-stage-skill-acquisition model. *Bulletin of Science, Technology and Society* 24: 251–264.
Duckworth, A., and M. Seligman. 2005. Self-discipline outdoes IQ in predicting academic performance of adolescents. *Psychological Science* 16(12): 939–944.
Duckworth, A., and M. Seligman. 2006. Self-discipline gives girls the edge: Gender in self-discipline, grades, and achievement test scores. *Journal of Educational Psychology* 98(1): 198–208.
Duckworth, A., C. Peterson, M. Matthews, and D. Kelly. 2010. Grit: Perseverance and passion for long-term goals. *Journal of Personality and Social Psychology* 92(6): 1087–1101.
Duffy, K.G., R.W. Wrangham, and J.B. Silk. 2007. Male Chimpanzees exchange political support for mating opportunities. *Current Biology* 17(15): R586.
Dugatkin, L.A. 1997. *Cooperation among animals: An evolutionary perspective*. New York: Oxford University Press.
Dunfee, T.W., and T. Donaldson. 2002. Untangling the corruption knot. Global bribery through the lens of integrative social contract theorie. In *The Blackwell guide to business ethics*, ed. N.E. Bowie. Malden/Oxford: Blackwell.
Düwell, M. 2008. *Bioethik. Methoden, Theorien und Bereiche*. Stuttgart/Weimar: Metzler.
Düwell, M., C. Hübenthal, and M.H. Werner. 2002. *Handbuch Ethik*. Stuttgart: Verlag J.B. Metzler.
Dworkin, R. 1978. *Taking right seriously*. Cambridge, MA: Harvard University Press.
Ebertz, R.P. 1993. Is reflective equilibrium a Coherentist model? *Canadian Journal of Philosophy* 23: 193–214.
Edel, A. 1961. *Science and the structure of ethics*. Chicago/London: University of Chicago Press.
Edmondson, R., and J. Pearce. 2007. The practice of health care: Wisdom as a model. *Medicine, Health Care, and Philosophy* 10: 233–244.
Edwards, J. 1993. Explicit connections. Ethnographic enquiry in north-west England. In *Technologies of procreation*, ed. J. Edwards, S. Franklin, E. Hirsch, F. Price, and M. Strathern, 42–66. Manchester: Manchester University Press.
Eisenberg, N., and N.D. Eggum. 2008. Empathic responding: Sympathy and personal distress. In *Cooperation: The political psychology of effective human interaction*, ed. B. Sullivan, M. Snyder, and J. Sullivan. Malden: Blackwell Publishing.
Engelhard, H.T. 1999. Moral knowledge, Moral narrative and K. Danner Clouser. The search for phronesis. In *Building bioethics*, ed. L.M. Kopelman, 51–67. Dordrecht: Kluwer Academic Press.
Epstein, S. 1991. Cognitive-experiential self-theory: An integrative theory of personality. In *The self with others: Convergences in psychoanalytical, social, and personality psychology*, ed. R. Curtis, 111–137. New York: Guilford.
Ericsson, K.A., and W. Kintsch. 1995. Long term working memory. *Psychological Review* 102: 211–245.
Ericsson, K.A., and J. Smith. 1991. *Toward a general theory of expertise*. New York: Cambridge University Press.
Eskine, K.J., N.A. Kacinik, and J.J. Prinz. 2011. A bad taste in the mouth: Gustatory disgust influences moral judgment. *Psychological Science* 22: 295–299.
Eslinger, P.J., and K.R. Biddle. 2000. Adolescent neuropsychological development after early right prefrontal cortex damage. *Developmental Neuropsychology* 18(3): 297–329.
Eslinger, P.J., and A.R. Damasio. 1985. Severe disturbance of higher cognition after bilateral frontal lobe ablation: Patient EVR. *Neurology* 35(12): 1731–1741.
Ewing, A.C. 1953. *Ethics*. London: English Universities Press.
Fazio, R.H. 1990. Multiple processes by which attitude guide behavior: The mode model as an integrative framework. *Advances in Experimental Social Psychology* 23: 75–109.
Fecteau, S., D. Knoch, F. Fregni, N. Sultani, P. Boggio, and A. Pascual-Leone. 2007a. Diminishing risk-taking behavior by modulating activity in the prefrontal cortex: A direct current stimulation study. *The Journal of Neuroscience* 27(46): 12500–12505.

Fecteau, S., A. Pascual-Leone, D.H. Zald, P. Liguori, H. Théoret, P.S. Boggio, et al. 2007b. Activation of prefrontal cortex by transcranial direct current stimulation reduces appetite for risk during ambiguous decision making. *The Journal of Neuroscience* 27(23): 6212–6218.

Fehr, E., and S. Gächter. 2002. Altruistic punishment in humans. *Nature* 415: 137–140.

Fehr, E., and U. Fischbacher. 2003. The nature of human altruism. *Nature* 423: 785–791.

Fehr, E., and B. Rockenbach. 2004. Human altruism: Economic, neural, and evolutionary perspectives. *Current Opinion in Neurobiology* 14: 784–790.

Feistner, A.T.C., and W.C. McGrew. 1989. Food-sharing in primates: A critical review. In *Perspectives in primate biology*, vol. 3, ed. P.K. Seth and S. Seth, 21–36. New Delhi: Today & Tomorrow's Printers and Publishers.

Feistner, A.T.C., and E.C. Price. 1991. Food offering in new world primates: Two species added. *Folia Primatologica* 57: 165–168.

Felitti, V.J., R.F. Anda, D. Nordenberg, D.F. Williamson, A.M. Spitz, V. Edwards, M.P. Koss, et al. 1998. The relationship of adult health status to childhood abuse and household dysfunction. *American Journal of Preventive Medicine* 14: 245–258.

Feltovich, P.J., K.M. Ford, and R.R. Hoffman. 1997. *Expertise in context*. Cambridge, MA: MIT Press.

Feltovich, P.J., M.J. Prietula, and K.A. Ericsson. 2006. Studies of expertise from psychological perspectives. In *Handbook of expertise and expert performance*, ed. K.A. Ericsson, 41–67. Cambridge: Cambridge University Press.

Ferrari, P.F., V. Gallese, G. Rizzolatti, and L. Fogassi. 2003. Mirror neurons responding to the observation of ingestive and communicative mouth actions in the ventral premotor cortex. *European Journal of Neuroscience* 17(8): 1703–1714.

Fesmire, S. 2003. *John Dewey and the moral imagination: Pragmatism in ethics*. Bloomington: Indiana University Press.

Fiala, B., A. Arico, and S. Nichols. 2011. On the psychological origins of dualism: Dual-process cognition and the explanatory gap. In *Creating consilience: Issues and case studies in the integration of the sciences and humanities*, ed. E. Slingerland and M. Collard, 88–110. Oxford: Oxford University Press.

Fine, S. 2010. Cross-cultural integrity testing as a marker of regional corruption rates. *International Journal of Selection and Assessment* 18: 251–259.

Finger, E.C., A.A. Marsh, N. Kamel, D.G.V. Mitchell, and J.R. Blair. 2006. Caught in the act: The impact of audience on the neural response to morally and socially inappropriate behavior. *NeuroImage* 33(1): 414–421.

Fins, J.J., F.G. Miller, and M.D. Bacchetta. 1997. Clinical pragmatism. A method of problem solving. *Kennedy Institute of Ethics Journal* 7: 129–142.

Fischer, J. 2009. Warum überhaupt ist Suizid ein ethisches Problem? Über Suizid und Suizidbeihilfe. *Zeitschrift für Ethik in der Medizin* 55: 243–256.

Fischer, J. 2010. Grundlagen der Moral aus ethischer Perspektive und aus der Perspektive der empirischen Moralforschung. In *Die Struktur der moralischen Orientierung. Interdisziplinäre Perspektiven*, ed. ibid., S. Gruden, 19–48. Berlin: LIT-Verlag.

Fischhoff, B., and R. Beyth. 1975. "I knew it would happen": Remembered probabilities of once-future things. *Organizational Behavior and Human Performance* 13: 1–16.

Fiske, S.T., and S.E. Taylor. 1991. *Social cognition*, 2nd ed. New York: McGraw-Hill.

Fitts, P.M., and M.I. Posner. 1967. *Human performance*. Belmont: Brookes Cole.

Fitzsimons, G.M., and J.A. Bargh. 2004. Automatic self-regulation. In *Handbook of self-regulation: Research, theory, and applications*, ed. R.F. Baumeister and K.D. Vohs, 151–170. New York: Guilford Press.

Fivush, R., J. Kuebli, and P.A. Chubb. 1992. The structure of event representations: A developmental analysis. *Child Development* 63: 188–201.

Flack, J., and F.B.M. de Waal. 2000. 'Any animal whatever': Darwinian building blocks of morality in monkeys and apes. *Journal of Consciousness Studies* 7(1–2): 1–29.

Flack, J., F.B.M. de Waal, and D.C. Krakauer. 2005. Social structure, robustness, and policing cost in a cognitively sophisticated species. *American Naturalist* 165: E126–E139.

Flack, J., M. Girvan, F.B.M. de Waal, and D.C. Krakauer. 2006. Policing stabilizes construction of social niches in primates. *Nature* 439: 426–429.
Flanagan, O. 1991. *Varieties of moral personality: Ethics and psychological realism*. Cambridge: Harvard University Press.
Fletcher, G.E. 2008. Attending to the outcome of others: Disadvantageous inequity aversion in male capuchin monkeys (Cebus apella). *American Journal of Primatology* 70: 901–905.
Fletcher, J.C., E.M. Spencer, and P.A. Lombardo (eds.). 2005. *Fletcher's introduction to clinical ethics*, 3rd ed. Hagerstown: University Publishing Group.
Flyvbjerg, B. 2001. *Making social science matter. Why social inquiry fails and how it can succeed again*. Cambridge/New York: Cambridge University Press.
Foot, P. 1972. Morality as a system of hypothetical imperatives. *The Philosophical Review* 81(3): 305–316.
Forgas, J.P. 1995. Mood and judgment: The affect infusion model (AIM). *Psychological Bulletin* 117(1): 39–66.
Foster, E.K. 2004. Research on gossip: Taxonomy, methods, and future directions. *Review of General Psychology* 8: 78–99.
Foucault, M. 1971/2004. *The order of things: An archaeology of the human sciences*. New York: Vintage Books.
Foucault, M. 1973. Archäologie des Wissens. Frankfurt a.M.: Suhrkamp.
Foucault, M. 2007. *The politics of truth*, 2nd ed. Los Angeles: Semiotext(e).
Fox, R.C. 2008. The bioethics that I would like to see. *Clinical Ethics* 3: 25–26.
Fox, E., S. Myers, and R.A. Pearlman. 2007. Ethics consultation in United States hospitals. A national survey. *The American Journal of Bioethics* 7: 13–25.
Fraedrich, J., O.C. Ferrell, and L. Ferrell. 2011. *Ethical decision making for business*, 8th ed. Mason/Andover: South-Western/Cengage Learning.
Frank, R.H. 1988. *Passion within reason: The strategic role of the emotions*. New York: Norton.
Franklin, S. 1997. *Embodied progress. A cultural account of assisted conception*. London: Routledge.
Fraser, D., D.M. Weary, E.A. Pajor, and B.N. Milligan. 1997. A scientific conception of animals that reflects ethical concerns. *Animal Welfare* 6: 187–205.
Freemann, R.E., J.S. Harrison, A.C. Wicks, B.L. Parmar, and S. de Colle. 2010. *Stakeholder theory. The state of the art*. Cambridge: Cambridge University Press.
French, P. 1998. *Individual and collective responsibility*. Rochester: Schenkman.
Freud, S. 1927/1961. *The future of an illusion*. Trans. J. Strachey. New York: Norton & Co.
Fricker, M. 1998. Rational authority and social power: Towards a truly social epistemology. *Proceedings of the Aristotelian Society* 98(2): 159–177.
Frijda, N.H. 1986. *The emotions*. Cambridge: Cambridge University Press.
Frimer, J.A., and L.J. Walker. 2009. Reconciling the self and morality: An empirical model of moral centrality development. *Developmental Psychology* 45: 1669–1681.
Friston, K.J., C.J. Price, P. Fletcher, C. Moore, R.S. Frackowiak, and R.J. Dolan. 1996. The trouble with cognitive subtraction. *NeuroImage* 4(2): 97–104.
Gächter, S., and B. Herrmann. 2006. Human cooperation from an economic perspective. In *Cooperation in primates and humans. Mechanisms and evolution*, ed. P.M. Kappeler and C.P. van Schaik, 275–301. Berlin: Springer.
Gardner, W.L., B.J. Avolio, F. Luthans, D.R. May, and F.O. Walumbwa. 2005. "Can you see the real me?" A self-based model of authentic leader and follower development. *The Leadership Quarterly* 16: 343–372.
Gauthier, D. 1986. *Morals by agreement*. Oxford: Clarendon.
Gibbard, A. 1992. *Wise choices, Apt feelings. A theory of normative judgments*. Cambridge: Harvard University Press.
Gibbs, J.C., K.S. Basinger, and R. Fuller. 1992. *Moral maturity: Measuring the development of sociomoral reflection*. Hillsdale: Erlbaum.
Gibson, R., C. Tanner, and A. Wagner. 2013. Preferences for truthfulness: Heterogenity among and within individuals. *American Economic Review* 103: 532–548.

Gieryn, T.F. 1983. Boundary work and the demarcation of science from non-science. *American Sociological Review* 38: 781–795.
Gigerenzer, G. 2010. Moral satisficing: Rethinking moral behavior as bounded rationality. *Topics in Cognitive Science* 2: 528–554.
Gigerenzer, G., P.M. Todd, and The ABC Research Group. 1999. *Simple heuristics that make us smart*. New York: Oxford University Press.
Gilligan, C. 1977. In a different voice – Womens conceptions of self and of morality. *Harvard Educational Review* 47: 481–517.
Gilligan, C. 1982. *In a different voice*. Cambridge, MA: Harvard University Press.
Gilligan, C., and J. Attanucci. 1988. Two moral orientations – Gender differences and similarities. *Merrill-Palmer Quarterly of Behavior and Development* 34: 223–237.
Gino, F., M.E. Schweitzer, N.L. Mead, and D. Ariely. 2011. Unable to resist temptation: How self-control depletion promotes unethical behavior. *Organizational Behavior and Human Decision Processes* 115(2): 191–203.
Gintis, H., S. Bowles, R.T. Boyd, and E. Fehr. 2005. *Moral sentiments and material interests: The foundations of cooperation in economic life*. Cambridge, MA: MIT-Press.
Glenn, A.L., and A. Raine. 2008. The neurobiology of psychopathy. *The Psychiatric Clinics of North America* 31(3): 463–475.
Glenn, A.L., R. Iyer, J. Graham, S. Koleva, and J. Haidt. 2009. Are all types of morality compromised in psychopathy? *Journal of Personality Disorders* 23(4): 384–398.
Goffman, E. 1959. *The presentation of self in everyday life*. New York: Anchor books.
Goldenberg, M.J. 2005. Evidence based ethics? On evidence-based practice and the "empirical turn" from normative bioethics. *BMC Medical Ethics* 6: E1–E9.
Goldman, A.I. 1979. What is justified belief? In *Justification and knowledge: New studies in epistemology*, ed. G. Pappas, 1–23. Dordrecht: Reidel.
Goleman, D. 1995. *Emotional intelligence*. New York: Bantam.
Gollan, T., and E.H. Witte. 2008. "It was right to do it, because…" Understanding justifications of actions as prescriptive attributions. *Social Psychology* 39(3): 189–196.
Gollan, T., A.C. Moser, M.L. Mendes-Teixeira, and V. Brandt. 2011. "It was the right decision because…" – Cultural and personal determinants of ethical justification. Paper presented at the 14th International Conference on Social Dilemmas, Amsterdam, July 6–9, 2011.
Gollenia, M.C. 1999. *Ethische Entscheidungen und Rechtfertigungen unter der besonderen Bedingung der sozialen Identität* [Ethical decisions and justification and their dependence on social identity]. Frankfurt am Main: Peter Lang.
Gomes, C.M., and C. Boesch. 2009. Wild Chimpanzees exchange meat for sex on a long-term basis. *PloS One* 4(4): e5116.
Gomes, C.M., R. Mundry, and C. Boesch. 2008. Long-term reciprocation of grooming in wild West African Chimpanzees. *Proceedings of the Royal Society B* 276: 699–706.
Goodall, J. 1986. *The Chimpanzees of Gombe: Patterns of behavior*. Cambridge, MA: Harvard University Press.
Goodpaster, K.E. 2002. Teaching and learning ethics by the case method. In *The Blackwell guide to business ethics*, ed. N. Bowie, 117–141. Malden: Blackwell.
Goody, J. 1980. Slavery in time and space. In *Asian and African systems of slavery*, ed. J.C. Watson. Berkeley: University of California Press.
Goody, J. 1983. *The development of the family and marriage in Europe*. Cambridge: Cambridge University Press.
Graham, J., J. Haidt, and B.A. Nosek. 2009. Liberals and conservatives rely on different sets of moral foundations. *Journal of Personality and Social Psychology* 96(5): 1029–1046.
Graham, J., B.A. Nosek, J. Haidt, R. Iyer, S. Koleva, and P.H. Ditto. 2011. Mapping the moral domain. *Journal of Personality and Social Psychology* 101(2): 366–385.
Greene, J.D. 2007. Why are VMPFC patients more utilitarian? A dual-process theory of moral judgment explains. *Trends in Cognitive Sciences* 11(8): 322–323.
Greene, J. 2008. The secret joke of Kant's soul. In *Moral psychology*, vol. 3, ed. W. Sinnott-Armstrong. Cambridge, MA: MIT Press.

Greene, J.D. manuscript. Notes on 'The normative significance of neuroscience' by Selim Berker. Online at http://www.wjh.harvard.edu/~jgreene/GreeneWJH/Greene-Notes-on-Berker-Nov10.pdf. Accessed 14 Nov 2012.
Greene, J., and J. Haidt. 2002. How (and where) does moral judgment work? *Trends in Cognitive Sciences* 6: 517–523.
Greene, J.D., and J.M. Paxton. 2009. Patterns of neural activity associated with honest and dishonest moral decisions. *Proceedings of the National Academy of Sciences of the United States of America* 106(30): 12506–12511.
Greene, J.D., R.B. Sommerville, L.E. Nystrom, J.M. Darley, and J.D. Cohen. 2001. An fMRI investigation of emotional engagement in moral judgment. *Science* 293: 2105–2108.
Greene, J.D., L.E. Nystrom, A.D. Engell, J.M. Darley, and J.D. Cohen. 2004. The neural bases of cognitive conflict and control in moral judgment. *Neuron* 44(2): 389–400.
Greene, J.D., S.A. Morelli, K. Lowenberg, L.E. Nystrom, and J.D. Cohen. 2008. Cognitive load selectively interferes with utilitarian moral judgment. *Cognition* 107(3): 1144–1154.
Greene, J.D., F.A. Cushman, L.E. Stewart, K. Lowenberg, L.E. Nystrom, and J.D. Cohen. 2009. Pushing moral buttons: The interaction between personal force and intention in moral judgment. *Cognition* 111(3): 364–371.
Greenspan, S.I., and S.I. Shanker. 2004. *The first idea*. Cambridge, MA: Da Capo Press.
Greenwald, A., D. McGhee, and J. Schwartz. 1998. Measuring individual differences in implicit cognition: The Implicit Association Test. *Journal of Personality and Social Psychology* 74: 1464–1480.
Grosjean, B., and G.E. Tsai. 2007. NMDA neurotransmission as a critical mediator of borderline personality disorder. *Journal of Psychiatry and Neuroscience* 32(2): 103–115.
Gross, P.R., and N. Levitt. 1994. *Higher superstition. The academic left and its quarrels with science*. Baltimore: John Hopkins University Press.
Guarini, M. 2007. Computation, coherence, and ethical reasoning. *Minds and Machines* 17: 27–46.
Gurven, M. 2004. To give and to give not: The behavioral ecology of human food transfers. *Behavioral and Brain Sciences* 27: 543–583.
Gurven, M., and K. Hill. 2009. Why do men hunt? A reevaluation of "Man the Hunter" and the sexual division of labor. *Current Anthropology* 50: 51–74.
Guse, B., P. Falkai, and T. Wobrock. 2010. Cognitive effects of high-frequency repetitive transcranial magnetic stimulation: A systematic review. *Journal of Neural Transmission* 117(1): 105–122.
Guth, W., R. Schmittberger, and B. Schwartze. 1982. An experimental analysis of ultimatum bargaining. *Journal of Economic Behavior and Organization* 3: 367–388.
Haack, S. 1993. *Evidence and inquiry. Towards reconstruction in epistemology*. Oxford/Cambridge: Blackwell Publishers.
Habermas, J. 1981. *Theorie kommunikativen Handelns*, Vol 1. Handlungsrationalität und gesellschaftliche Rationalisierung, Vol. 2. Zur Kritik der funktionalistischen Vernunft. Frankfurt: Suhrkamp.
Habermas, J. 1987. *The philosophical discourse of modernity*. Cambridge, MA: MIT Press.
Habermas, J. 2001. *Die Zukunft der menschlichen Natur. Auf dem Weg zu einer liberalen Eugenik?* Frankfurt: Suhrkamp.
Hackel, S. 1995. *Zur beruflichen Sozialisation und Identität ost- und westdeutscher Arbeitnehmer* [On the professional socialization of employees in East and West Germany]. Dresden: University of Dresden.
Haidt, J. 2001. The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review* 108: 814–834.
Haidt, J. 2003. The moral emotions. In *The handbook of affective sciences*, ed. R.J. Davidson, K.R. Scherer, and H.H. Goldsmith, 852–870. Oxford: Oxford University Press.
Haidt, J. 2007. The new synthesis in moral psychology. *Science* 316: 998–1002.
Haidt, J. 2008. Morality. *Perspectives on Psychological Science* 3: 65–72.
Haidt, J., and J. Graham. 2007. When morality opposes justice: Conservatives have moral intuitions that liberals may not recognize. *Social Justice Research* 20: 98–116.

Haidt, J., and C. Joseph. 2007. The moral mind: How five sets of innate intuitions guide the development of many culture-specific virtues, and perhaps even modules. In *The innate mind*, vol. 3, ed. P. Carruthers, S. Laurence, and S. Stich, 367–391. New York: Oxford University Press.

Haidt, J., and S. Kesebir. 2010. Morality. In *Handbook of social psychology*, vol. 2, ed. S.T. Fiske, D.T. Gilbert, and G. Lindzey, 797–832. Hoboken: Wiley.

Haidt, J., S.H. Koller, and M.G. Dias. 1993. Affect, culture, and morality, or is it wrong to eat your dog? *Journal of Personality and Social Psychology* 65(4): 613–628.

Haimes, E., and R. Williams. 1998. Social constructionism and the new technologies of reproduction. In *The politics of constructionism*, ed. I. Velody and R. Williams, 132–146. London: Sage.

Haley, K.J., and D.M.T. Fessler. 2005. Nobody's watching? Subtle cues affect generosity in an anonymous economic game. *Evolution and Human Behavior* 26: 245–256.

Halverscheid, S., and E.H. Witte. 2007. Inhaltsanalytische Modelle zur Identifikation und Analyse von ethischen Rechtfertigungen politischer Gewalt [Identifying and analyzing ethical justification of political violence with content analytic models]. *Sicherheit und Frieden* 25(2): 85–91.

Halverscheid, S., and E.H. Witte. 2008. Justification of war and terrorism. A comparative case study analyzing ethical positions based on prescriptive attribution theory. *Social Psychology* 39(1): 26–36.

Ham, J., K. van den Bos, and E.A. van Doorn. 2009. Lady justice thinks unconsciously: Unconscious thought can lead to more accurate justice judgments. *Social Cognition* 27: 509–521.

Hamlin, J.K., K. Wynn, and P. Bloom. 2007. Social evaluation by preverbal infants. *Nature* 450: 557–560.

Hammond, K. 1996. *Human judgment and social policy*. New York: Oxford University Press.

Hanselmann, M., and C. Tanner. 2008. Taboos and conflicts in decision making: Sacred values, decision difficulty, and emotions. *Judgment and Decision Making* 3: 51–63.

Haraway, D. 1989. *Primate visions. Gender, race and nature in the world of modern science*. New York/London: Routledge.

Harbaugh, W.Z., U. Mayr, and D.R. Burghart. 2007. Neural responses to taxation and voluntary giving reveal motives for charitable donations. *Science* 326: 1622–1625.

Hardin, G. 1968. The tragedy of the commons. *Science* 162: 1243–1248.

Hardy, S.A., and G. Carlo. 2005. Identity as a source of moral motivation. *Human Development* 48: 232–256.

Hare, B., A.P. Melis, V. Woods, S. Hastings, and R. Wrangham. 2007. Tolerance allows Bonobos to outperform Chimpanzees on a cooperative task. *Current Biology* 17: 619–623.

Harenski, C.L., and S. Hamann. 2006. Neural correlates of regulating negative emotions related to moral violations. *NeuroImage* 30(1): 313–324.

Harenski, C.L., and K.A. Kiehl. 2010. Reactive aggression in psychopathy and the role of frustration: Susceptibility, experience, and control. *British Journal of Psychology* 101(3): 401–406.

Harenski, C.L., S.H. Kim, and S. Hamann. 2009. Neuroticism and psychopathy predict brain activation during moral and nonmoral emotion regulation. *Cognitive, Affective & Behavioral Neuroscience* 9(1): 1–15.

Harenski, C.L., O. Antonenko, M.S. Shane, and K.A. Kiehl. 2010a. A functional imaging investigation of moral deliberation and moral intuition. *NeuroImage* 49(3): 2707–2716.

Harenski, C.L., K.A. Harenski, M.S. Shane, and K.A. Kiehl. 2010b. Aberrant neural processing of moral violations in criminal psychopaths. *Journal of Abnormal Psychology* 119(4): 863–874.

Harlow, J.M. 1848. Passage of an iron rod through the head. *The Boston Medical and Surgical Journal* 39: 389–393.

Harman, G. 1986. *Change in view*. Cambridge, MA: MIT Press.

Harman, G. 1999. Moral philosophy meets social psychology: Virtue ethics and the fundamental attribution error. *Proceedings of the Aristotelian Society* 99: 315–331.

Harris, M. 1977/1991. *Cannibals and kings. Origins of cultures*. New York: Vintages.

Hart, B.L., and L.A. Hart. 1992. Reciprocal allogrooming in Impala, Aepyceros melampus. *Animal Behavior* 44: 1073–1083.

Hauser, M.D. 2006a. *Moral minds: How nature designed our sense of right and wrong*. New York: Ecco Press.

Hauser, M.D. 2006b. The liver and the moral organ. *Social Cognitive and Affective Neuroscience* 1(3): 214–220.

Hauser, M.D., M.K. Chen, F. Chen, and E. Chuang. 2003. Give unto others: Genetically unrelated cotton-top tamarin monkeys preferentially give food to those who altruistically give food back. *Proceedings of the Royal Society of London, Series B* 270: 2363–2370.

Hawkes, K., J.F. O'Connell, N.G. Blurton Jones, H. Alvarez, and E.L. Charnov. 1998. Grandmothering, menopause, and the evolution of human life histories. *Proceedings of the National Academy of Sciences of the United States of America* 95: 1336–1339.

Hawkins, S., and R. Hastie. 1990. Hindsight: Biased judgements of past events after the outcomes are known. *Psychological Bulletin* 107: 311–327.

Heekeren, H.R., I. Wartenburger, H. Schmidt, H.-P. Schwintowski, and A. Villringer. 2003. An fMRI study of simple ethical decision-making. *Neuroreport* 14(9): 1215–1219.

Heekeren, H.R., I. Wartenburger, H. Schmidt, K. Prehn, H.-P. Schwintowski, and A. Villringer. 2005. Influence of bodily harm on neural correlates of semantic and moral decision-making. *NeuroImage* 24(3): 887–897.

Heider, F. 1958. *The psychology of interpersonal relations*. New York: Wiley.

Heinrich, B. 1999. *Mind of the Raven*. New York: Harper Collins.

Heitkamp, I. 2007. *Die Entwicklung einer Moderationsmethode für Ethikkommissionen* [The development of a moderation technique for ethics committees]. Retrieved January 3, 2008, from http://www.sub.uni-hamburg.de/opus/volltexte/2007/3313/

Helbing, D., A. Szolnoki, M. Perc, and G. Szabó. 2010. Evolutionary establishment of moral and double moral standards through spatial interactions. *PLOS Computational Biology* 6(4): e1000758.

Held, V. 1970. Can a random collection of individuals be responsible? *Journal of Philosophy* 67: 471–481.

Henrich, J., R. Boyd, S. Bowles, C. Camerer, E. Fehr, H. Gintis, R. McElreath, M. Alvard, A. Barr, J. Ensminger, N. Smith Henrich, K. Hill, F. Gil-White, M. Gurven, F.W. Marlowe, J.Q. Patton, and D. Tracer. 2005. "Economic man" in cross-cultural perspective: Behavioral experiments in 15 small-scale societies. *Behavioral and Brain Sciences* 28: 795–855.

Henrich, J., J. Ensminger, R. McElreath, A. Barr, C. Barrett, A. Bolyanatz, J.C. Cardenas, M. Gurven, E. Gwako, N. Henrich, C. Lesorogol, F.W. Marlowe, D. Tracer, and J. Ziker. 2010. Markets, religion, community size, and the evolution of fairness and punishment. *Science* 327: 1480–1484.

Henry, J.P., and S. Wang. 1998. Effects of early stress on adult affiliative behavior. *Psychoneuroendocrinology* 23(8): 863–875.

Henson, R. 2006. Forward inference using functional neuroimaging: Dissociations versus associations. *Trends in Cognitive Sciences* 10(2): 64–69.

Herdt, J.A. 2008. *Putting on virtue: The legacy of the splendid vices*. Chicago: University of Chicago Press.

Herrera, C. 2008. Is it time for bioethics to go empirical? *Bioethics* 22: 137–146.

Hewstone, M., W. Stroebe, and K. Jonas. 2007. *Introduction to social psychology. A European perspective*, 4th ed. Oxford: Blackwell Publishing.

Higgins, E.T. 1996. Knowledge activation: Accessibility, applicability, and salience. In *Social psychology: Handbook of basic principles*, ed. E.T. Higgins and A.W. Kruglanski, 133–168. New York: Guilford Press.

Hill, K. 2002. Altruistic cooperation during foraging by the Ache, and the evolved human predisposition to cooperate. *Human Nature* 13(1): 105–128.

Hill, K.R. 2009. Animal "Culture". In *The question of animal culture*, ed. K.N. Laland and B.G. Galef, 269–287. Cambridge, MA: Harvard University Press.

Hill, K.R., and A.M. Hurtado. 2009. Cooperative breeding in South American hunter-gatherers. *Proceedings of the Royal Society B* 276: 3863–3870.

Hill, K.R., R.S. Walker, M. Bozicevic, J. Eder, T. Headland, B. Hewlett, A.M. Hurtado, F. Marlowe, P. Wiessner, and B. Wood. 2011. Co-residence patterns in hunter-gatherer societies show unique human social structure. *Science* 331: 1286–1289.

Hinde, R.A. 1970. *Animal behaviour: A synthesis of ethology and comparative psychology*, 2nd ed. Tokyo: McGraw-Hill Kokagusha, Ltd.
Hockings, K.J., T. Humle, J.R. Anderson, D. Biro, C. Sousa, G. Ohashi, and T. Matsukawa. 2007. Chimpanzees share forbidden fruit. *PloS One* 2(9): e886.
Höffe, O. 1984. Sittlichkeit als Rationalität des Handelns? In *Rationalität. Philosophische Beiträge*, ed. H. Schnädelbach, 141–174. Frankfurt: Suhrkamp.
Hoffman, R.R. 1987. The problem of extracting the knowledge of experts from the perspective of experimental psychology. *The AI Applications* 1(2): 35–48.
Hoffman, M. 2000. *Empathy and moral development: Implications for caring and justice*. Cambridge: Cambridge University Press.
Hoffmaster, B. 1992. Can ethnography save the life of ethics? *Social Science & Medicine* 35: 1421–1431.
Hoffmaster, B. 1994. The forms and limits of medical ethics. *Social Science & Medicine* 35: 1155–1164.
Hogarth, R.M. 2001. *Educating intuition*. Chicago: University of Chicago Press.
Hogarth, R.M. 2002. *Deciding analytically or trusting your intuition? The advantages and disadvantages of analytic and intuitive thought*. Retrieved on February 15, 2010, from http://www.econ.upf.edu/docs/papers/downloads/654.pdf. A shorter version is published in Betsch, T., and S. Haberstroh (eds.). 2005. *The routines of decision making*, 67–82. Mahwah: Erlbaum.
Holm, S. 2008. Background paper on Article 14 of the Universal Declaration on Bioethics and Human Rights from a philosophical perspective. Report of the meeting of the working group of IBC on social responsibility and health (Annex V), 13, UNESCO, Paris.
Horner, V., J.D. Carter, M. Suchak, and F.B.M. de Waal. 2011. Spontaneous prosocial choice by Chimpanzees. *Proceedings of the National Academy of Sciences of the United States of America* 108(33): 13847–13851. doi:10.1073/pnas.1111088108.
Horton, K. 2004. Aid and bias. *Inquiry* 47: 545–561.
Hrdy, S.B. 1977. *The Langurs of Abu. Female and male strategies of reproduction*. Cambridge, MA/London: Harvard University Press.
Hrdy, S.B. 2009. *Mothers & others: The evolutionary origins of mutual understanding*. Cambridge: Harvard University Press.
Huemer, M. 2001. *Skepticism and the veil of perception*. Lanham: Rowman & Littlefield.
Huff, C., and W. Frey. 2005. Moral pedagogy and practical ethics. *Science and Engineering Ethics* 11: 389–408.
Hume, D. 1739–1740/1978/2003. *A treatise of human nature*. Introd. and ed. L.A. Selby-Bigge and P.H. Nidditch. Oxford/Mineola: Oxford University Press/Dover Publications (2003 edition).
Hume, D. 1751. *An enquiry concerning the principles of morals*. Publ. for A Millar Online see David Hume, D Banach, St Anselm College. http://www.anselm.edu/homepage/dbanach/Hume-Enquiry%20Concerning%20Morals.htm#sec1. [German version: Hepfer, K. (ed.). 2002. *Eine Untersuchung der Grundlagen der Moral*. Göttingen: Vandenhoek & Ruprecht].
Huxley, T.H. 1894. *Evolution and ethics*. London: MacMillan and Co.
Ives, J. 2008. Encounters with experience: Empirical bioethics and the future. *Health Care Analysis* 16: 1–6.
Ives, J., and H. Draper. 2009. Appropriate methodologies for empirical bioethics: It's all relative. *Bioethics* 23: 249–258.
Jacobson, D. 2005. Seeing by feeling: Virtues, skills and moral perception. *Ethical Theory and Moral Practice* 8(2): 387–409.
Jaeggi, A.V., J.M. Burkart, and C.P. van Schaik. 2010a. On the psychology of cooperation in humans and other primates: Combining the natural history and experimental evidence of prosociality. *Philosophical Transactions of the Royal Society of London* 365: 2723–2735.
Jaeggi, A.V., J.M.G. Stevenson, and C.P. Van Schaik. 2010b. Tolerant food sharing and reciprocity precluded by despotism among Bonobos but not Chimpanzees. *American Journal of Physical Anthropology* 143: 41–51.

James Jr., S.H. 2000. Reinforcing ethical decision making through organizational structure. *Journal of Business Ethics* 28: 43–58.

Jamieson, D. 1991. Method and moral theory. In *A companion to ethics*, ed. P. Singer, 476–487. Oxford: Blackwell.

Janis, I.L., and L. Mann. 1977. *Decision making: A psychological analysis of conflict, choice, and commitment*. New York: Free Press.

Jasanoff, S. 1990. *The fifth branch. Science advisers as policy makers*. Cambridge: Harvard University Press.

Jensen, K., B. Hare, J. Call, and M. Tomasello. 2006. What's in it for me? Self-regard precludes altruism and spite in Chimpanzees. *Proceedings of the Royal Society B: Biological Sciences* 273: 1013–1021.

Johnson, A.W., and T. Earle. 2000. *The evolution of human societies: From foraging group to Agrarian state*, 2nd ed. Stanford: Stanford University Press.

Jones, T.M. 1991. Ethical decision making by individuals in organizations: An issue-contingent model. *Academy of Management Review* 16: 366–395.

Jordan, J. 2009. A social cognition framework for examining moral awareness in managers and academics. *Journal of Business Ethics* 84: 237–258.

Joyce, R. 2006. *The evolution of morality*. Cambridge, MA: MIT Press.

Kahane, G., and N. Shackel. 2008. Do abnormal responses show utilitarian bias? *Nature* 452: E5–E6.

Kahane, G., K. Wiech, N. Shackel, M. Farias, J. Savulescu, and I. Tracey. 2012. The neural basis of intuitive and counterintuitive moral judgment. *Social Cognitive and Affective Neuroscience* 7(4): 393–402.

Kahneman, D. 2003. A perspective on judgment and choice: Mapping bounded rationality. *American Psychologist* 58: 697–720.

Kahneman, D. 2011. *Thinking fast and slow*. New York: MacMillan.

Kahneman, D., and G. Klein. 2009. Conditions for intuitive expertise: A failure to disagree. *American Psychologist* 64(6): 515–526.

Kant, I. 1781/1974. *Kritik der reinen Vernunft*. Frankfurt: Suhrkamp.

Kant, I. 1785/1983/2002. *Grundlegung zur Metaphysik der Sitten*. Darmstadt: Wissenschaftliche Buchgesellschaft. [English edition 2002. *Groundwork for the metaphysics of morals*. Binghamton: Vail-Ballou Press].

Kant, I. 1788/1983. *Kritik der praktischen Vernunft*. Darmstadt: Wissenschaftliche Buchgesellschaft.

Kant, I. 1798/1983. *Anthropologie in pragmatischer Hinsicht*. Stuttgart: Reclam.

Kanungo, R.N. 2001. Ethical values of transactional and transformational leaders. *Canadian Journal of Administrative Sciences* 18: 257–265.

Kaplan, H., P.L. Hooper, and M. Gurven. 2009. The evolutionary and ecological roots of human social organization. *Philosophical Transactions of the Royal Society B: Biological Sciences* 364: 3289–3299.

Kappel, K. 2006. The meta-justification of reflective equilibrium. *Ethical Theory and Moral Practice* 9: 131–147.

Karim, A.A., M. Schneider, M. Lotze, R. Veit, P. Sauseng, C. Braun, et al. 2010. The truth about lying: Inhibition of the anterior prefrontal cortex improves deceptive behavior. *Cerebral Cortex* 20(1): 205–213.

Kauppinen, A. 2007. The rise and fall of experimental philosophy. *Philosophical Explorations* 10(2): 95–118.

Kauppinen, A. Forthcoming. Intuition and belief in moral motivation. In *Motivational Internalism*, ed. G. Björnsson. Oxford: Oxford University Press.

Kawabata, H., and S. Zeki. 2004. Neural correlates of beauty. *Journal of Neurophysiology* 91: 1699–1705.

Keeley, L.H. 1988. Hunter-gatherer economic complexity and "population pressure": A cross-cultural analysis. *Journal of Anthropological Archaeology* 7: 373–411.

Keeney, R.L. 1992. *Value-focused thinking: A path to creative decision making*. Cambridge, MA: Harvard University Press.

Kelly, D., S. Stich, D. Fessler, K. Haley, and S. Eng. 2007. Harm, affect and the moral/conventional distinction. *Mind and Language* 22: 117–131.

Keltner, D., E.J. Horberg, and C. Oveis. 2006. Emotions as moral intuitions. In *Affect in social thinking and behavior*, ed. J.P. Forgas, 161–175. New York: Psychology Press.

Kennett, J., and C. Fine. 2009. Will the real moral judgment please stand up? The implications of social intuitionist models of cognition for meta-ethics and moral psychology. *Ethical Theory and Moral Practice* 12: 77–96.

Kiehl, K.A. 2006. A cognitive neuroscience perspective on psychopathy: Evidence for paralimbic system dysfunction. *Psychiatry Research* 142(2–3): 107–128.

Kihlstrom, J.F., V.A. Shames, and J. Dorfman. 1996. Intimations of memory and thought. In *Implicit memory and metacognition*, ed. L. Reder, 1–23. Mahwah: Erlbaum.

Kitcher, P. 1998. Psychological altruism, evolutionary origins, and moral rules. *Philosophical Studies* 89: 283–316.

Kitcher, P. 2011. *The ethical project*. Cambridge, MA: Harvard University Press.

Klaus, M.H., and J.H. Kennell. 1976/1983. *Maternal-infant bonding: The impact of early separation or loss on family development*. St. Louis: C V Mosby.

Klein, G.A. 2008. Naturalistic decision making. *Human Factors* 50(3): 456–460.

Klein, C. 2011. The dual track theory of moral decision-making: A critique of the neuroimaging evidence. *Neuroethics* 4: 143–162.

Klein, G.A., S. Wolf, L. Militello, and C. Zsambok. 1995. Characteristics of skilled option generation in chess. *Organizational Behavior and Human Decision Processes* 62(1): 63–69.

Kliemann, D., L. Young, J. Scholz, and R. Saxe. 2008. The influence of prior record on moral judgment. *Neuropsychologia* 46(12): 2949–2957.

Knobe, J. 2007. Experimental philosophy and philosophical significance. *Philosophical Explorations* 10(2): 119–121.

Knobe, J., and S. Nichols (eds.). 2008. *Experimental philosophy*. Oxford: Oxford University Press.

Knoch, D., and E. Fehr. 2007. Resisting the power of temptations: The right prefrontal cortex and self-control. *Annals of the New York Academy of Sciences* 1104: 123–134.

Knoch, D., A. Pascual-Leone, K. Meyer, V. Treyer, and E. Fehr. 2006. Diminishing reciprocal fairness by disrupting the right prefrontal cortex. *Science* 314(5800): 829–832.

Knoch, D., M.A. Nitsche, U. Fischbacher, C. Eisenegger, A. Pascual-Leone, and E. Fehr. 2008. Studying the neurobiology of social interaction with transcranial direct current stimulation – The example of punishing unfairness. *Cerebral Cortex* 18(9): 1987–1990.

Kochanska, G. 2002. Mutually responsive orientation between mothers and their young children: A context for the early development of conscience. *Current Directions in Psychological Science* 11: 191–195.

Koenigs, M., and D. Tranel. 2007. Irrational economic decision-making after ventromedial prefrontal damage: Evidence from the Ultimatum Game. *Journal of Neuroscience* 27(4): 951–956.

Koenigs, M., L. Young, R. Adolphs, D. Tranel, F. Cushman, M. Hauser, et al. 2007. Damage to the prefrontal cortex increases utilitarian moral judgements. *Nature* 446(7138): 908–911.

Kohlberg, L. 1964. Development of moral character and moral ideology. In *Review of child development research*, ed. M.L. Hoffman and L.W. Hoffman, 381–431. New York: Russell Sage Foundation.

Kohlberg, L. 1969. Stage and sequence: The cognitive-developmental approach to socialization. In *Handbook of socialization theory and research*, ed. D.A. Goslin, 347–480. Chicago: Ran McNally.

Kohlberg, L. 1971. From is to ought. How to commit the naturalistic fallacy and get away with it in the study of moral development. In *Cognitive development and epistemology*, ed. T. Mischel, 151–235. New York: Academic.

Kohlberg, L. 1980. High school democracy and education for a just society. In *Moral education: A first generation in theory and development*, ed. R. Mosher, 20–57. New York: Praeger.

Kohlberg, L. 1984/2003. Moral stages and moralization: The cognitive developmental approach. In *Essays on moral development, Vol. 2. The psychology of moral development: The nature and*

validity of moral stages, ed. L. Kohlberg, 170–205. San Francisco: Harper & Row/Langan-Fox & Shirley.
Koski, S., and E.H.M. Sterck. 2007. Triadic postconflict affiliation in captive Chimpanzees: Does consolation console? *Animal Behaviour* 73: 133–142.
Krebs, D.L. 2008. Morality: An evolutionary account. *Perspectives on Psychological Science* 3: 149–172.
Kripke, S. 1980. *Naming and necessity*. Cambridge, MA: Harvard University Press.
Kripke, S. 1981. *Wittgenstein on rules and private language*. Cambridge, MA: Harvard University Press.
Krones, T. 2006. The scope of the recent bioethics debate in Germany: Kant, crisis and no confidence in society. *Cambridge Quarterly of Healthcare Ethics* 15(3): 273–281.
Krones, T. 2008a. *Kontextsensitive Bioethik, Wissenschaftstheorie und Medizin als Praxis*. Frankfurt am Main: Campus.
Krones, T. 2008b. Pränatal- und Präimplantationsdiagnostik. Diskriminierung von Menschen mit Behinderungen? In *Normal-anders-krank?* ed. D. Gross, S. Müller, and J. Steinmetzer, 435–454. Berlin: Medizinisch wissenschaftliche Verlagsgesellschaft.
Krones, T., E. Schlüter, K. Manolopoulos, K. Bock, H.R. Tinneberg, M.C. Koch, M. Lindner, G.F. Hoffmann, E. Mayatepek, G. Huels, E. Neuwohner, S. El Ansari, T. Wissner, and G. Richter. 2005. Public, expert and patients opinions towards Preimplantation genetic diagnosis in Germany. *Reproductive Biomedicine Online* 10(1): 116–123.
Krones, T., E. Schlüter, E. Neuwohner, S. El Ansari, T. Wissner, and G. Richter. 2006. What is the preimplantation embryo? *Social Science and Medicine* 63(1): 1–20.
Krueger, F., K. McCabe, J. Moll, N. Kriegeskorte, R. Zahn, M. Strenziok, et al. 2007. Neural correlates of trust. *Proceedings of the National Academy of Sciences of the United States of America* 104(50): 20084–20089.
Kryazhimskiy, S., G. Tkacik, and J.B. Plotkin. 2009. The dynamics of adaptation on correlated fitness landscapes. *Proceedings of the National Academy of Sciences of the United States of America* 106(44): 18638–18643.
Kuhn, T. 1962. *The structure of scientific revolution*. Chicago: University of Chicago Press.
Kummer, H., and M. Cords. 1990. Cues of ownership in long-tailed macaques, Macaca fascicularis. *Animal Behavior* 42: 529–549.
Kvanvig, J. 1989. Conservatism and its virtues. *Synthese* 79: 143–163.
La Folette, H. 2000. Pragmatic ethics. In *The Blackwell guide to ethical theory*, ed. H. La Folette, 400–419. Oxford: Oxford University Press.
Lakatos, I. 1977. *The methodology of scientific research programmes: Philosophical papers*, vol. 1. Cambridge: Cambridge University Press.
Lakoff, G., and M. Johnson. 1999. *Philosophy in the flesh: The embodied mind and its challenge to western thought*. New York: HarperCollins Publishers.
Lakshminarayanan, V., and L.R. Santos. 2008. Capuchin monkeys are sensitive to others' welfare. *Current Biology* 18: R999–R1000.
Langford, D.J., S.E. Crager, Z. Shehzad, S.B. Smith, S.G. Sotocinal, J.S. Levenstadt, M.L. Chanda, D.J. Levitin, and J.S. Mogil. 2006. Social modulation of pain as evidence for empathy in mice. *Science* 312(5782): 1967–1970.
Lanius, R.A., E. Vermetten, and D. Pain (eds.). 2010. *The impact of early life trauma on health and disease*. New York: Cambridge University Press.
Lapsley, D.K., and P. Hill. 2008. On dual processing and heuristic approaches to moral cognition. *Journal of Moral Education* 37(3): 313–332.
Lapsley, D.K., and D. Narvaez. 2004. A social-cognitive approach to the moral personality. In *Moral development, self and identity*, ed. D.K. Lapsley and D. Narvaez, 189–212. Mahwah: Erlbaum.
Lapsley, D.K., and D. Narvaez. 2005. The psychological foundations of everyday morality and moral expertise. In *Character psychology and character education*, ed. D.K. Lapsley and F.C. Power, 140–165. Notre Dame: University of Notre Dame Press.
Lapsley, D.K., and D. Narvaez. 2006. Character education. In *Handbook of child psychology*, vol. 4, ed. A. Renninger and I. Siegel, 248–296. New York: Wiley.

Larkin, J.H. 1979. Information processing and science instruction. In *Cognitive process instruction*, ed. J. Lochhead and J. Clements, 109–119. Philadelphia: Franklin Institute Press.
Larkin, J.H., J. McDermott, D. Simon, and H.A. Simon. 1980. Expert and novice performance in solving physics problems. *Science* 208: 1335–1342.
Lecky, W.E.H. 1869/2002. *History of European morals from Augustus to Charlemagne*. Honolulu: University Press of the Pacific.
Leisinger, K.M. 2006. *On corporate responsibility for human rights.* http://www.novartisfoundation.org/. Accessed 15 July 2010.
Lennick, D., and F. Kiel. 2005. *Moral intelligence: Enhancing business performance and leadership success*. Upper Saddle River: Wharton Business Press.
Lerner, J., and P.E. Tetlock. 1999. Accounting for the effects of accountability. *Psychological Bulletin* 125: 255–275.
Lerner, J., J. Goldberg, and P. Tetlock. 1998. Sober second thought: The effects of accountability, anger, and authoritarianism on attributions of responsibility. *Personality and Social Psychology Bulletin* 24: 563–574.
Leung, K., and M.H. Bond. 1984. The impact of cultural collectivism on reward allocation. *Journal of Personality and Social Psychology* 47: 793–804.
Levine, R.V., A. Norenzayan, and K. Philbrick. 2001. Cross-cultural differences in helping strangers. *Journal of Cross-Cultural Psychology* 32: 543–560.
Lewis, H.D. 1948. Collective responsibility. *Philosophy* 23: 3–18.
Lewis, T., F. Amini, and R. Lannon. 2000. *A general theory of love*. New York: Vintage.
Lieberman, M.D. 2007. Social cognitive neuroscience: A review of core processes. *Annual Review of Psychology* 58: 259–289.
Lind, G. 2006. The moral judgment test: Comments on Villegas de Posada's critique. *Psychological Reports* 98(2): 580–584.
Lind, G. 2008. The meaning and measurement of moral judgment competence – A dual aspect theory. In *Contemporary philosophical perspectives on moral development and education*, ed. D. Fasko and W. Willis, 185–220. Cresskill: Hampton Press.
Lind, G., and R.H. Wakenhut. 1980. The assessment of moral judgment competence with a standardized questionnare. *Diagnostica* 26: 312–334.
Lindemann Nelson, J. 2000. Moral teachings from unexpected quarters. Lessons from the social sciences and managed care. *The Hastings Center Report* 30: 12–17.
LoBue, V., T. Nishida, C. Chiong, J.S. DeLoache, and J. Haidt. 2009. When getting something good is bad: Even three-year-olds react to inequality. *Social Development* 20: 154–170.
Loeb, D., and M. Alfano. Forthcoming. Experimental moral philosophy. *Stanford encyclopedia of philosophy*.
Loewenstein, G., and J.S. Lerner. 2003. The role of affect in decision making. In *Handbook of affective sciences*, ed. R.J. Davidson, K.R. Scherer, and H.H. Goldsmith, 619–642. Oxford/New York: Oxford University Press.
Logothetis, N.K. 2008. What we can do and what we cannot do with fMRI. *Nature* 453(7197): 869–878.
London, A.J. 2000. Amenable to reason: Aristotle's Rhetoric and the moral psychology of practical ethics. *Kennedy Institute of Ethics Journal* 10: 287–305.
Lord, R.G., and D.J. Brown. 2001. Leadership, values, and subordinate self-concepts. *The Leadership Quarterly* 12: 133–152.
Lotze, M., R. Veit, S. Anders, and N. Birbaumer. 2007. Evidence for a different role of the ventral and dorsal medial prefrontal cortex for social reactive aggression: An interactive fMRI study. *NeuroImage* 34(1): 470–478.
Luce, M.F., J.R. Bettman, and J.W. Payne. 1997. Choice processing in emotionally difficult decisions. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 23: 384–405.
Lycan, W. 1988. *Judgement and justification*. Cambridge: Cambridge University Press.
Maak, T., and P. Ulrich. 2007. *Integre Unternehmensführung. Ethisches Orientierungswissen für die Wirtschaftspraxis*. Stuttgart: Schäffer-Poeschel Verlag.

MacFarlane, J. 2005. Making sense of relative truth. *Proceedings of the Aristotelian Society* 105(3): 321–339.
Machery, E. 2010. The bleak implications of moral psychology. *Neuroethics* 3(3): 223–231.
Machida, S. 2006. Threat calls in alliance formation by members of a captive group of Japanese macaques. *Primates* 31(2): 205–211.
Macklin, R. 2000. Against relativism. Cultural diversity and the search for ethical universal in medicine. *Theoretical Medicine and Bioethics* 21: 385–392.
MacLean, P.D. 1990. *The triune brain in evolution: Role in paleocerebral functions*. New York: Springer.
Macmillan, M. 2000. *An odd kind of fame. Stories of Phineas Gage*. Cambridge, MA: MIT Press.
Maeng, Y-J. 1996. *Ethische Grundpositionen als Handlungsrechtfertigung interpersonaler Handlungen: Ein Kulturvergleich zwischen Korea (ROK) und Deutschland* [Ethical principles as justification of interpersonal actions: A comparison of the ROK and Germany]. Münster: Waxmann.
Malinowski, B. 1925/1975. *Magic, science, and religion and other essays*. New York: Waveland Press.
Manz, C.C., V. Anand, M. Joshi, and K.P. Manz. 2008. Emerging paradoxes in executive leadership: A theoretical interpretation of the tensions between corruption and virtuous values. *The Leadership Quarterly* 19: 385–392.
Marlowe, F.W. 2003. A critical period for provisioning by Hadza men: Implications for pair bonding. *Evolution and Human Behavior* 24: 217–229.
Marlowe, F.W. 2005. Hunter-gatherers and human evolution. *Evolutionary Anthropology* 14: 54–67.
Marlowe, F.W. 2009. Hadza cooperation: Second-party punishment, yes; third-party punishment, no. *Human Nature* 20: 417–430.
Marlowe, F.W. 2010. *The Hadza Hunter-Gatherers of Tanzania*. London: University of California Press.
Marquardt, N., and R. Hoeger. 2009. The effect of implicit moral attitudes on managerial decision-making: An implicit social cognition approach. *Journal of Business Ethics* 85: 157–171.
Martin, D.E., and B. Austin. 2010. Moral competency inventory validation: Content, construct, convergent and discriminant approaches. *Management Research Review* 33(5): 437–451.
Masserman, J., M.S. Wechkin, and W. Terris. 1964. Altruistic behavior in rhesus monkeys. *The American Journal of Psychiatry* 121: 584–585.
Matusall, S., M. Christen, and I. Kaufmann. 2011. The emergence of social neuroscience as an academic discipline. In *The Oxford handbook of social neuroscience*, ed. J. Decety and J. Cacioppo, 9–27. Oxford: Oxford University Press.
May, D.R., A.Y.L. Chan, T.D. Hodges, and B.J. Avolio. 2003. Developing the moral component of authentic leadership. *Organizational Dynamics* 32: 247–260.
Mayer, R.E. 1983. *Thinking, problem solving, cognition*. New York: Freeman.
Mazar, N., O. Amir, and D. Ariely. 2008. The dishonesty of honest people: A theory of self-concept maintenance. *Journal of Marketing Research* 45(6): 633–644.
McClelland, J.L. 1989. Parallel distributed processing – Implications for cognition and development. In *Parallel distributed processing – Implications for psychology and neurobiology*, ed. R. Morris. Oxford: Clarendon Press.
McGilchrist, I. 2009. *The master and his emissary: The divided brain and the making of the western world*. New Haven: Yale University Press.
McGuire, J., R. Langdon, M. Coltheart, and C. Mackenzie. 2009. A reanalysis of the personal/impersonal distinction in moral psychology research. *Journal of Experimental Social Psychology* 45: 577–580.
Mead, N.L., R.F. Baumeister, F. Gino, M.E. Schweitzer, and D. Ariely. 2009. Too tired to tell the truth: Self-control resource depletion and dishonesty. *Journal of Experimental Social Psychology* 45: 594–597.
Meister, U., C. Finck, Y. Stobel-Richter, G. Schmutzer, and E. Brahler. 2005. Knowledge and attitudes towards preimplantation genetic diagnosis in Germany. *Human Reproduction* 20: 231–238.
Mele, A.R. (ed.). 1997/2003. *The philosophy of action*. Oxford: Oxford University Press.

Melis, A.P., B. Hare, and M. Tomasello. 2008. Do Chimpanzees reciprocate received favours? *Animal Behavior* 76: 951–962.
Melis, A.P., F. Warneken, K. Jensen, A.C. Schneider, J. Call, and M. Tomasello. 2010. Chimpanzees help conspecifics obtain food and non-food items. *Proceedings of the Royal Society B: Biological Sciences* 278: 1405–1413.
Mendez, M., E. Anderson, and J. Shapira. 2005. An investigation of moral judgment in frontotemporal Dementia. *Cognitive and Behavioral Neurology* 18(4): 193–197.
Mepham, B. 2008. *Bioethics. An introduction for the biosciences*, 2nd ed. Oxford/New York: Oxford University Press.
Metzger, M., D.R. Dalton, and J.W. Hill. 1993. The organization of ethics and the ethics of organizations: The case for expanded organizational ethics audits. *Business Ethics Quarterly* 3(1): 27–43.
Meulman, E.J.M., C.M. Sanz, E. Visalberghi, and C.P. Van Schaik. 2012. The role of terrestriality in promoting primate technology. *Evolutionary Anthropology* 21(2): 58–68.
Mikhail, J. 2000. *Rawls' linguistic analogy: A study of the 'generative grammar' model of moral theory described by John Rawls in 'A Theory of Justice.'* Doctoral dissertation, Department of Philosophy, Cornell University.
Mikhail, J. 2007. Universal moral grammar: Theory, evidence and the future. *Trends in Cognitive Sciences* 11(4): 143–152.
Milgram, S. 1963. Behavioral study of obedience. *Journal of Abnormal and Social Psychology* 67(4): 371–378.
Milinski, M., D. Pfluger, D. Kulling, and R. Kettler. 1990. Do sticklebacks cooperate repeatedly in reciprocal pairs? *Behavioral Ecology and Sociobiology* 27: 17–21.
Milinski, M., D. Semmann, and H.-J. Krambeck. 2002. Reputation helps solve the 'tragedy of the commons'. *Nature* 415: 424–426.
Miller, D. 2008. *Grundsätze sozialer Gerechtigkeit (übersetzt von Ulrike Berger)*. Frankfurt/New York: Campus Verlag.
Miller, J.G., and D.M. Bersoff. 1998. The role of liking in perceptions of the moral responsibility to help: A cultural perspective. *Journal of Experimental Social Psychology* 34: 443–469.
Miller, J.G., D.M. Bersoff, and R.L. Harwood. 1990. Perceptions of social responsibilities in India and in the United States: Moral imperatives or personal decisions? *Journal of Personality and Social Psychology* 58: 33–47.
Miner, M., and A. Petocz. 2003. Moral theory in ethical decision making: Problems, clarifications and recommendations from a psychological perspective. *Journal of Business Ethics* 42: 11–25.
Mischel, W., and Y. Shoda. 1995. A cognitive-affective system theory of personality: Reconceptualizing the invariances in personality and the role of situations. *Psychological Review* 102(2): 246–268.
Mitani, J.C. 2006. Reciprocal exchange in Chimpanzees and other primates. In *Cooperation in primates and humans: Evolution and mechanisms*, ed. P. Kapeller and C.P. van Schaik, 101–113. Berlin: Springer.
Mitchell, J.P. 2009. Inferences about mental states. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 364(1521): 1309–1316.
Moll, J., and R. de Oliveira-Souza. 2007. Moral judgments, emotions and the utilitarian brain. *Trends in Cognitive Sciences* 11(8): 319–321.
Moll, J., P.J. Eslinger, and R. de Oliveira-Souza. 2001. Frontopolar and anterior temporal cortex activation in a moral judgment task: Preliminary functional MRI results in normal subjects. *Arquivos de neuro-psiquiatria* 59(3-B): 657–664.
Moll, J., R. de Oliveira-Souza, I. Bramati, and J. Grafman. 2002a. Functional networks in emotional moral and nonmoral social judgments. *NeuroImage* 16: 696–703.
Moll, J., R. de Oliveira-Souza, P.J. Eslinger, I.E. Bramati, J. Mourão-Miranda, P.A. Andreiuolo, and L. Pessoa. 2002b. The neural correlates of moral sensitivity: A functional magnetic resonance imaging investigation of basic and moral emotions. *The Journal of Neuroscience* 22: 2730–2736.
Moll, J., R. de Oliveira-Souza, and P.J. Eslinger. 2003. Morals and the human brain: A working model. *Neuroreport* 14(3): 299–305.

Moll, J., R. de Oliveira-Souza, F.T. Moll, F.A. Ignácio, I.E. Bramati, E.M. Caparelli-Dáquer, et al. 2005a. The moral affiliations of disgust: A functional MRI study. *Cognitive and Behavioral Neurology* 18(1): 68–78.

Moll, J., R. Zahn, R. de Oliveira-Souza, F. Krueger, and J. Grafman. 2005b. The neural basis of human moral cognition. *Nature Reviews Neuroscience* 6(10): 799–809.

Moll, J., F. Krueger, R. Zahn, M. Pardini, R. de Oliveira-Souza, and J. Grafman. 2006. Human fronto-mesolimbic networks guide decisions about charitable donation. *Proceedings of the National Academy of Sciences of the United States of America* 103(42): 15623–15628.

Moll, J., R. de Oliveira-Souza, and R. Zahn. 2008a. The neural basis of moral cognition: Sentiments, concepts, and values. *Annals of the New York Academy of Sciences* 1124: 161–180.

Moll, J., R. de Oliveira-Souza, R. Zahn, and J. Grafman. 2008b. The cognitive neuroscience of moral emotions. In *Moral psychology*, vol. 3, ed. W. Sinnot-Armstrong, 1–17. Cambridge, MA: MIT-Press.

Monin, B., D.A. Pizarro, and J.S. Beer. 2007. Deciding versus reacting: Conceptions of moral judgment and the reason-affect debate. *Review of General Psychology* 11: 99–111.

Monroe, K. 1994. But what else could I do? Choice, identity and a cognitive-perceptual theory of ethical political behavior. *Political Psychology* 15: 201–226.

Monroe, K.R., A. Martin, and P. Ghosh. 2009. Politics and an innate moral sense: Scientific evidence for an old theory? *Political Research Quarterly* 62: 614–634.

Montague, P.R., G.S. Berns, J.D. Cohen, S.M. McClure, G. Pagnoni, M. Dhamala, et al. 2002. Hyperscanning: Simultaneous fMRI during linked social interactions. *NeuroImage* 16(4): 1159–1164.

Moore, G.E. 1903. *Principia Ethica*. Accessible online: http://fair-use.org/g-e-moore/principia-ethica. Accessed 3 Nov 2011.

Moreno, J.D. 1999. Bioethics is a naturalism. In *Pragmatic bioethics*, ed. G. McGee, 5–17. Nashville/London: Vanderbilt University Press.

Moretto, G., E. Làdavas, F. Mattioli, and G. di Pellegrino. 2010. A psychophysiological investigation of moral judgment after ventromedial prefrontal damage. *Journal of Cognitive Neuroscience* 22(8): 1888–1899.

Moscovici, S., and M. Zavalloni. 1969. The group as a polarizer of attitudes. *Journal of Personality and Social Psychology* 12: 125–135.

Moser, P.K., and T.L. Carson. 2000. *Moral relativism: A reader*. New York: Oxford University Press.

Murdock, G.P., and D.R. White. 1969. Standard cross-cultural sample. *Ethnology* 9: 329–369.

Musschenga, A.W. 2001. Naturalness: Beyond animal welfare. *Journal of Agricultural and Environmental Ethics* 15: 171–186.

Musschenga, A.W. 2005. Empirical ethics, context-sensitivity, and contextualism. *The Journal of Medicine and Philosophy* 30: 467–490.

Musschenga, A.W. 2008. Moral judgement and moral reasoning: A critique of Jonathan Haidt. In *The contingent nature of life. Bioethics and the limits of human existence*, ed. M. Düwell et al., 131–147. Dordrecht: Springer.

Musschenga, A.W. 2009. Moral intuitions, moral expertise, and moral reasoning. *The Journal of Philosophy of Education* 43: 597–613.

Musschenga, A.W. 2010a. Empirical ethics and the special place of the practitioner's moral judgements. In *Ethics and empirics. Strange and fragile bedfellows*, special issue of ethical perspectives 17(2), ed. V. Draulans et al., 231–258. Leuven: Peeters.

Musschenga, A.W. 2010b. The epistemic value of psychological moral intuitions. *Philosophical Explorations* 13: 113–128.

Nadelhoffer, T. 2006. Bad acts, blameworthy agents, and intentional actions. Some problems for juror impartiality. *Philosophical Explorations* 9(2): 203–219.

Nagel, T. 1979. *Moral luck. Mortal questions*. Cambridge: Cambridge University Press.

Nakamura, M., and N. Itoh. 2001. Sharing of wild fruits among male Chimpanzees: Two cases from Mahale, Tanzania. *Pan African News* 8: 67–70.

Narvaez, D. 1998. The effects of moral schemas on the reconstruction of moral narratives in 8th grade and college students. *Journal of Educational Psychology* 90(1): 13–24.

Narvaez, D. 1999. Using discourse processing methods to study moral thinking. *Educational Psychology Review* 11(4): 377–394.
Narvaez, D. 2001. Moral text comprehension: Implications for education and research. *Journal of Moral Education* 30(1): 43–54.
Narvaez, D. 2005. The neo-Kohlbergian tradition and beyond: Schemas, expertise and character. In *Nebraska symposium on motivation, Vol. 51. Moral motivation through the lifespan*, ed. G. Carlo and C. Pope-Edwards, 119–163. Lincoln: University of Nebraska Press.
Narvaez, D. 2006. Integrative ethical education. In *Handbook of moral development*, ed. M. Killen and J. Smetana, 703–733. Mahwah: Erlbaum.
Narvaez, D. 2008. Triune ethics: The neurobiological roots of our multiple moralities. *New Ideas in Psychology* 26: 95–119.
Narvaez, D. 2009. *Nurturing character in the classroom*, EthEx series, Book 4: Ethical action. Notre Dame: ACE Press.
Narvaez, D. 2010a. Moral complexity: The fatal attraction of truthiness and the importance of mature moral functioning. *Perspectives on Psychological Science* 5(2): 163–181.
Narvaez, D. 2010b. The emotional foundations of high moral intelligence. In *Children's moral emotions and moral cognition: Developmental and educational perspectives*, New directions for child and adolescent development 129, ed. B. Latzko and T. Malti, 77–94. San Francisco: Jossey-Bass.
Narvaez, D. 2013a. Development and socialization within an evolutionary context: Growing up to become "A good and useful human being". In *War, peace and human nature: The convergence of evolutionary and cultural views*, ed. D. Fry, 643–672. New York: Oxford University Press.
Narvaez, D. 2013b. *The neurobiology and development of human morality*. New York: W.W. Norton. Book (forthcoming).
Narvaez, D., and T. Bock. 2002. Moral schemas and tacit judgement or how the defining issues test is supported by cognitive science. *Journal of Moral Education* 31(3): 297–314.
Narvaez, D., and T. Bock. 2009. *Nurturing character in the classroom*, EthEx series, Book 2: Ethical judgment. Notre Dame: ACE Press.
Narvaez, D., and L. Endicott. 2009. *Nurturing character in the classroom*, EthEx series, Book 1: Ethical sensitivity. Notre Dame: ACE Press.
Narvaez, D., and T. Gleason. 2007. The Influence of moral judgment development and moral experience on comprehension of moral narratives and expository texts. *The Journal of Genetic Psychology* 168(3): 251–276.
Narvaez, D., and T. Gleason. 2013. Developmental optimization. In *Human nature, early experience and the environment of evolutionary adaptedness*, ed. D. Narvaez, J. Panksepp, A. Schore, and T. Gleason, 307–325, New York: Oxford University Press.
Narvaez, D., and D. Lapsley. 2005. The psychological foundations of everyday morality and moral expertise. In *Character psychology and character education*, ed. D. Lapsley and C. Power, 140–165. Notre Dame: University of Notre Dame Press.
Narvaez, D., and D.K. Lapsley (eds.). 2009. *Personality, identity, and character: Explorations in moral psychology*. New York: Cambridge University Press.
Narvaez, D., and J. Lies. 2009. *Nurturing character in the classroom*, EthEx series, Book 3: Ethical motivation. Notre Dame: ACE Press.
Narvaez, D., and J. Rest. 1995. The four components of acting morally. In *Moral behavior and moral development: An introduction*, ed. W. Kurtines and J. Gewirtz, 385–400. New York: McGraw-Hill.
Narvaez, D., I. Getz, J.R. Rest, and S. Thoma. 1999a. Individual moral judgment and cultural ideologies. *Developmental Psychology* 35: 478–488.
Narvaez, D., T. Gleason, C. Mitchell, and J. Bentley. 1999b. Moral theme comprehension in children. *Journal of Educational Psychology* 91(3): 477–487.
Narvaez, D., D. Lapsley, S. Hagele, and B. Lasky. 2006. Moral chronicity and social information processing: Tests of a social cognitive approach to the moral personality. *Journal of Research in Personality* 40: 966–985.

Narvaez, D., J. Brooks, and B. Mattan. 2011a. *Attachment-related variables predict moral mindset and moral action*. Montreal: Society for Research in Child Development.
Narvaez, D., K. Mrkva, B. Bettonville, A. Prister, E. Mullen, and K. Delgado. 2011b. The crying baby: Moral identity influences moral perception. Association for Moral Education annual meeting, Nanjing, China.
Narvaez, D., J. Panksepp, A. Schore, and T. Gleason (eds.). 2013a. *Human nature, early experience and the environment of evolutionary adaptedness*. New York: Oxford University Press.
Narvaez, D., T. Gleason, L. Wang, J. Brooks, J. Lefever, A. Cheng, and Centers for the Prevention of Child Neglect. 2013b. The evolved development Niche: Longitudinal effects of caregiving practices on early childhood psychosocial development. *Early Childhood Research Quarterly* 28(4): 759–773.
Narveson, J. 2002. Collective responsibility. *The Journal of Ethics* 6: 179–198.
Nationaler Ethikrat. 2003. *Stellungnahme genetische Diagnostik vor und während der Schwangerschaft*. Online. http://www.ethikrat.org/publikationen/stellungnahmen. Accessed 20 Sept 2008.
Neisser, U. 1976. *Cognition and reality*. New York: W.H. Freeman and Company.
Nelkin, D. 2008. Moral luck. *Stanford encyclopedia of philosophy*. http://plato.stanford.edu/
Nelkin, D., and S. Lindee. 1995. *The DNA-mystique. The gene as cultural icon*. New York: Freeman.
Nelson, K., and J. Gruendel. 1981. Generalized event representations: Basic building blocks of cognitive development. In *Advances in developmental psychology*, ed. M. Lamb and A. Brown, 131–158. Hillsdale: Erlbaum.
Nichols, S. 2002. Norms with feeling: Towards a psychological account of moral judgment. *Cognition* 84(2): 221–236.
Nichols, S. 2004. *Sentimental rules: On the natural foundations of moral judgment*. New York: Oxford University Press.
Nichols, S. 2008. Moral rationalism and empirical immunity. In *Moral psychology, Vol. 3. The neuroscience of morality*, ed. W. Sinnott-Armstrong, 395–407. Cambridge, MA: MIT Press.
Nichols, S., and J. Knobe. 2007. Moral responsibility and determinism: The cognitive science of folk intuitions. *Noûs* 41(4): 663–685.
Nida-Rümelin, J. 2006. Theoretische und angewandte Ethik: Paradigmen, Begründungen, Bereiche. In *Angewandte Ethik. Die Bereichsethiken und ihre theoretische Fundierung*, ed. ibid., 2–85. Stuttgart: Alfred Kröner Verlag.
Nielsen, K. 1982a. Grounding rights and a method of reflective equilibrium. *Inquiry* 25: 277–306.
Nielsen, K. 1982b. On needing a moral theory. *Metaphilosophy* 13: 97–111.
Nietzsche, F. 1887/2009. *On the genealogy of morals: A Polemic*. Oxford: Oxford University Press.
Nisan, M. 1987. Moral norms and social conventions: A cross-cultural comparison. *Developmental Psychology* 23: 719–725.
Nishida, T., and K. Hosaka. 1996. Coalition strategies among adult male Chimpanzees of the Mahale Mountains, Tanzania. In *Great Ape Societies*, ed. W.C. McGrew, L.F. Marchant, and T. Nishida, 114–134. Cambridge: Cambridge University Press.
Nishida, T., T. Hasegawa, H. Hayaki, Y. Takahata, and S. Uehara. 1992. Meat-sharing as a coalition strategy by an alpha male Chimpanzee? In *Topics in primatology: Human origins*, vol. 1, ed. T. Nishida, W.C. McGrew, P. Marler, M. Pickford, and F.B.M. de Waal, 159–174. Tokyo: University of Tokyo Press.
Nitsche, M.A., D. Liebetanz, N. Lang, A. Antal, F. Tergau, and W. Paulus. 2003. Safety criteria for transcranial direct current stimulation (tDCS) in humans. *Clinical Neurophysiology* 239(114): 2220–2222.
Norman, W. 1998. 'Inevitable and unacceptable'. Methodological Rawlsianism in Anglo-American political philosophy. *Political Studies* 46: 276–294.

Nozick, R. 1981. *Philosophical explanations*. Cambridge, MA: Belknap Press.
Nussbaum, M.C. 1986. *The fragility of goodness. Luck and ethics in Greek tragedy and philosophy*. Cambridge/New York: Cambridge University Press.
Nussbaum, M.C. 2000. Why practice needs ethical theory: Particularism, principle, and bad behaviour. In *Moral particularism*, ed. B. Hooker and M.O. Little, 227–256. Oxford: Clarendon Press.
Palagi, E., D. Antonacci, and I. Norscia. 2008. Peacemaking on treetops: First evidence of reconciliation from a wild prosimian (Propithecus verreauxi). *Animal Behavior* 76: 737–747.
Palazzo, G. 2007. Organizational integrity – Understanding the dimensions of ethical and unethical behavior in corporations. In *Corporate ethics and corporate governance*, ed. W.C. Zimmerli, K. Richter, and M. Holzinger, 113–128. Berlin: Springer.
Panksepp, J. 1998. *Affective neuroscience. The foundations of human and animal emotions*. New York: Oxford University Press.
Panksepp, J., and J. Watt. 2011. Why does depression hurt? Ancestral primary-process separation-distress (PANIC) and diminished brain reward (SEEKING) processes in the genesis of depressive affect. *Psychiatry* 74: 5–14.
Parfit, D. 1984. *Reasons and persons*. Oxford: Clarendon.
Parkinson, C., W. Sinnott-Armstrong, P. Koralus, A. Mendelovici, V. McGeer, and T. Wheatley. 2011. Is morality unified? Evidence that distinct neural systems underlie moral judgments of harm, dishonesty, and disgust. *Journal of Cognitive Neuroscience* 32(10): 3162–3180.
Payne, L.S. 2006. A compass for decision making. In *Responsible leadership*, ed. T. Maak and N. Pless. London: Routledge.
Pedersen, L.J.T. 2009. See no evil: Moral sensitivity in the formulation of business problems. *Business Ethics: A European Review* 18: 335–348.
Pellegrino, E.D. 1995. The limitation of empirical research in ethics. *The Journal of Clinical Ethics* 6: 162.
Perry, S. 1997. Male-female social relationships in wild white-faced capuchin monkeys, Cebus capucinus. *Behaviour* 134: 477–510.
Perugini, M., and L. Leone. 2009. Implicit self-concept and moral action. *Journal of Research in Personality* 43: 747–754.
Pessoa, L. 2008. On the relationship between emotion and cognition. *Nature Reviews Neuroscience* 9(2): 148–158.
Petty, R.E., and J.T. Cacioppo. 1986. The elaboration likelihood model of persuasion. In *Advances in experimental social psychology*, vol. 19, ed. L. Berkowitz, 123–205. New York: Academic.
Piaget, J. 1932/1965/1997. *The moral judgment of the child*. New York: Free Press (earlier issues: New York/London: Simon & Schuster/Kegan, Paul).
Pincoffs, E.L. 1986. *Quandaries and virtues. Against reductivism in ethics*. Lawrence: University of Kansas Press.
Pinker, J. 2007. A history of violence: We're getting nicer every day. *New Republic*, March 19: 1–4.
Pizarro, D.A., and P. Bloom. 2003. The intelligence of the moral intuitions: Comment on Haidt (2001). *Psychological Review* 110(1): 193–196.
Plato. 1873. The laws. In *The dialogues of Plato*, vol. IV. Trans. B. Jowett. New York: Scribner, Armstrong, and Co.
Plessner, H., and S. Czenna. 2008. The benefits of intuition. In *Intuition in judgment and decision making*, ed. H. Plessner, C. Betsch, and T. Betsch, 251–266. Mahwah: Lawrence Erlbaum.
Pobiner, B.L., M.J. Rogers, C.M. Monahan, and J.W.K. Harris. 2008. New evidence for hominin carcass processing strategies at 1.5 Ma, Koobi Fora, Kenya. *Journal of Human Evolution* 55: 103–130.
Poldrack, R.A. 2006. Can cognitive processes be inferred from neuroimaging data? *Trends in Cognitive Sciences* 10(2): 59–63.
Porges, S.P. 2011. *Polyvagal theory*. New York: WW Norton.
Posner, M., and C. Snyder. 1975. Attention and cognitive control. In *Information processing and cognition: The Loyola symposium*, ed. R.L. Solso. Hillsdale: Erlbaum.
Prehn, K., and H.R. Heekeren. 2009. Moral judgment and the brain: A functional approach to the question of emotion and cognition in moral judgment integrating psychology, neuroscience and

evolutionary biology. In *The moral brain*, ed. J. Verplaetse, J. de Schrijver, S. Vanneste, and J. Braeckman, 129–154. Dordrecht: Springer.

Prehn, K., I. Wartenburger, K. Mériau, C. Scheibe, O.R. Goodenough, A. Villringer, et al. 2008. Individual differences in moral judgment competence influence neural correlates of socio-normative judgments. *Social Cognitive and Affective Neuroscience* 3(1): 33–46.

Prehn, K., F. Schlagenhauf, L. Schulze, C. Berger, K. Vohs, M. Fleischer, et al. 2013. Neural correlates of risk taking in violent criminal offenders characterized by emotional hypo- and hyper-reactivity. *Social Neuroscience* 8(2): 136–147.

Preuschoft, S., and J.A.R.A.M. van Hooff. 1997. The social function of "smile" and "laughter": Variations across primate species and societies. In *Nonverbal communication: Where nature meets culture*, ed. U. Segerstråle and P. Molnar, 171–189. Mahwah: Lawrence Erlbaum Assoc., Inc.

Preuschoft, S., and C.P. van Schaik. 2000. Dominance, social relationships and conflict management. In *Conflict management*, ed. F. Aureli and F.B.M. de Waal, 77–105. Berkeley: California University Press.

Price, T. 2008. *Leadership ethics. An introduction*. Cambridge: Cambridge University Press.

Prichard, H.A. 1912. Does moral philosophy rest on a mistake? *Mind* 21: 21–37.

Prinz, J.J. 2007a. Is morality innate? In *Moral psychology, Vol. 1. Evolution of morals*, ed. W. Sinnott-Armstrong, 367–406. Cambridge, MA: MIT Press.

Prinz, J.J. 2007b. *The emotional construction of morals*. New York: Oxford University Press.

Prinz, J.J. 2009. Summary. *Analysis* 69(4): 701–704.

Prinz, J.J. in press. Naturalizing metaethics. *Ethics*.

Pryor, J. 2000. The sceptic and the Dogmatist. *Noûs* 34(4): 517–549.

Pullman, D. 2005. Ethics first Aid. Reframing the role of 'Principlism' in clinical ethics education and practice. *The Journal of Clinical Ethics* 16: 223–229.

Putnam, H. 1975. *Mind, language and reality*, Philosophical papers, vol. 2. Cambridge: Cambridge University Press.

Quante, M. 2003. *Einführung in die Allgemeine Ethik*. Darmstadt: Wissenschaftliche Buchgesellschaft.

Rabbie, J.M. 1992. The effects of intragroup cooperation and intergroup competition on in-group cohesion and out-group hostility. In *Coalitions and alliances in humans and other animals*, ed. A.H. Harcourt and F.B.M. de Waal, 175–205. Oxford: Oxford University Press.

Radzik, L.A. 2002. Coherentist theory of normative authority. *The Journal of Ethics* 6: 21–42.

Railton, P. 1984. Alienation, consequentialism, and the demands of morality. *Philosophy and Public Affairs* 13: 134–171.

Raine, A., and Y. Yang. 2006. Neural foundations to moral reasoning and antisocial behavior. *Social Cognitive and Affective Neuroscience* 1(3): 203–213.

Range, F., L. Horn, Z. Viranyi, and L. Huber. 2008. The absence of reward induces inequity aversion in dogs. *Proceedings of the National Academy of Sciences of the United States of America* 106(1): 340–345.

Rapp, R. 2000. *Testing women, testing the fetus. The social impact of amniocentesis in America*. New York: Routledge.

Rasmussen, L. (ed.). 2005. *Ethics expertise: History, contemporary perspectives, and applications*. Dordrecht: Springer.

Rauprich, O. 2008. Common morality: Comment on Beauchamp and Childress. *Theoretical Medicine and Bioethics* 29: 43–71.

Rawls, J. 1951. Outline of a decision procedure in ethics. *Philosophical Review* 60: 177–197.

Rawls, J. 1971/1999. *A theory of justice*. Cambridge, MA: Harvard University Press (revised edition).

Rawls, J. 2005. *Political liberalism, expanded version*. New York: Columbia University Press.

Reber, A.S. 1993. *Implicit learning and tacit knowledge: An essay on the cognitive unconscious*. New York: Oxford University Press.

Resick, C.J., P.J. Hanges, M.W. Dickson, and J.K. Mitchelson. 2006. A cross-cultural examination of the endorsement of ethical leadership. *Journal of Business Ethics* 63: 345.

Rest, J.R. 1974. *Manual for the defining issue test: An objective test for moral judgment development*. Minneapolis: University of Minneapolis Press.
Rest, J.R. 1983. Morality. In *Cognitive development. Manual of child psychology*, vol. 3, 4th ed, ed. J.H. Flavell and E. Markman, 556–629. New York: Wiley.
Rest, J.R. 1986. *Moral development: Advances in research and theory*. New York: Praeger.
Rest, J.R., and D. Narvaez (eds.). 1994. *Moral development in the professions: Psychology and applied ethics*. Hillsdale: Lawrence Erlbaum.
Rest, J.R., D. Narvaez, M.J. Bebeau, and S.J. Thoma. 1999. *Postconventional moral thinking: A Neo-Kohlbergian approach*. Mahwah: Lawrence Erlbaum.
Reynolds, S.J. 2006. A neurocognitive model of the ethical decision-making process: Implications for study and practice. *Journal of Applied Psychology* 91(4): 737–748.
Reynolds, S.J. 2008. Moral attentiveness: Who pays attention to the moral aspects of life? *Journal of Applied Psychology* 93(5): 1027–1041.
Richards, N. 1986. Luck and desert. *Mind* 65: 198–209.
Richardson, H.S. 2000. Specifying, balancing and interpreting bioethical principles. *The Journal of Medicine and Philosophy* 25: 285–307.
Richter, G. 2007. Greater patient, family, and surrogate involvement in clinical ethics consultation: The model of clinical ethics liaison service as a measure for preventive ethics. *HEC Forum* 19: 324–337.
Ridley, M. 1996. *The origins of virtue. Human instincts and the evolution of cooperation*. London: Penguin Books.
Rilling, J., D. Gutman, T. Zeh, G. Pagnoni, G. Berns, and C. Kilts. 2002. A neural basis for social cooperation. *Neuron* 35(2): 395–405.
Rippe, K.P. 1998. Ethikkommissionen als Expertengremien? In *Angewandte Ethik in der pluralistischen Gesellschaft*, ed. ibid. Freiburg: Academic.
Rizzolatti, G., and M. Fabbri-Destro. 2010. Mirror neurons: From discovery to autism. *Experimental Brain Research* 200(3–4): 223–237.
Roberts, R. 2003. *Emotions. An essay in aid of moral psychology*. Cambridge: Cambridge University Press.
Robertson, D., J. Snarey, O. Ousley, K. Harenski, F. DuBois Bowman, R. Gilkey, et al. 2007. The neural processing of moral sensitivity to issues of justice and care. *Neuropsychologia* 45(4): 755–766.
Romero, T., and F. Aureli. 2008. Reciprocity of support in coaties (Nasua nasua). *Journal of Comparative Psychology* 122(1): 19–25.
Rorty, R. 1979. *Philosophy and the mirror of nature*. Princeton: Princeton University Press.
Rorty, R. 1984. Habermas and Lyotard on Post-Modernity. *Praxis International* 1: 32–44.
Rorty, R. 1989. *Contingency, irony, and solidarity*. Cambridge: Cambridge University Press.
Rosas, A. 2007. Beyond the sociobiological dilemma: Social emotions and the evolution of morality. *Zygon* 42: 685–699.
Rosch, E. 1973. Natural categories. *Cognitive Psychology* 4: 324–350.
Rosebury, B. 1995. Moral responsibility and moral luck. *Philosophical Review* 104: 499–524.
Roskies, A. 2002. Neuroethics for the new millennium. *Neuron* 35: 21–23.
Rossi, S., M. Hallett, P.M. Rossini, and A. Pascual-Leone. 2009. Safety, ethical considerations, and application guidelines for the use of transcranial magnetic stimulation in clinical practice and research. *Clinical Neurophysiology* 120(12): 2008–2039.
Rowe, G., J.B. Hirsh, and A.K. Anderson. 2007. Positive affect increases the breadth of attentional selection. *Proceedings of the National Academy of Sciences of the United States of America* 104(1): 383–388.
Royzman, E., and R. Kumar. 2004. Is consequential luck morally inconsequential? Empirical psychology and the reassessment of moral luck. *Ratio* 17: 329–344.
Royzman, E., K.W. Cassidy, and J. Baron. 2003. 'I Know, You Know': Epistemic egocentrism in children and adults. *Review of General Psychology* 7: 38–65.
Rozin, P., L. Lowery, S. Imada, and J. Haidt. 1999. The CAD triad hypothesis: A mapping between three moral emotions (contempt, anger, disgust) and three moral codes (community, autonomy, divinity). *Journal of Personality and Social Psychology* 76: 574–586.

Rudolf von Rohr, C., J.M. Burkart, and C.P. van Schaik. 2011. Evolutionary precursors of social norms in Chimpanzees: A new approach. *Biology and Philosophy* 26(1): 1–30.

Rudolf von Rohr, C., S.E. Koski, J.M. Burkart, C. Caws, O.N. Fraser, A. Ziltener, and C.P. van Schaik. 2012. Impartial third-party interventions in captive Chimpanzees: A reflection of community concern. *PloS One* 7(3): e32494.

Russell, D. 2009. *Practical intelligence and the virtues*. Oxford: Oxford University Press.

Russell, R.F. 2001. The role of values in servant leadership. *Leadership & Organization Development Journal* 22: 76–83.

Russell, Y.I., J. Call, and R.I.M. Dunbar. 2008. Image scoring in great apes. *Behavioural Processes* 78: 108–111.

Rutgers, B., and R. Heeger. 1999. Inherent worth and respect for animal integrity. In *Recognizing the intrinsic value of animals*, ed. M. Dol, M. Fentener van Vlissingen, S. Kasanmoentalib, T. Visser, and H. Zwart, 41–53. Assen: Van Gorcum.

Rutte, C., and M. Taborsky. 2007. Generalized reciprocity in rats. *PLoS Biology* 5(7): e196.

Rypma, B., J.S. Berger, V. Prabhakaran, B.M. Bly, D.Y. Kimberg, B.B. Biswal, et al. 2006. Neural correlates of cognitive efficiency. *NeuroImage* 33(3): 969–979.

Sahlins, M. 2008. *The Western illusion of human nature*. Chicago: Prickly Paradigm Press.

Salovey, P., and J.D. Mayer. 1990. Emotional intelligence. *Imagination, Cognition, and Personality* 9: 185–211.

Sanday, P.R. 1986. *Divine hunger: Cannibalism as a cultural system*. Cambridge: Cambridge University Press.

Sanfey, A.G., J.A. Rilling, J.K. Aronson, L. Nystrom, and J.D. Cohen. 2003. The neural basis of economic decision making in the Ultimatum Game. *Science* 300: 1755–1757.

Sapolsky, R., and L. Share. 2004. A pacific culture among wild baboons: Its emergence and transmission. *PLoS Biology* 2(4): e106.

Saver, J.L., and A.R. Damasio. 1991. Preserved access and processing of social knowledge in a patient with acquired sociopathy due to ventromedial frontal damage. *Neuropsychologia* 29(12): 1241–1249.

Saxe, R., and N. Kanwisher. 2003. People thinking about thinking people. The role of the temporo-parietal junction in "theory of mind". *NeuroImage* 19(4): 1835–1842.

Saxe, R., S. Carey, and N. Kanwisher. 2004. Understanding other minds: Linking developmental psychology and functional neuroimaging. *Annual Review of Psychology* 55: 87–124.

Sayre-McCord, G. 1996. Coherentist epistemology and moral theory. In *Moral knowledge? New readings in moral epistemology*, ed. W. Sinnot-Armstrong and M. Timmons, 137–189. New York: Oxford University Press.

Scanlon, T.M. 1998. *What we owe to each other*. Cambridge, MA: Harvard University Press.

Scanlon, T.M. 2003. Rawls on justification. In *The Cambridge companion to Rawls*, ed. S. Freeman, 139–168. Cambridge/New York: Cambridge University Press.

Schaber, P. 2003. Menschenwürde als das Recht, nicht erniedrigt zu werden. In *Menschenwürde – Annäherung an einen Begriff*, ed. R. Stoecker, 119–131. Wien: ÖBV + HPT Verlagsgesellschaft GmbH.

Schaber, P. 2011. Was moralische von altruistischen Motiven unterscheidet. In *Naturalismus in der Ethik. Perspektiven und Grenzen*, ed. T. Schmidt and T. Tarkian. Paderborn: Mentis-Verlag.

Schaich Borg, J., C. Hynes, J. Van Horn, S. Grafton, and W. Sinnott-Armstrong. 2006. Consequences, action, and intention as factors in moral judgments: An FMRI investigation. *Journal of Cognitive Neuroscience* 18(5): 803–817.

Schino, G. 1998. Reconciliation in domestic goats. *Behaviour* 135(3): 343–356.

Schino, G., and F. Aureli. 2008. Grooming reciprocation among female primates: A meta-analysis. *Biology Letters* 4: 9–11.

Schino, G., and F. Aureli. 2010. Primate reciprocity and its cognitive requirements. *Evolutionary Anthropology* 19: 130–135.

Schlenker, B.R. 2008. Integrity and character: Implications of principled and expedient ethical ideologies. *Journal of Social and Clinical Psychology* 27(10): 1078–1125.

Schmidt, W.H., and B.Z. Posner. 1982. *Managerial values and expectations: The silent power in personal and organizational life*. New York: Amacom.

Schmitz, T.W., E. DeRosa, and A.K. Anderson. 2009. Opposing influences of affective state valence on visual cortical encoding. *Journal of Neuroscience* 29(22): 7199–7207.
Schnall, S., J. Haidt, G. Clore, and A. Jordan. 2008. Disgust as embodied moral judgment. *Personality and Social Psychology Bulletin* 34: 1096–1109.
Scholl, B. 2007. Object persistence in philosophy and psychology. *Mind and Language* 22: 563–591.
Schore, A.N. 1994. *Affect regulation and the origin of the self: The neurobiology of emotional development*. Mahwah: Erlbaum.
Schore, A.N. 2013. Bowlby's "Environment of evolutionary adaptedness": Recent studies on the interpersonal neurobiology of attachment and emotional development. In *Evolution, early experience and human development: From research to practice and policy*, ed. D. Narvaez, J. Panksepp, A. Schore, and T. Gleason. New York: Oxford University Press.
Schwartz, S.H. 1992. Universals in the content and structure of values: Theoretical advances and empirical tests in 20 countries. In *Advances in experimental social psychology*, vol. 25, ed. M.P. Zanna, 1–65. New York: Academic.
Schwartz, S.H., G. Melech, A. Lehmann, S. Burgess, M. Harris, and V. Owens. 2001. Extending the cross-cultural validity of the theory of basic human values with a different method of measurement. *Journal of Cross-Cultural Psychology* 32(5): 519–542.
Schwarz, N., and G.L. Clore. 1983. Mood, misattribution and judgments of well-being: Informative and directive functions of affective states. *Journal of Personality and Social Psychology* 45: 513–523.
Schwitzgebel, E., and F. Cushman. 2012. Expertise in moral reasoning? Order effects on moral judgments in professional philosophers and non-philosophers. *Mind and Language* 27(2): 135–153.
Sear, R., and R. Mace. 2008. Who keeps children alive? A review of the effects of kin on child survival. *Evolution and Human Behavior* 29: 1–18.
Seidel, A., and J.J. Prinz. 2013. Mad and glad: Musically induced emotions have divergent moral impact. *Motivation and Emotion* 37(3): 629–637.
Sekerka, L.E., and R.P. Bagozzi. 2007. Moral courage in the workplace: Moving to and from the desire and decision to act. *Business Ethics: A European Review* 16(2): 132–149.
Sellars, W. 1956. *Empiricism and the philosophy of mind*. Cambridge, MA: Harvard University Press.
Shafer-Landau, R. 2003. *Moral realism: A defence*. New York: Oxford University Press.
Shaw, B. 1992. Explaining incest: Brother-sister marriage in Graeco-Roman Egypt. *Man* 27: 267–299.
Shaw, R.E., M.T. Turvey, and W.M. Mace. 1982. Ecological psychology. The consequence of a commitment to realism. In *Cognition and the symbolic processes*, vol. 2, ed. W. Weimer and D. Palermo, 159–226. Hillsdale: Erlbaum.
Shiffrin, R.M., and W. Schneider. 1977. Controlled and automatic human information processing: II. Perceptual learning, automatic attending and a general theory. *Psychological Review* 84: 127–189.
Shweder, R.A., N.C. Much, M. Mahapatra, and L. Park. 1997. The "Big Three" of morality (autonomy, community, divinity): and the "Big Three" explanations of suffering. In *Morality and health*, ed. P. Rozin and A. Brandt. New York: Routledge.
Sigg, H., and J. Falett. 1985. Experiments on the respect of possession and property in hamadryas baboons (Papio hamadryas). *Animal Behavior* 33: 978–984.
Silk, J.B. 1979. Feeding, foraging and food sharing behavior of immature Chimpanzees. *Folia Primatologica* 31: 123–142.
Silk, J.B., S.F. Brosnan, J. Vonk, J. Henrich, D.J. Povinelli, A.S. Richardson, et al. 2005. Chimpanzees are indifferent to the welfare of unrelated group members. *Nature* 437: 1357–1359.
Simon, H.A. 1955. A behavioral model of rational choice. *Quarterly Journal of Economics* 69: 99–118.
Simon, H.A. 1980. *Models of thought*. Vols. 1 and 2. New Haven: Yale University Press.
Simpson, E. 1999. Between internalism and externalism in ethics. *The Philosophical Quarterly* 49(195): 201–214.

Singer, P. 1974. Sidgwick and reflective equilibrium. *The Monist* 58: 490–517.
Singer, P. 1975/2002. *Animal liberation*. New York: Harper Collins.
Singer, P. 1981. *The expanding circle: Ethics and sociobiology*. New York: Farrar, Straus & Giroux.
Singer, M.S. 2000. Ethical and fair work behavior: A normative-empirical dialogue concerning ethics and justice. *Journal of Business Ethics* 28: 187–209.
Singer, P. 2005. Ethics and intuitions. *The Journal of Ethics* 9: 331–352.
Singer, P. 2007. Hunger, Wohlstand und Moral. In *Weltarmut und Ethik*, ed. B. Bleisch and P. Schaber, 37–52. Paderborn: Mentis-Verlag.
Singer, P.A., E.D. Pellegrino, and M. Siegler. 2001. Clinical ethics revisited. *BMC Medical Ethics* 2: e1.
Singer, T., B. Seymour, J.P. O'Doherty, K.E. Stephan, R.J. Dolan, and C.D. Frith. 2006. Empathic neural responses are modulated by the perceived fairness of others. *Nature* 439: 466–469.
Sinnott-Armstrong, W. 2006. Moral intuitionism meets empirical psychology. In *Metaethics after Moore*, ed. T. Horgan and M. Timmons, 339–365. Oxford: Clarendon Press.
Sinnott-Armstrong, W. 2008a. Framing moral intuitions. In *Moral psychology, Vol. 2. The cognitive science of morality: Intuition and diversity*, ed. W. Sinnott-Armstrong, 47–77. Cambridge, MA: MIT Press.
Sinnott-Armstrong, W. 2008b. Is moral phenomenology unified? *Phenomenology and the Cognitive Sciences* 7(1): 85–97.
Sinnott-Armstrong, W., and T. Wheatley. 2012. The disunity of morality and why it matters to philosophy. *The Monist* 95(3): 355–377.
Skitka, L.J., C.W. Bauman, and E.G. Sargis. 2005. Moral conviction: Another contributor to attitude strength or something more? *Journal of Personality and Social Psychology* 88: 895–917.
Slatter, P.E. 1987. *Building expert systems: Cognitive emulation*. Chichester: Ellis Horwood.
Sloboda, J.A. 1991. Musical expertise. In *Toward a general theory of expertise*, ed. K.A. Ericsson and J. Smith, 153–171. New York: Cambridge University Press.
Slocombe, K.E., and N.E. Newton-Fisher. 2005. Fruit sharing between wild adult Chimpanzees (Pan troglodytes schweinfurthii): A socially significant event? *American Journal of Primatology* 65(4): 385–391.
Sloman, S.A. 1996. The empirical case for two systems of reasoning. *Psychological Bulletin* 119: 3–22.
Sloman, S.A. 2002. Two systems of reasoning. In *Heuristics and biases*, ed. T. Gilovich, D. Griffin, and D. Kahneman, 379–396. New York: Cambridge University Press.
Slovic, P., M. Finucane, E. Peters, and D.G. MacGregor. 2002. The affect heuristic. In *Heuristics and biases*, ed. T. Gilovich, D. Griffin, and D. Kahneman, 397–420. New York: Cambridge University Press.
Smith, A. 1759/2000/2009. *The theory of moral sentiments*. Amherst/London: Prometheus Books/Penguin Books Ltd (2009 edition).
Smith, M. 1987. The Humean theory of motivation. *Mind* 96(381): 36–61.
Smith, M. 2004. *Ethics and the A Priori*. Cambridge: Cambridge University Press.
Smith, E.R., and J. DeCoster. 2000. Dual process models in social and cognitive psychology: Conceptual integration and links to underlying memory systems. *Personality and Social Psychology Review* 4: 108–131.
Smith, E., and D. Medin. 1981. *Categories and concepts*. Cambridge, MA: Harvard University Press.
Snarey, J.R. 1985. Cross-cultural universality of social-moral development: A critical review of Kohlbergian research. *Psychological Bulletin* 97(2): 202–232.
Solomon, R.C. 1992. *Ethics and excellence. Cooperation and integrity in business*. Oxford: Oxford University Press.
Solomon, R.C. 2003. Victims of circumstances? A defense of virtue ethics in business. *Business Ethics Quarterly* 13(1): 43–62.
Solomon, R.C. 2009. Business ethics and virtue. In *A companion to business ethics*, ed. R.E. Frederick, 30–37. Malden: Blackwell.

Solter, D., D. Beyleveld, M.B. Friele, J. Holówka, R. Pardo Avellaneda, R. Lovell-Badge, C. Mandla, U. Martin, and H. Lilie. 2003. *Embryo research in pluralistic Europe*. Berlin: Springer.

Sosa, E. 2007. *A virtue epistemology I: Apt belief and reflective knowledge*. New York: Oxford University Press.

Sparks, J.R., and S.D. Hunt. 1998. Marketing researcher ethical sensitivity: Conceptualization, measurement, and exploratory investigation. *The Journal of Marketing* 62(2): 92–109.

Sparks, J.R., and Y. Pan. 2010. Ethical judgements in business ethics research: Definition, and research agenda. *Journal of Business Ethics* 91: 405–418.

Spelke, E. 2000. Core knowledge. *American Psychologist* 55: 1233–1243.

Spiro, R.J. 1980. Constructive processes in prose comprehension and recall. In *Theoretical issues in reading comprehension*, ed. R.J. Spiro, B.C. Bruce, and W.F. Brewer, 245–258. Hillsdale: Erlbaum.

Spitzer, M., U. Fischbacher, B. Herrnberger, G. Grön, and E. Fehr. 2007. The neural signature of social norm compliance. *Neuron* 56(1): 185–196.

Sreenivasan, G. 2002. Errors about errors: Virtue theory and trait attribution. *Mind* 111(441): 47–68.

Sripada, C., and S. Stich. 2005. A framework for the psychology of norms. In *The innate mind: Structure and content*, ed. P. Carruthers, S. Laurence, and S. Stich, 280–301. New York: Oxford University Press.

Steger, M.F., T.B. Kashdan, and S. Oishi. 2008. Being good by doing good: Daily eudaimonic activity and well-being. *Journal of Research in Personality* 42: 22–42.

Steinkamp, N.L., B. Gordijn, and H. ten Have. 2008. Debating ethical expertise. *Kennedy Institute of Ethics Journal* 18: 173–192.

Steutel, J., and B. Spiecker. 2004. Cultivating sentimental dispositions through Aristotelian habituation. *Journal of the Philosophy of Education* 38(4): 531–549.

Stevens, J.R. 2010. Donor payoffs and other-regarding preferences in cotton-top tamarins (Saguinus oedipus). *Animal Cognition* 13: 663–670.

Stich, S. 1978. Beliefs and Subdoxastic States. *Philosophy of Science* 45: 499–518.

Stichter, M. 2007a. Ethical expertise: The skill model of virtue. *Ethical Theory and Moral Practice* 10(2): 183–194.

Stichter, M. 2007b. The skill model of virtue. *Philosophy in the Contemporary World* 14(2): 39–49.

Strawson, P.F. 1974/2008. *Freedom and resentment. And other essays*. London: Routledge.

Strong, C. 2010. Theoretical and practical problems with wide reflective equilibrium in bioethics. *Theoretical Medicine and Bioethics* 31: 123–140.

Subiaul, F., J. Vonk, S. Okamoto-Barth, and J. Barth. 2008. Do Chimpanzees learn reputation by observation? Evidence from direct and indirect experience with generous and selfish strangers. *Animal Cognition* 11: 611–623.

Sugarman, J., and D.P. Sulmasy. 2001. *Methods in medical ethics*. Washington, DC: Georgetown University Press.

Sugiyama, L.S., and R. Chacon. 2000. Effects of illness and injury on foraging among the Yora and Shiwiar: Pathology risk as adaptive problem. In *Adaptation and human behavior: An anthropological perspective*, ed. L. Cronk, N. Chagnon, and W. Irons, 371–395. New York: Aldine de Gruyter.

Suhler, C.L., and P. Churchland. 2011. Can Innate, modular "Foundations" explain morality? Challenges for Haidt's Moral Foundations Theory. *Journal of Cognitive Neuroscience* 23(9): 2103–2116.

Sulmasy, D.P., and J. Sugarman. 2001. The many methods of medical ethics (or, thirteen ways of looking at a blackbird). In *Methods in medical ethics*, ed. J. Sugarman and D.P. Sulmasy, 6–10. Washington, DC: Georgetown University Press.

Suter, R.S., and R. Hertwig. 2011. Time and moral judgment. *Cognition* 119(3): 454–458.

Swanton, C. 1991. The role played by the method of wide reflective equilibrium in moral epistemology. *Dialogue* 30: 575–589.

Sytsma, J., and E. Machery. 2010. Two conceptions of subjective experience. *Philosophical Studies* 151(2): 299–327.

Takezawa, M., M. Gummerum, and M. Keller. 2006. A stage for the rational tail of the emotional dog: Roles of moral reasoning in group decision making. *Journal of Economic Psychology* 27(1): 117–139.

Talmi, D., and C. Frith. 2007. Neurobiology: Feeling right about doing right. *Nature* 446(7138): 865–866.

Tangney, J.P., J. Stuewig, and D.J. Mashek. 2007. Moral emotions and moral behavior. *Annual Review of Psychology* 58: 345–372.

Tanner, C. 2008. Zur Rolle von Geschützten Werten bei Entscheidungen. In *Sozialpsychologie und Werte. Beiträge des 23. Hamburger Symposiums zur Methodologie der Sozialpsychologie*, Hrsg. E.H. Witte, 172–188. Lengerich: Pabst.

Tanner, C. 2009. To act or not to act: Nonconsequentialism in environmental decision making. *Ethics and Behavior* 19: 479–495.

Tanner, C., and D.L. Medin. 2004. Protected values: No omission bias and no framing effects. *Psychonomic Bulletin & Review* 11: 185–191.

Tanner, C., D.L. Medin, and R. Iliev. 2008. Influence of deontological vs. consequentialist orientations on act choices and framing effects: When principles are more important than consequences. *European Journal of Social Psychology* 38: 757–769.

Tanner, C., B. Ryf, and M. Hanselmann. 2009. Geschützte Werte Skala (GWS): Konstruktion und Validierung eines Messinstrumentes. *Diagnostica* 55: 174–183.

Tanner, C., A. Brügger, S. van Schie, and C. Lebherz. 2010. Actions speak louder than words. The benefits of ethical behaviors of leaders. *Journal of Psychology* 218(4): 225–233.

Tassy, S., O. Oullier, Y. Duclos, O. Coulon, J. Mancini, C. Deruelle, et al. 2012. Disrupting the right prefrontal cortex alters moral judgement. *Social Cognitive and Affective Neuroscience* 7(3): 282–288.

Taylor, C. 1989. *Sources of the self*. Cambridge, MA: Harvard University Press.

Tetlock, P.E. 1986. A value pluralism model of ideological reasoning. *Journal of Personality and Social Psychology* 50(4): 819–827.

Tetlock, P.E., R.S. Peterson, and J.S. Lerner. 1996. Revising the value pluralism model: Incorporating social content and context postulates. In *The psychology of values: The Ontario symposium*, vol. 8, ed. M.P. Zanna, C. Seligman, and J.M. Olson, 25–51. Hillsdale: Lawrence Erlbaum Associates, Inc.

Tetlock, P.E., O.V. Kristel, S.B. Elson, M. Green, and J.S. Lerner. 2000. The psychology of the unthinkable. Taboo trade-offs, forbidden base rates, and heretical counterfactuals. *Journal of Personality & Social Psychology* 5: 853–870.

Thagard, P. 1998. Ethical coherence. *Philosophical Psychology* 11(4): 405–422.

Thagard, P. 2000. *Coherence in thought and action*. Cambridge, MA: MIT-Press.

Thoma, S.J. 1994. Moral judgment and moral action. In *Moral development in the professions: Psychology and applied ethics*, ed. J. Rest and N. Narvaez, 121–146. Hillsdale: Erlbaum.

Thoma, S., R. Barnett, J. Rest, and D. Narvaez. 1999. What does the DIT measure? *British Journal of Social Psychology* 38(1): 103–111.

Thoma, S.J., W.P. Derryberry, and D. Narvaez. 2009. The distinction between moral judgment development and verbal ability: Some relevant data using socio-political outcome variables. *High Ability Studies* 20(2): 173–185.

Thomas, B.C., K.E. Croft, and D. Tranel. 2011. Harming Kin to Save Strangers: Further evidence for abnormally utilitarian moral judgments after ventromedial prefrontal damage. *Journal of Cognitive Neuroscience* 23(9): 2186–2196.

Thompson, R.A. 1998. Early sociopersonality development. In *Handbook of child psychology, Vol. 3. Social, emotional and personality development*, ed. W. Damon and N. Eisenberg, 25–104. New York: Wiley.

Thomson, J.J. 1993. Morality and bad luck. In *Moral luck*, ed. D. Statman, 195–215. Albany: State University of New York Press.

Thornhill, N.W. 1991. An evolutionary analysis of rules regulating human inbreeding and marriage. *Behavioral and Brain Sciences* 14: 247–293.

Thurston, R.W. 2000. The rise and fall of judicial torture: Why it was used in early modern Europe and the Soviet Union. *Human Rights Review* 1: 26–49.
Tinbergen, N. 1951. *The study of Instinct*. Oxford: Oxford University Press.
Tinbergen, N. 1963. On aims and methods of ethology. *Zeitschrift für Tierpsychologie* 20: 410–433.
Tomasello, M. 2009. *Why we cooperate*. Cambridge, MA: The MIT Press.
Tomasello, M., B. Hare, H. Lehmann, and J. Call. 2007. Reliance on head versus eye in the gaze following of great apes and human infants: The cooperative eye hypothesis. *Journal of Human Evolution* 52: 314–320.
Tomkins, S. 1965. Affect and the psychology of knowledge. In *Affect, cognition, and personality*, ed. S.S. Tomkins and C.E. Izard. New York: Springer.
Tooby, J., and L. Cosmides. 2005. Conceptual foundations of evolutionary psychology. In *The handbook of evolutionary psychology*, ed. D.M. Buss. Hoboken: Wiley.
Treviño, L.K. 2009. Business ethics and the social science. In *A companion to business ethics*, ed. R.E. Frederick, 218–230. Malden: Blackwell.
Treviño, L.K., and M.E. Brown. 2004. Managing to be ethical: Debunking five business ethics myths. *Academy of Management Executive* 18(2): 69–83.
Treviño, L.K., M.E. Brown, and M. Hartman. 2003. A qualitative investigation of perceived executive ethical leadership: Perceptions from inside and outside the executive suite. *Human Relations* 56(1): 5–37.
Treviño, L.K., K. Butterfield, and D. McCabe. 1998. The ethical context in organizations: Influences on employee attitudes and behaviors. *Business Ethics Quarterly* 8(3): 447–476.
Treviño, L.K., L.P. Hartman, and M. Brown. 2000. Moral person and moral manager: How executives develop a reputation for ethical leadership. *California Management Review* 42: 128–142.
Treviño, L.K., G.R. Weaver, and S.J. Reynolds. 2006. Behavioral ethics in organizations: A review. *Journal of Management* 32: 951–990.
Trivers, R.L. 1971. The evolution of reciprocal altruism. *Quarterly Review of Biology* 46: 35–57.
Trivers, R.L. 1985. *Social evolution*. Menlo Park: Benjamin/Cummings.
Trout, J.D. 2009. *The empathy gap*. New York: Viking/Penguin.
Tsukiura, T., and R. Cabeza. 2010. Shared brain activity for aesthetic and moral judgments: Implications for the beauty-is-good stereotype. *Social Cognitive and Affective Neuroscience* 6(1): 138–148.
Tugendhat, E. 1984. Antike und moderne Ethik. In *Probleme der Ethik*, ed. ibid. Stuttgart: Reclam.
Tugendhat, E. 1993. *Vorlesungen über Ethik*. Frankfurt am Main: Suhrkamp.
Turiel, E. 1983. *The development of social knowledge: Morality and convention*. Cambridge: Cambridge University Press.
Urmson, J.O. 1988. *Aristotle's ethics*. Oxford: Blackwell.
Vahid, H. 2004. Varieties of epistemic conservatism. *Synthese* 141: 97–122.
Vaish, A., M. Carpenter, and M. Tomasello. 2009. Sympathy through affective perspective taking and its relation to prosocial behavior in toddlers. *Developmental Psychology* 45: 534–543.
Valdesolo, P., and D.A. DeSteno. 2006. Manipulations of emotional context shape moral decision making. *Psychological Science* 17(6): 476–477.
Van Delden, J.J.M., and G.J.M.W. van Thiel. 1998. Reflective equilibrium as a normative-empirical model in bioethics. In *Reflective equilibrium*, ed. W. van der Burg and T. van Willigenburg, 251–259. Dordrecht: Kluwer Academic Publishers.
Van Delden, J.J.M., L. van der Scheer, G. van Thiel, and G. Widdershoven. 2005. *Ethiek en Empirie. Theorie en methodologie van empirisch ethisch onderzoek*, 106–107. Maastricht: NWO/Caphri.
Van den Daele, W. 2003. *Empirische Befunde zu den gesellschaftlichen Folgen der Pränataldiagnostik. Vorgeburtliche Selektion und die Auswirkungen auf die Lage behinderter Menschen*. Berlin: Berlin-Brandenburgische Akademie der Wissenschaften.
Van den Hoven, M.A. 2006. *A claim for a reasonable morality*. Utrecht: Zeno.

Van der Burg, W. 1997. The importance of ideals. *Journal of Value Inquiry* 31: 23–37.
Van der Burg, W., and T. van Willigenburg (eds.). 1998. *Reflective equilibrium. Essays in Honour of Robert Heeger*. Dordrecht: Kluwer Academic Publishers.
Van der Ploeg, I. 2004. Only Angels can co without skin: On reproductive technology´s hybrids and the politics of body boundaries. *Body & Society* 10: 153–181.
Van der Steen, W.J., and B. Musschenga. 1992. The issue of generality. *Journal of Value Inquiry* 26: 511–524.
Van Willigenburg, T. 1991. *Inside the ethical expert*. Kampen: Kok Pharos.
Van Luijk, H.E.M., and W. Dubbink. 2011. Moral competence. In *European business ethics cases in context*, Issues in business ethics, vol. 28, ed. W.D. van Liederkerke and H. van Luijk, 11–17. Dordrecht/New York: Springer.
Van Luijn, H.E.M., A.W. Musschenga, R.B. Keus, W.M. Robinson, and N.K. Aaronson. 2002. Assessment of the risk/benefit ratio of phase II cancer clinical trials by institutional review boards (IRB) members. *Annals of Oncology* 13: 1307–1313.
Van Schaik, C.P. 2000. Infanticide by male primates: The sexual selection hypothesis revisited. In *Infanticide by males and its implications*, ed. C.P. van Schaik and C.H. Janson, 27–60. Cambridge: Cambridge University Press.
Van Schaik, C.P., and C.H. Janson. 2000. *Infanticide by males and its implications*. Cambridge: Cambridge University Press.
Van Thiel, G.J.M.W., and J.J.M. van Delden. 2010. Reflective equilibrium as a normative empirical model. *Ethical Perspectives* 17(2): 183–202.
Van Wolkenten, M., S.F. Brosnan, and F.B.M. de Waal. 2007. Inequity responses in monkeys modified by effort. *Proceedings of the National Academy of Sciences* 104(47): 18854–18859.
Vartanian, O., and V. Goel. 2004. Neuroanatomical correlates of aesthetic preference for paintings. *Neuroreport* 15: 893–897.
Vasquez, K., D. Keltner, D.H. Ebenbach, and T.L. Banaszynski. 2001. Cultural variation and similarity in moral rhetorics: Voices from the Philippines and United States. *Journal of Cross-Cultural Psychology* 32: 93–120.
Verplanken, B., and R.W. Holland. 2002. Motivated decision making: Effects of activation and self-centrality of values on choices and behavior. *Journal of Personality and Social Psychology* 82(3): 434–447.
Victor, B., and J.B. Cullen. 1988. The organizational bases of ethical work climates. *Administrative Science Quarterly* 33: 101–125.
Von Uexküll, and W. Wesiack. 1998. *Theorie der Humanmedizin*, 3rd ed. München/Wien/Baltimore: Urban & Schwarzenberg.
Von Uexküll, R. Adler, J.M. Herrmann, K. Köhle, O.W. Schonecke, and W. Wesiack (eds.). 1996. *Psychosomatic medicine*. München/Wien/Baltimore: Urban & Schwarzenberg.
Vonk, J., S.F. Brosnan, J.B. Silk, J. Henrich, A.S. Richardson, S. Lambeth, S.J. Schapiro, and D.J. Povinelli. 2008. Chimpanzees do not take advantage of very low cost opportunities to deliver food to unrelated group members. *Animal Behaviour* 75: 1757–1770.
Wagner, U., K. N'Diaye, T. Ethofer, and P. Vuilleumier. 2011. Guilt-specific processing in the prefrontal cortex. *Cerebral Cortex* 21: 2461–2470.
Wahaj, S.A., K.R. Guse, and K.E. Holekamp. 2001. Reconciliation in spotted hyenas (Crocuta crocuta). *Ethology* 107: 1057–1074.
Walker, L.J., and J. Frimer. 2009. Moral personality exemplified. In *Personality, identity and character: Explorations in moral psychology*, ed. D. Narvaez and D.K. Lapsley, 232–255. New York: Cambridge University Press.
Walster, E. 1966. Assignment of responsibility for an accident. *Journal of Personality and Social Psychology* 3: 73–79.
Walster [Hatfield], E., G.W. Walster, and E. Berscheid. 1978. *Equity: Theory and research*. Boston: Allyn and Bacon.
Warneken, F., and M. Tomasello. 2006. Altruistic helping in human infants and young Chimpanzees. *Science* 311: 1301–1303.

Warneken, F., B. Hare, A.P. Melis, D. Hanus, and M. Tomasello. 2007. Spontaneous altruism by Chimpanzees and young children. *PLoS Biology* 5(7): e184.
Watanabe, K. 2001. A review of 50 years of research on the Japanese monkeys of Koshima: Status and dominance. In *Primate origins of human cognition and behavior*, ed. T. Matsuzawa, 405–417. Tokyo: Springer.
Waterman, A.S. 1988. On the uses of psychological theory and research in the process of ethical inquiry. *Psychological Bulletin* 103(3): 283–298.
Watts, D.P. 2002. Reciprocity and interchange in the social relationships of wild male Chimpanzees. *Behaviour* 139: 343–370.
Wear, S. 2005. Ethical expertise in the clinical setting. In *Ethics expertise: History, contemporary perspectives, and applications*, ed. L.M. Rasmussen, 243–258. Dordrecht: Springer.
Weaver, G.R.M., and L.K. Treviño. 1994. Normative and empirical business ethics: Separation, marriage of convenience, or marriage of necessity? *Business Ethics Quarterly* 4(29): 129–143.
Weber, M. 1918. *Science as a vocation.* Online: http://www.wisdom.weizmann.ac.il/~oded/X/WeberScienceVocation.pdf. Accessed 20 Sept 2008.
Weingart, P. 2002. Verlust der Distanz – Verlust des Vertrauens? Kommunikation gesicherten Wissens unter Bedingungen der Medialisierung. In *Ideale Akademie*, ed. W. Vosskamp, 95–112. Berlin: Akademie Verlag.
Weinstein, B. 1994. The possibility of ethical expertise. *Theoretical Medicine* 15: 61–75.
Wendell, S. 1996. *The rejected body. Feminist philosophical reflections on disability*. New York: Routledge.
Werner, D. 1979. A cross-cultural perspective on theory and research on male homosexuality. *Journal of Homosexuality* 4: 345–362.
Wertz, D.C., and J.C. Fletcher. 2004. *Genetics and ethics in global perspective*. Dordrecht: Kluwer Academic Publishers.
Weston, A. 2008. *A 21st century ethical toolbox*, 2nd ed. Oxford: Oxford University Press.
Wheatley, T., and J. Haidt. 2005. Hypnotically induced disgust makes moral judgments more severe. *Psychological Science* 16: 780–784.
White, D.R., and M.L. Burton. 1988. Causes of polygyny: Ecology, economy, kinship, and warfare. *American Anthropologist* 90: 871–887.
Wiesemann, C. 2006. *Von der Verantwortung ein Kind zu bekommen. Eine Ethik der Elternschaft*. München: Beck.
Wilkinson, G.S. 1984. Reciprocal food sharing in the vampire bat. *Nature* 308: 181–184.
Wilkinson, R., and K. Pickett. 2010. *The spirit level: Why equality is better for everyone*. London: Penguin Books.
Williams, B. 1985. *Ethics and the limits of philosophy*. London: Fontana.
Williams, G.C. 1989. A sociobiological expansion of evolution and ethics. In *Evolution and ethics, with new essays on its Victorian and sociobiological context*, ed. J. Paradis, 179–214. Princeton: Princeton University Press.
Williams, B. 1995. Moral Luck: A postscript. In *Making sense of humanity*, ed. B. Williams, 241–247. Cambridge: Cambridge University Press.
Williamson, T. 2007. *The philosophy of philosophy*. Oxford: Blackwell.
Winch, P. 1987. Who is my neighbour? In *Trying to make sense*, ed. ibid, 154–166. Oxford: Blackwell.
Witte, E.H. 1994. *Lehrbuch Sozialpsychologie* [Social psychology: A textbook]. Weinheim: Beltz-PVU.
Witte, E.H., and J. Doll. 1995. Soziale Kognition und empirische Ethikforschung: Zur Rechtfertigung von Handlungen [Social cognition and empirical research on ethics: On justification of behaviour]. In *Soziale Kognition und empirische Ethikforschung*, ed. E.H. Witte, 97–115. Lengerich: Pabst.
Witte, E.H., and I. Heitkamp. 2005. Empirical research on ethics: The influence of social roles on decisions and on their ethical justification [Hamburger Forschungsbericht zur Sozialpsychologie Nr. 61]. Hamburg: Universität Hamburg, Arbeitsbereich Sozialpsychologie.

Repr. in: G.N. Galanis (ed.). *Eleftherna. scientific yearbook,* 55–83. University of Crete: Department of Psychology.

Witte, E.H., and C. Mölders. 2007. Einkommensteuergesetz: Begründung der vorhandenen Ausnahmetatbestände ethisch bedenklich [Income tax law: Justification of exception rules are questionable]. *Wirtschaftspsychologie* 2: 65–81.

Witte, E.H., G. Aßmann, and S. Lecher. 1995. Ethik-Kodizes aus Psychologie und Soziologie und ihre Verbindung zu ethischen Grundpositionen [Codices of ethics from psychology and sociology and their relation to the classical positions of ethics]. In *Soziale Kognition und empirische Ethikforschung* [Social cognition and empirical research on ethics], ed. E.H. Witte, 116–120. Lengerich: Pabst.

Wobber, V., R. Wrangham, and B. Hare. 2010. Bonobos exhibit delayed development of social behavior and cognition relative to Chimpanzees. *Current Biology* 20(3): 226–230.

Wood, A. 2011. Humanity as an end in itself. In *Derek Parfit: On What Matters*, vol. 2. Oxford: Oxford University Press.

Woods, M. 1999. A nursing ethic: The moral voice of experienced nurses. *Nursing Ethics* 6: 423–432.

Woodward, J., and J. Allman. 2007. Moral intuition: Its neural substrates and normative significance. *Journal of Physiology–Paris* 101: 179–202.

Wrangham, R.W. 1999. Evolution of coalitionary killing. *Yearbook of Physical Anthropology* 42: 1–30.

Wrangham, R.W. 2004. Killer species. *Daedalus* 133: 25–35.

Wrangham, R.W., M.L. Wilson, and M.N. Muller. 2006. Comparative rates of violence in Chimpanzees and humans. *Primates* 47: 14–26.

Yamamoto, S., and M. Tanaka. 2009a. Do Chimpanzees (Pan troglodytes) spontaneously take turns in a reciprocal cooperation task? *Journal of Comparative Psychology* 123(3): 242–249.

Yamamoto, S., and M. Tanaka. 2009b. How did altruism and reciprocity evolve in humans? *Journal of Interaction Studies* 10(2): 150–182.

Young, L., and J. Dungan. 2011. Where in the brain is morality? Everywhere and maybe nowhere. *Social Neuroscience* 7: 1–10.

Young, L., and M. Koenigs. 2007. Investigating emotion in moral cognition: A review of evidence from functional neuroimaging and neuropsychology. *British Medical Bulletin* 84: 69–79.

Young, L., and R. Saxe. 2008. The neural basis of belief encoding and integration in moral judgment. *NeuroImage* 40(4): 1912–1920.

Young, L., and R. Saxe. 2009. An FMRI investigation of spontaneous mental state inference for moral judgment. *Journal of Cognitive Neuroscience* 21(7): 1396–1405.

Young, L., F. Cushman, M. Hauser, and R. Saxe. 2007. The neural basis of the interaction between theory of mind and moral judgment. *Proceedings of the National Academy of Sciences of the United States of America* 104(20): 8235–8240.

Young, L., J.A. Camprodon, M. Hauser, A. Pascual-Leone, and R. Saxe. 2010a. Disruption of the right temporoparietal junction with transcranial magnetic stimulation reduces the role of beliefs in moral judgments. *Proceedings of the National Academy of Sciences of the United States of America* 107(15): 6753–6758.

Young, L., A. Bechara, D. Tranel, H. Damasio, M. Hauser, and A. Damasio. 2010b. Damage to ventromedial prefrontal cortex impairs judgment of harmful intent. *Neuron* 65(6): 845–851.

Young, L., S. Nichols, and R. Saxe. 2010c. Investigating the neural and cognitive basis of moral luck: It's not what you do but what you know. In *European review of philosophy*, Special issue on psychology and experimental philosophy 1, ed. J. Knobe, T. Lombrozo, and E. Machery, 333–349.

Young, L., J. Scholz, and R. Saxe. 2011. Neural evidence for "intuitive prosecution": The use of mental state information for negative moral verdicts. *Social Neuroscience* 6(3): 302–315.

Zagzebski, L. 2006. The admirable life and the desirable life. In *Values and virtues: Aristotelianism in contemporary ethics*, ed. T.D.J. Chappell, 53–66. Oxford: Oxford University Press.

Authors

Mark Alfano (philosophy) is at the Center of Human Values of Princeton University, Princeton, USA.

Sarah F. Brosnan (primatology) is at the Departments of Psychology and Philosophy, Georgia State University, Georgia, USA.

Judith M. Burkart (anthropology) is at the Anthropological Institute and Museum of the University of Zurich, Switzerland.

Markus Christen (neuroethics) is at the Institute of Biomedical Ethics, University of Zurich, Zurich, Switzerland.

Johannes Fischer (theological ethics, emeritus) was at the Institute of Social Ethics of the University of Zurich, Switzerland.

Tobias Gollan (social psychology) is at the Research Institute for Quality in Child and Youth Support at the foundation 'Die Gute Hand', Kürten, Germany.

Hauke R. Heekeren (affective neuroscience) is at the Cluster of Excellence "Languages of Emotion" of the Freie Universität Berlin, Germany.

Markus Huppenbauer (ethics) is at the University Priority Program Ethics, University of Zurich, Zurich, Switzerland.

Adrian V. Jaeggi (anthropology) is at the Department of Anthropology of the University of California Santa Barbara, California, USA.

Antti Kauppinen (philosophy) is at the Department of Philosophy of the Trinity College Dublin, Ireland.

Tanja Krones (clinical ethics and sociology) is at the University Hospital Zurich and the Institute of Biomedical Ethics of the University of Zurich, Switzerland.

Daniel Lapsley (moral psychology and adolescent development) is at the Psychology Department of the University of Notre Dame, Indiana, USA.

M. Christen et al. (eds.), *Empirically Informed Ethics: Morality between Facts and Norms*, Library of Ethics and Applied Philosophy 32, DOI 10.1007/978-3-319-01369-5,

Theresa Lopez (moral philosophy) is at the Department of Philosophy of the Hamilton College, New York, USA.

Bert Musschenga (moral philosophy) is at the Philosophy Department of the VU University, Amsterdam, Netherlands.

Darcia Narvaez (moral psychology) is at the Department of Psychology, University of Notre Dame, Indiana, USA.

Adriano Naves de Brito (moral philosophy) is at the School of Humanities, Philosophy, Unisinos University, Brazil.

Shaun Nichols (experimental philosophy) is at the Department of Philosophy, University of Arizona, Arizona, USA.

Kristin Prehn (cognitive neuroscience) is at the Cluster of Excellence "Languages of Emotion", Freie Universität Berlin, Germany.

Jesse J. Prinz (moral philosophy) is at the Philosophy Program at the Graduate Center, City University of New York, New York, USA.

Claudia Rudolf von Rohr (anthropology) is at the Anthropological Institute & Museum of the University of Zurich, Switzerland.

Carmen Tanner (social psychology) is at the Department of Banking and Finance, Center for Responsibility in Finance, University of Zurich, Zurich, Switzerland.

Mark Timmons (moral philosophy) is at the Department of Philosophy, University of Arizona, Arizona, USA.

Johannes J.M. van Delden (medical ethics) is at the Julius Center for Health Sciences and Primary Care of the University Medical Center Utrecht, The Netherlands.

Carel van Schaik (anthropology) is at the Anthropological Institute & Museum of the University of Zurich, Switzerland.

Ghislaine J.M.W. van Thiel (medical ethics) is at the Julius Center for Health Sciences and Primary Care of the University Medical Center Utrecht, The Netherlands.

Erich H. Witte (social psychology) is at the Psychology Department of the University of Hamburg, Germany.

Index

M. Christen et al. (eds.), *Empirically Informed Ethics: Morality between Facts and Norms*, Library of Ethics and Applied Philosophy 32, DOI 10.1007/978-3-319-01369-5,