**A Research Program for Empirically Informed Ethics**

*Markus Christen & Mark Alfano*

**1. Introduction**

“What is the right thing to do?” This question[[1]](#footnote--1) echoes through the centuries and millennia of human history. It refers 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 align with, and the states of affairs people should aim at. It is therefore unsurprising that human behavior in moral contexts has become a topic of empirical research, although it was to some respect deliberately excluded as a legitimate research topic in the advent of modern science.[[2]](#footnote-0) This, however, has changed. Notably, in the last two decades there has been a substantial increase in empirical research on morality—in particular using psychological and neuroscientific methods—which is confirmed by bibliometric studies (Christen 2010). This research also influences moral philosophy; as a matter of fact, empirical research on morality 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 the increasing amount of empirical data on the evolutionary “origin”, the biological “foundation”, the psychological suggestibility, and cultural diversity of human morality have been ambivalent. One strand of argument 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 acknowledges empirical insighst for theory building within ethics (Edel 1961)—but with contradictory 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 started using empirical methods themselves in order to inform their normative thinking (for an overview see Appiah 2008; Knobe & Nichols 2008).

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 on 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 section 2.4.).

Thus, the endeavor of promoting an *empirically informed* ethics raises various questions. This chapter intends to structure them with respect to the subject-matter (i.e. morality), 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 should also outline what is on stake when empirical insights are taken seriously by normative theorists—a point that may affect a genuine 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 endeavor 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 in a distinct 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, p. 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 2 presents the phenomenon empirically informed ethics tackles, which is, we propose, an exhaustive 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. In section 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 section 4, we provide an overview of the different kinds data that can inform ethics in various ways. The other eleven contributions of this book will be assigned to the paragraphs of this section. Next, in section 5, we discuss and respond to some of the more trenchant critics of empirically informed ethics. Finally, in section 6, we present several problems that we consider to be of particular importance for an empirical approach to ethics.

**2. The phenomenon under investigation**

*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[[3]](#footnote-1) thought, feeling, deliberation, and behavior[[4]](#footnote-2) 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[[5]](#footnote-3) of actions, judgments, negotiations, and other kinds of expressions social entities are able to perform when they interact. 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 would not consider to be of a moral kind (Haidt & Graham 2007).

This straightforward observation already leads to a difficult question: whether there are uncontroversial criteria in order 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 universability 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 Carson & Moser 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 valuable 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. 2008), but are then confronted with the large variability of individual emotional excitability or “affective 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 (Joice 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 either as being ‘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.[[6]](#footnote-4)

This classification problem is complicated by a further aspect: 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 the historical development) with respect to what is called ‘immoral’ (see the contribution of Prinz 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 morality.[[7]](#footnote-5) 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.

*2.2. Moral agency*

Having set a first, basic frame of the problem, we will now outline the subject matter 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 5 and 12 and the following explanations).[[8]](#footnote-6)

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 certainly requires further specification depending on the concrete issue under investigation, as well as empirical data of various kinds (see section 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 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 section 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[[9]](#footnote-7) is 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)[[10]](#footnote-8). 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 (because, for instance, there are different time scales involved). Therefore, some moral behaviors may be collective 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).

The structure of moral agency as we construe it here is threefold. First, moral agency requires a specified *set of abilities* 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). Abilities, 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.

Agent

Abilities

Normative

Frame

Context

**Figure 1:** The structural components of moral agency.

First, specifying the set of abilities certainly is closely related to the defining the moral agent—and the spectrum of proposals for necessary and/or sufficient abilities 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 actors, such that the behavior of the agent is not completely controlled by factors outside of the agent. 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.[[11]](#footnote-9) All these terms employ 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 which includes the implication that this comparison has some effect for the agent and its (future) behavior. Again, the definition 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. The definitions of these elements such that the normative frame can be called a properly *moral* frame are 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 one of the central questions to which empirical research can contribute (see section 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 can be an open question whether a boundary condition is understood as inherent necessity or just as practical constraint that can be overcome. It’s also not implausible to expect that some aspect of the normative frame may be attributed to the context—to what degree, is again an issue of definition.

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 incimpleteness, 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 abilities 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 component 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.

*2.3. Spatial and temporal scales*

Moral agency develops on various spatial and temporal scales, which allows us 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 upon a specific issue), a group of directly interacting agents over a longer timescale (allowing, for instance, the relationships of mutual trust and dependency), and a collective of agents who interact in an 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).[[12]](#footnote-10) Again, many real-world problems involve an entanglement of several time-scales, for instance, with respect to actions-consequences relations on a collective spatial scale.[[13]](#footnote-11) However, this classification also allows to pigeonhole 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).

|  |  |
| --- | --- |
|  | **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 upon a moral decision | Collective decision making in medical ethics | Institutionalized processes of praise 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 |

**Table 1:** Examples of behavioral patterns of moral agency, classified along spatial and temporal scales.

For each of these behavioral patterns, abilities, normative frame and context will have to be specified, if they are to become an object of systematic investigation. This will be outlined in more detail in sections 3 and 4.

*2.4. What does ‘being informed’ mean?*

Our discussion so far has focused on the phenomena of moral agency that happen to be 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.[[14]](#footnote-12) To do we, we first outline in more detail the difficulties that arise when attempting to draw clear distinctions 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 considers 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 the field is dealing with. However, there are two (interrelated) analytical distinctions that are considered central the task of ethics: the is-ought-dichotomy[[15]](#footnote-13) (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” (but see footnote 18) is often devastating in any practical discussion of ethics (an example is Arn 2009). We are, however, skeptical about both to 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 to relate empirical data with 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 been argued that many seemingly value-based conflicts 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 high, which complicates the task of identifying hidden normative assumptions between some facts. This problem may be overcome by allowing sufficient time for deliberation,[[16]](#footnote-14) but this solution may fail when taking the 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; 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) 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 have been analyzed (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). The data are therefore not independent of the investigators' normative frame, but involved in a complex feedback loop with it. Third, philosophical interpreters of scientific results are often unaware of raging controversies within the scientific discipline over the validity of those results.[[17]](#footnote-15) Such 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 & Nichols 2008; see also the contribution of Shaun Nichols and colleagues 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]](#footnote-16) 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 equilibirium; Rawls 1971/1999). We have to expect various similarity relations between such entities (Thagard 1998), and they will rely on intuitions, too (Christen & Ott, submitted). 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 judgements. He comes to the conclusion that moral reflection from this orientation cannot do justice to moral phenomena. Furthermore, this view can neither 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 agentive 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 constituent 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 offers in that way various results of case studies, an empirically informed ethics could provide.

These considerations and contributions demonstrate that an empirically informed ethics has a different relationship to facts and data compared to ethical theorizing that basically uses data as an exploitable resource. It involves both a 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 section explore this point in more detail, taking the other contributions of the book emerging from various disciplines as exemplars.

**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 three of the more important methodological distinctions relevant to empirically informed ethics.

*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.

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, learning, and markets).

Quantitative research methods offer many benefits. Their results can be analyzed statistically, replicated across time and researcher 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, 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.

*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 XXX). 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, or 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. 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, McGhee, & Schwartz (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 are always easy. What’s tested is 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 & 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 economics Nina Mazar, On Amir, and Dan Ariely (2008) did so, for instance, by providing people opportunities to cheat.

These methods can and in many cases probably 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.

*3.3. Individual / social*

One final 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’s 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 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 answered by a smorgasbord of methodologies.

**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 many other disciplines in the humanities and social sciences.[[19]](#footnote-17) The same holds also for some disciplines within the natural sciences (in particular: neuroscience, anthropology, primatology). Even in medicine, “deviations” in moral behavior have increasingly become an issue (Christen & Regard, in press). 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.

*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 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 & Richerson 1985, Caplan 1978, 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 abilities the agent should have. 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 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 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 section 2.2., a crucial issue for any phylogenetic explanation of moral agency is the identification of the abilities 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 abilities required for them. The contribution of Sara Brosnan is an exemplar 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 settings) of primates living both in a natural or an “artificial” (zoo etc.) environment, 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 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 by other social species.

With respect to the relevant context for the phylogeny of moral agency, there is little disagreement among the researchers in that field: the specific environmental conditions and the lifestyle of human foragers—i.e., the spatial scale of the (small) group with strong mutual interdependenties and relations—shaped (human) moral agency in a decisive way. In the contribution of Carel P. van Schaik, Judith Burkart, Adrian Jaeggi and Claudia Rudolf von Rohr, 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 in respect of 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. 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, political science, among others.

*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 evident with respect 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, which abilities a moral agent should develop in the course of the first few years (or perhaps even decades) of life. Carmen Tanner and Markus Christen investigate this issue in their contribution by presenting a broad overview of the various competences that have been proposed in the literature (in particular, moral psychology) as essential for moral agency. In their contribution these abilities are re-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 competences 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 through the competence of moral commitment. The framework of “moral intelligence” stresses the importance of making such competences 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 with respect to the ontogeny of moral agency refers to 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 Dan Lapsley outlines this point by presenting both empirical data on early childhood development and a theoretical framework called “triune ethics,” which distinguishes between three basic ethical systems with respect to their functional effect that 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 abilities 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. An question of particular importance is whether this research will be able to distinguish between domain-specific and general capacities needed for moral agency, and how individual differences reflected by the variance of this neuronal infrastrucure 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 empirical research helps to disentangle the different 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 whether one can speak of moral expertise in a similar way. Bert Musschenga 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, who address “ethical leadership”: the abilities and values that should be fostered by people who occupy positions of responsibility in companies and other institutions. In their contribution, they sketch various areas of intersections where ethics and psychology can learn and benefit from each other when exploring ethical leadership and ethical decision-making. They thus show how empirically informed ethics can become very practical indeed.

*4.3. Reasons and moral agency*

A third group of relevant questions within ethics, for which empirical data is of importance, refers to the role of reasons and their founding in intuitions and standards of rationality —an aspect that requires a sophisticated understanding of morality with respect to basic abilities 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 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 understand 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. 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 give harsher blame judgments to unlucky agents than to 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 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 empirically test hypotheses, 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, the negative expression of particularistic utilitarianism, 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.[[20]](#footnote-18)

**5. Criticisms of empirically informed ethics**

XXX

**6. Some focus questions for empirically informed ethics**

The contributions of this volume do not of course paint a complete picture of the empirical investigation of normative issues. They serve instead as exemplars of the many ways in which empirical information can be brought to bear on ethical questions. In the following, we briefly suggest how some of these questions can be aggregated to more general questions empirically informed ethics 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 beliefs, whereas the kind of interrelations allow for discerning the topology of the space and—using a clustering approach (Christen & Ott, submitted)—the classes present. In the literature, various classifications have been proposed (Autonomy, Community, Divinity: Shweder et al. 1997; harm, fairness, ingroup, authority, purity: 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 discernation of the topology of the moral space 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 dynamics, whereas “mavericks” in the ethical discourse (e.g., honor; Appiah 2010) may be of more importance. In this context, also the effect of deliberations in advancing the morality of societies may be an important, but often neglected aspect of empirical research. This refers, for example, to the inflated appeal to “responsibility” towards function owners within social, political or economic systems that may conflict with the owner’s actual intervention possibilities to control the system – in particular, if these systems emerge from self-organized dynamics of interacting agents. This could lead to responsibility gaps, i.e. a divide between the appeals of responsibility on the one hand and the control-possibilities founded in explicit responsibility-relations on the other hand. In other words: an explicite advocacy for moral change towards a “better world” may have effects that actually undermine this goal, making the understanding of moral change an important topic of investigation.
* *Dealing with moral complexity:* A further interesting phenomenon refers to the complexity-simplicity relation with respect of 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 & Medin 2004); it makes a seemingly complex problem easier to decide. In short we may say: ethics makes things more complex, morals makes 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 deeper 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 they grasp 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]](#footnote-19)—could serve as a guideline for future research.

**New Citations:**

Gilligan, C. (1982). *In a Different Voice*. Cambridge, MA: Harvard University Press.

Greenwald, A., McGhee, D., & Schwartz, J. (1998). Measuring individual differences in implicit cognition: The Implicit Association Test. *Journal of Personality and Social Psychology*, 74, 1464-80.

Mazar, N., Amir, O., & Ariely, D. (2008). The dishonesty of honest people: A theory of self-concept maintenance. *Journal of Marketing Research*, 45:6, 633-44.

Parkinson, C., Sinnott-Armstrong, W., Koralus, P., Mendelovici, A., McGeer, V., & Wheatley, T. (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.

Perugini, M. & Leone, L. (2009). Implicit self-concept and moral action. Journal of Research in

Personality, 43, 747-54.

1. Allowedly, the human concern with morality is not represented by a single question. Actually, the focus on moral decision making and moral action, for which this question stands, is typical for a very recent understanding of ethics understood 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) as an early critique on this tendency to narrow the focus of ethics. Recently, however, an increase in interest in virtue ethics can be observed. [↑](#footnote-ref--1)
2. A well-known evidence for this point is the draft of the credo of the Royal Society written by Robert Hooke in 1663, where he described the role of the Royal 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 (explicitely and implicitely) in various parts of the charter (quoted after Weingart 2002: 96). [↑](#footnote-ref-0)
3. I use the term ‘guide’ not in the sense that the agent necessarily requires conscious avareness of this frame. There may be various physical implementations of the normative frame some of which are directly involved in the mechanisms that produce behaviors. [↑](#footnote-ref-1)
4. I use the term ‘behaviors’ to denote any observable expression of interacting social entities that includes communicative expressions of both verbal and 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 constrainted 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—i.a. because the issue of responsibility attribution is more clear in that case. However, I will not follow this tradition (see section 2.4.). [↑](#footnote-ref-2)
5. Again, the term ‘social’ is used 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 unvestigation, 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 in this volume). [↑](#footnote-ref-3)
6. This ambiguity probably results from the basic fact, that normativity is interwoven in the very expression of life: all life-forms have built-in “desired states” or “goal states” with respect to basic needs and threaths, sensors to detect them and actors to seek or avoid them (including plants, fungi, protozoa, and bacteria). Although there is surely a consensus that most goal-seeking behaviors of life-forms are non-moral, this certitude decreases when social life forms come into the focus of investigation. And although we have good reason to couple (sophisticated) morality to language and the ability for conscious reasoning, this criterion may be of little use when the question is focused, how something like morality evolved. [↑](#footnote-ref-4)
7. See Haidt & Kesebir (2010) as an example of how the classifier shaped the types of research questions with respect to morality within social psychology. [↑](#footnote-ref-5)
8. One may object that a different term should be used instead, e.g. ‘moral behaver’ instead of ‘moral agent’ and the coinage ‘moral behavory’ instead of ‘moral agency’. However, ‘behavior’ is too broad in that respect, as only a subclass of behaviors is relevant for moral agency, as social impact of the behavior (pattern) is required and some kind of feedback mechanism back 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 (pattern). [↑](#footnote-ref-6)
9. Here, I further have to note that the concepts ‘agent’ and ‘action’ should be distinguished as well in the sense that not all moral behaviors agents may produce are actions. As indicated earlier (footnote 5), the term’action’ refers to rather constraint behaviors operating on a rather 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 unter 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’). [↑](#footnote-ref-7)
10. There is a longstanding discussion about collective responsibility that I will not outline here (see e.g. Held 1970; Lewis 1948; Narveson 2002). [↑](#footnote-ref-8)
11. For an overview see Bandura (2001), Christen (2009), or the entry on “action” in the Stanford Encyclopedia of Philosophy (http://plato.stanford.edu/entries/action; accessed on October 31st 2011). [↑](#footnote-ref-9)
12. The issue of the relevant timescale of evolutionaly processes surely depends on the type of phenomenon one analyses. 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). [↑](#footnote-ref-10)
13. One issue is the possibility of a 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. [↑](#footnote-ref-11)
14. In the following, I will not go into much detail with respect to meta-ethical issues that focus on the logical, semantical and pragmatic structures of ethical argumentation. However, one has to take into acocunt, that also within moral philosophy the distinction between ethics and meta-ethics is not easy to draw (e.g. Düwell et al. 2002: 3) and some of the issues discussed further in this classfications may also be called meta-ethical. [↑](#footnote-ref-12)
15. This dichotomy goes back to David Hume (1739-40) and should be distinguished from the natural fallacy problem that has been discussed by George E. Moore (1903). See also the contribution of Krones in this volume. [↑](#footnote-ref-13)
16. To what extend such clarifications are practicable is, however, disputable. Some “hidden” values probably are not “hidden” in the sense of the word, but reflect the fact that there are widely shared by one dialoge group, i.e. are self-evident and don’t need to be made explicit (from the groups’ perspective). Making them explicit may then be something that widens the gap between disputants—and not a first step to close it. [↑](#footnote-ref-14)
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. [↑](#footnote-ref-15)
18. See Krist Vaesen & Martin Peterson, The Reliability of Armchair Intuitions (unpublished manuscript). I thank Mark Alfano for indicating this paper to me. [↑](#footnote-ref-16)
19. Examples can, e.g., be found in behavioral economics (Gintis et al. 2005), political science (Haidt & Graham 2007), and pedagogy (Huff & Frey 2005). [↑](#footnote-ref-17)
20. I add that one recent methodology that is used to inform ethical theorizing as well, but is not presented in this book: computer simulation methods. It is a well-established phenomenon that various scientific disciplines currently experience a tremendous transition through the use of computer simulations (e.g., solid-state physics, chemistry, molecular biology, and climate physics)—and various disciplines in humanities and social sciences will surely be transformed by this methodology in the near future as well (an example is the FuturICT project, a mega-science-proposal currently under investigation in the EU 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). See Danielson (1992) as an early example of using computer simulation methods in ethics. [↑](#footnote-ref-18)
21. On August 8th 1900, the German mathematitian David Hilbert presented a list of 23 unsolved mathematical problems at the International Congress of Mathematics in Paris that turned ot to become a programmatic outlook influencing the research within the field for many years. [↑](#footnote-ref-19)