Rationality and Morality

The Handbook of Rationality

Edited by Markus Knauff and Wolfgang Spohn

The MIT Press
Cambridge, Massachusetts
London, England

The MIT Press would like to thank the anonymous peer reviewers who provided comments on drafts of this book. The generous work of academic experts is essential for establishing the authority and quality of our publications. We acknowledge with gratitude the contributions of these otherwise uncredited readers.

This book was set in Stone Serif and Stone Sans by Westchester Publishing Services. Printed and bound in the United States of America.

Library of Congress Cataloging-in-Publication Data

Names: Knauff, Markus, editor. | Spohn, Wolfgang, editor.
Title: The handbook of rationality / edited by Markus Knauff and Wolfgang Spohn.
Description: Cambridge : The MIT Press, 2021. | Includes bibliographical references and index.
Identifiers: LCCN 2020048455 | ISBN 9780262045070 (hardcover)
Subjects: LCSH: Reasoning (Psychology) | Reason. | Cognitive psychology. | Logic. | Philosophy of mind.
Classification: LCC BF442 .H36 2021 | DDC 153.4/3—dc23
LC record available at https://lccn.loc.gov/2020048455

10 9 8 7 6 5 4 3 2 1

Contents

Preface

This book is for students and scholars who want to think about the big questions: Are we rational? What at all does it mean to be rational? Why do we sometimes deviate from the norms of rationality? What happens in our brain when we reason and decide? What is a good or bad decision? How do we come to sensible beliefs and knowledge? These are questions that people have wondered about since the beginning of humankind. And these are questions about which, since ancient times, thinkers have had much to say. The aim of *The Handbook of Rationality* is to show how (analytic) philosophers and (cognitive) psychologists think about these big questions. In the past decades, both disciplines have made tremendous advances in better understanding the very nature of human rationality. Yet, the interaction between the fields is developed only in rudimentary forms. The consequence is that no publications exist that provide students and scholars with an easily accessible integrated overview about the state of the art in the psychology and philosophy of rationality. *The Handbook of Rationality* seeks to fill this gap and to be a novel resource for students and scholars in psychology and philosophy, but also in neighboring areas, such as economics, neuroscience, artificial intelligence, linguistics, law, sociology, anthropology, or education.

This handbook has not become so voluminous because we wanted it to be. A little less work with the publication of this volume would have been all right. But a less complete treatment would not have done justice to the importance of the subject. Of course we were fully aware of the complexity and breadth of the topic of rationality when we started to think about the chapters and authors that we wanted to collect in this handbook. However, during our work, we repeatedly encountered new topics and issues that we had not considered at the beginning because they only arose from our interdisciplinary cooperation. If we had merely concentrated on our respective home disciplines, we would have missed many of these interesting questions lying in the no man's land between the disciplines.

What has greatly facilitated our interdisciplinary collaboration and made it so productive is the fact that we already collaborated for many years within the Priority Program SPP 1516, *New Frameworks of Rationality*, which has been funded by the German Research Foundation (Deutsche Forschungsgemeinschaft [DFG]) from 2011 to 2018. The program consisted of 15 research projects from psychology, philosophy, and artificial intelligence and often combined methods from these disciplines within highly interdisciplinary projects. The program was headed by Markus Knauff and coinitiated by Wolfgang Spohn, the cognitive psychologists Ralph Hertwig and Michael Waldmann, and the philosopher Gerhard Schurz. Later, the computer scientist Gabriele Kern-Isberner joined this governing board. In this priority program, we gathered many times, worked together very intensively, worked on many joint publications, and spent many, many hours of intensive discussion (www.spp1516.de). All this resulted in enormous shared benefit and progress in mutual understanding.

At some point, we thought we should share our progress with a broader scientific public, and so the idea for this handbook was born in the summer of 2015. We are very grateful to the DFG for the generous funding of the interdisciplinary research program. We are aware that Germany is one of the few countries that still invest so much in basic research without expecting results that are directly economically exploitable. This is a privilege we greatly appreciate. Many scientists from this program have also contributed to this handbook, although we have taken great care to include many other experts from the worldwide community of rationality research. The handbook is a truly international enterprise. We are very pleased that most of the world's leading experts were very enthusiastic and immediately agreed to contribute a chapter to this handbook. The different disciplines use quite different formats for footnotes, indexing, references, and other stylistic matters. To establish uniformity throughout the handbook, we have chosen to use the rules from the *Publication Manual of the American Psychological Association* (6th edition, 2010).

Our list of acknowledgments is long. First, we thank all our authors. In order to cover our field comprehensively, we had to engage the best authors across the disciplines. We expected that many authors would deliver too late, that we would have to wait a very long time for many contributions, that we would have to send out dozens of reminders, and so on. But we were too pessimistic. Most of our authors were very reliable, delivered their contributions fairly punctually, and were also very cooperative and responsive. We are very sad that one of our initial authors, Sieghard Beller, passed away before he could deliver his chapter and that another author, Arthur Merin, died shortly after finishing his chapter. We are most grateful to all of the authors for the exemplary cooperation.

We are also indebted to all our reviewers. In general, each chapter was reviewed by two reviewers, one from psychology and one from philosophy. In most cases, this was possible, and sometimes we even received three or four reviews from colleagues with different disciplinary backgrounds. For each chapter, at least one review came from an author of another chapter and the other from an external reviewer. We are very thankful to the following external reviewers:

Robert Audi
Anthony Blair
Thomas Blanchard
Luc Bovens
Matteo Colombo
Vincenzo Crupi
David Danks
Lorraine Daston
Eddie Dekel
Emmanuelle-Anna Dietz Saldanha
Igor Douven
Frank Fischer
Christian Freksa
Peter Gärdenfors
Valeria Giardino
Itzhak Gilboa
Anthony Gillies
Herbert Gintis
Geoffrey Goodwin
Werner Güth
Jaap Hage
Simon Handley
Sven Ove Hansson
Tim Henning
Risto Hilpinen
Franz Huber
Carsten Koenneker
Jakob Koscholke
Annette Lessmöllmann
Tania Lombrozo
Hugo Mercier
Philippe Mongin
Martin Monti
Josef Perner
Martin Peterson
Eric Raidl
Christian Ruff
Giovanni Sartor
Walter Schaeken
Mark Schroeder
Armin Schulz
Keith Stenning
Bob Sternberg
Jakub Szymanik
Raimo Tuomela
Gottfried Vosgerau
Thomas Voss
Peter Wakker
Paul Weirich
Gregory Wheeler
Kevin Zollman

We are very sad that Christian Freksa died in an accident before the book came out.

We mentioned that we could build on a long previous cooperation within the Priority Program *New Frameworks of Rationality*. Here we want to thank all members of the program for the lively and productive cooperation. It was most instructive.

We are deeply grateful to Christopher von Bülow, who checked all the manuscripts and brought them into the final uniform shape. He did so with unsurpassable care, at times at which we editors were already exhausted. It is also his merit that this project came to a good end.

We also thank Lupita Estefania Gazzo Castañeda, Andreas Kemmerling, and Christoph Klauer for their comments on an earlier version of the introductory chapter. Knauff wrote parts of the long introductory chapter at the University of California, Santa Barbara (UCSB), and thus wants to thank Daniel Montello for his hospitality and the perfect environment for focused thinking and writing. Knauff also wants to thank Estefania and his whole team for the many interesting and lively discussions. Spohn is doubly grateful to the DFG for support: in 2016–2018 through the research unit FOR 1614 at the University of Konstanz and in 2019–2020 through the excellence cluster EXC 2064/1 (project number 390727645) at the University of Tübingen.

We are also most grateful to Philip Laughlin and the team from MIT Press. The interaction with all of them was always pleasant and professional. This handbook was a huge enterprise and thus, in the final phase, an enormous work for the publisher, which was performed in an excellent way.

Normally, the editors form a unit that cannot thank itself. Still, we feel the urge to do this. We can only thank each other for the untiring willingness to work together over such a long period of time on such a large project. It was a great scientific experience that neither of us wants to miss, even though we spent so many days and nights discussing and Skyping, as well as exchanging (about 3,000) emails. Our spouses, Erika and Ulli, have faithfully accompanied our work with their understanding, charity, and support. We dedicate our work to them!

Markus Knauff and Wolfgang Spohn

Gießen and Konstanz, January 2021

Overview of the Handbook

Markus Knauff and Wolfgang Spohn

This handbook seeks to cover its topic, rationality, in an interdisciplinary way, more than any other existing handbook on this topic. Surely it cannot fulfill this intention completely. The topic is just too large. But it intends to reach broad representativity at least in its two main disciplines, philosophy and psychology. In addition, it presents some insights from disciplines such as artificial intelligence, behavioral and microeconomics, and neuroscience, although these areas are only partially covered here. The current program is already very ambitious.

The handbook is not organized along disciplinary lines of psychology and philosophy. This would be in clear contradiction to our interdisciplinary endeavor. Rather, it is organized along the common themes that occur in both fields, albeit often in different forms and under different names. In our introductory chapter, we propose various distinctions for classifying the overwhelming research on rationality. Certainly, the most important one is that between theoretical and practical rationality. This distinction is mirrored in the headings of part II and part III, the two main parts of this handbook. These parts are preceded by part I, which addresses some fundamental or propaedeutic matters concerning the history, some paradigms, and some key issues of research on rationality. The handbook concludes with part IV, which presents further facets of rationality that are relevant for both theoretical and practical rationality.

Each of the four parts consists of several sections, which in turn each contain several chapters, so that we ended up with 15 sections and 65 chapters. The order of chapters within a section does not have a particular significance. We just had to find one that creates connections between the different disciplinary backgrounds of the authors. The assignment of chapters to sections was more difficult. The field is so variegated and interlinked that no classification can be perfect. Still, we hope that we have organized the handbook in a way that makes it easy to read.

In our **introductory chapter**, we systematize the field by means of four binary distinctions: theoretical versus practical rationality, normative versus descriptive theories of rationality, individual versus social rationality, and outcome- versus process-oriented accounts of rationality. Since the distinctions can be combined with each other, we end up with a system of 16 compartments. Our chapter says something about most of the 16 compartments and uses this system not only to give guidance to the readers of this handbook but also to suggest a new classification system for researchers in human rationality. We recommend all readers to study this introductory chapter. This will certainly make it easier to orient oneself in, and to contextualize, the rich material collected in this handbook.

Part I: Origins and Key Issues of Rationality

The chapters in **section 1** deal with the history of rationality research and the cortical and evolutionary foundations of rational thinking. In chapter 1.1, *Thomas Sturm* describes the history of philosophical conceptions of rationality, which were just as much psychological conceptions. Since this history is so enormously rich, he just focuses on the emergence of the distinction between a descriptive and a normative perspective, which is so important for the entire handbook. This philosophical chapter is then complemented by chapter 1.2, in which *Jonathan Evans* summarizes the younger history of the psychology of reasoning from his personal point of view. Then, in chapter 1.3, *Gerhard Schurz* looks even farther back in time by speculating about the evolution of rationality. In chapter 1.4, *Vinod Goel* deals with the material basis of rationality, the brain. He presents findings from cognitive neuroscience and brain imaging on which cortical networks are involved in different kinds of rational thinking.

The chapters in **section 2** are concerned with some substantial philosophical and psychological topics of rationality. The section begins with chapter 2.1, in which

John Broome deals with the relation between rationality and reasoning. Both keywords are extensively used in this handbook, but they do not seem to denote exactly the same matter. The terms are also used quite differently in philosophy and psychology. In chapter 2.2, *Ralph Wedgwood* discusses the distinction between theoretical and practical rationality. This is also the top-level distinction in this handbook. The next chapters present various cognitive approaches to rationality. In chapter 2.3, *Philip Johnson-Laird* gives an overview of the theory of mental models, which is a far-reaching framework for explaining accurate and fallacious human reasoning. Another cognitive framework is the heuristics and biases approach, which is presented and critically discussed in chapter 2.4 by *Klaus Fiedler, Johannes Prager*, and *Linda McCaughey*. An equally influential framework is established by dual-process theories of reasoning. This framework and the related empirical evidence are presented by *Karl Christoph Klauer* in chapter 2.5. We could have placed the Bayesian reasoning theories here, too, as these theories also have a quite universal ambition. However, the relevant chapter by Chater and Oaksford fits even better in the section that is particularly dedicated to probabilistic reasoning. The last chapter in section 2 is a big jump toward artificial intelligence. In their chapter 2.6, *Johan van Benthem, Fenrong Liu*, and *Sonja Smets* outline the logico-computational perspective on rationality. They show how this approach can be efficiently used to solve rationality problems in computers. This chapter also creates a bridge to section 3 in part II, which is about logical and deductive reasoning.

Part II: Theoretical Rationality

Sections 3–7 are concerned with the key issues of theoretical (or “epistemic” or “doxastic”) rationality and reasoning. The first term originates from philosophy and is, in fact, not as common in psychology, in which cognitive theories of reasoning are a very active research field. However, the philosophical term is much broader and thus serves as the heading for the following sections. Not surprisingly, this part of the handbook starts with the two dominating reasoning paradigms: deductive logic as already conceived in ancient philosophy and probabilistic thinking as developed since the middle of the 17th century. These two paradigms are represented in sections 3 and 4, respectively.

The chapters in **section 3** are concerned with deductive logic and reasoning. Of course, the principles of propositional and first-order logic have long been part of the basic knowledge of our disciplines. Yet we wanted to explain at least in one place of this handbook what a logical proof is and how it is related to rational belief. This is done in chapter 3.1 by *Florian Steinberger*, who added some novel ideas to this classical topic. Propositional and first-order logic are also important in cognitive research on human reasoning. In chapter 3.2, *David O’Brien* presents the natural-logic account and some evidence in support of this account. In chapter 3.3, *Sangeet Khemlani* explains the cognitive foundations of syllogistic reasoning, which is a fragment of first-order logic. His approach is based on mental models as described in chapter 2.3.

Then, **section 4** is concerned with approaches from probability theory. In chapter 4.1, *Alan Hájek* and *Julia Staffel* explain the normative foundations of Bayesianism. Chapter 4.2 by *Stephan Hartmann* complements this by explaining the usefulness of the theory of so-called Bayes nets. A further complement is given in chapter 4.3 by *Arthur Merin*, which unfolds the probabilistic core of all considerations of relevance. This is important for accounts of reasoning, since premises or arguments are usually assumed to be relevant for their conclusions. In chapter 4.4, *Niki Pfeifer* presents his probability logic. In chapter 4.5, *Nick Chater* and *Mike Oaksford* defend their psychological perspective on rationality as conceived within Bayesianism. This, in turn, is complemented by chapter 4.6 by *Klaus Oberauer* and *Danielle Pessach*, who explain what, for (some) psychologists, conditionals have to do with probability. Chapter 4.7 by *Didier Dubois* and *Henri Prade* also proceeds from the assumption that beliefs, or epistemic states in general, come in degrees. This is why the chapter is placed in this section. Yet, in their account, rational degrees of belief do not behave like probabilities but in certain other ways. We think it is important to acknowledge that such alternatives exist.

Logic and probability theory are, of course, not the only paradigms of theoretical rationality and reasoning. In fact, many other frameworks exist, and many of them are not restricted to quantitative conceptions of degrees of belief, as in Bayesianism. These alternative accounts are represented in the remaining sections of part II.

The chapters in **section 5** are concerned with qualitative representations of belief and the related accounts of reasoning. In chapter 5.1, *Hans van Ditmarsch* presents the basic theory of doxastic and epistemic logic. Then, belief revision theory is concerned with the rational change of epistemic states. This normative account of belief revision is described in chapter 5.2 by *Hans Rott*. The dynamic account of belief revision theory is completed by ranking theory. This theory is laid out in chapter 5.3 by *Gabriele Kern-Isberner, Niels Skovgaard-Olsen*, and *Wolfgang Spohn*. The dynamics of belief is also important in cognitive theories of defeasible reasoning. In chapter 5.4, *Lupita Estefania Gazzo Castañeda* and *Markus Knauff* describe the empirical results on human belief revision and defeasible reasoning. They also explain how qualitative and quantitative theories try to account for the empirical findings. Defeasible reasoning, finally, is closely related to argumentation, where (rational) arguments drive (rational) epistemic change. The psychological perspective on argumentation is presented in chapter 5.5 by *Ulrike Hahn* and *Peter Collins*. Chapter 5.6 by *John Woods* deals with the same matter from a philosophical perspective.

There is also a close relation between epistemic change, on the one hand, and conditional and counterfactual reasoning, on the other, although the relation is not easy to specify. The broad range of approaches to reasoning with conditionals and counterfactuals is the topic of **section 6**. It begins with chapter 6.1 by *William Starr*, who reviews the different attempts to capture the logic of conditional and counterfactual constructions. In psychology, so-called supposition theory has become prominent. This attempt to explain human conditional reasoning is represented in chapter 6.2 by *David Over* and *Nicole Cruz*. In chapter 6.3, *Ruth Byrne* and *Orlando Espino* explain how human reasoners deal with counterfactual inferences. A special, but psychologically particularly relevant, case of conditional reasoning, so-called utility conditionals, is treated by *Jean-François Bonnefon* in chapter 6.4.

From counterfactual reasoning, it is just a small step to causal reasoning and its subform, diagnostic reasoning. The importance of causal reasoning among our many reasoning activities cannot be overemphasized. This is a huge topic on its own. Thus, we devote the complete **section 7** to it. In chapter 7.1, *Judea Pearl* presents his influential account of causal (and counterfactual) inference in probabilistic or statistical terms. How this kind of reasoning actually works in humans is discussed by *Michael Waldmann* in chapter 7.2. The special case of diagnostic reasoning is dealt with by *Björn Meder* and *Ralf Mayrhofer* in chapter 7.3. This is what this handbook offers concerning theoretical or epistemic rationality and reasoning.

Part III: Practical Rationality

Sections 8–12 deal with the core topics of practical rationality and decision making. Again, the first term is familiar in philosophy, and the second is mostly used in psychology. The terms are not equivalent, but they are close enough to be treated under one heading. We think that in this way, we can best highlight the interdisciplinary links and show how both research areas are connected.

This part of the handbook should of course start with accounts of individual decision making, which is the topic of **section 8**. In chapter 8.1, *Till Grüne-Yanoff* explains the main ideas of preference and utility theory. Next, chapter 8.2 by *Martin Peterson* presents standard decision theory and the normative arguments in its favor. However, this account has been criticized from the psychological as well as from the economic side. Indeed, prospect theory arose from the observation that people often deviate from expected utility maximization. This theory is explained by *Andreas Glöckner* in chapter 8.3. However, there are still further ways how beliefs and desires can be combined to determine a rational decision. Some of them are presented in chapter 8.4 by *Brian Hill*. Another critical view on economic decision theory comes from the theory of bounded rationality, which is influential in psychology and economics. This approach is reviewed in chapter 8.5 by *Ralph Hertwig* and *Anastasia Kozyreva*. The section ends with a chapter 8.6 by *Valerie Thompson, Shira Elqayam*, and *Rakefet Ackerman*, which is concerned with the connection between rationality and metacognition. The authors report empirical results that show that metacognitive control and monitoring processes are important for promoting our rational performance. Thinking about our own thinking makes us more rational. This chapter is quite general, and it might have fit under other headlines, but it also connects well to the previous chapters on practical rationality.

Then, **section 9** deals with game theory, the other standard theory of practical rationality. We treat here only so-called noncooperative game theory, which is, strictly speaking, still about individual practical rationality but in a social context. Chapter 9.1 by *Max Albert* and *Hartmut Kliemt* presents the classical theory, as it has been developed since the 1940s. In the 1980s, epistemic game theory emerged, which promises a more rigorous rationalization of the norms of game theory. This theory is described in chapter 9.2 by *Andrés Perea*. A different perspective is offered by evolutionary game theory, which is outlined by *J. McKenzie Alexander* in chapter 9.3. Again, these theories are criticized from the empirical point of view. This led to the rise of behavioral and psychoeconomics, which are sometimes (mistakenly) seen as fields of psychology but are still driven by the interests of economists. The many different ideas in this area are presented in chapter 9.4 by *Sanjit Dhami* and *Ali al-Nowaihi*.

In our introductory chapter, we explain why most research on rationality focuses on individual rationality.

In fact, **section 10** is the only section that really deals with social rationality in the proper sense. Therefore, it is not strictly limited to practical matters. In chapter 10.1, *Franz Dietrich* and *Kai Spiekermann* deal with social epistemology, which discusses the normative standards for group belief formation. Recently, related topics such as we-intentionality and collective rationality have developed into a larger philosophical field. These topics are presented in chapter 10.2 by *Hans Bernhard Schmid*. Communication, particularly linguistic communication, is a very special social and rational activity, which is the topic of chapter 10.3 by *Georg Meggle*. Rational choice theory has also become a strong paradigm in the social sciences. This line of research is explained by *Werner Raub* in chapter 10.4. A look into the history of philosophy, into political philosophy, and so on reveals that the standard theories of practical rationality go far beyond the scope of instrumental decision making. At least chapter 10.5 by *Julian Nida-Rümelin, Rebecca Gutwald*, and *Niina Zuber* discusses such extensions and under the label of "structural rationality." Finally, in chapter 10.6, *Leda Cosmides* and *John Tooby* present their adaptationist account of rationality. Although their evolutionary account seeks to explain all forms of human rationality, their chapter mainly focuses on practical rationality.

Practical reasoning need not proceed in quantitative terms of probabilities, utilities, or the like. Deontic and legal reasoning, for instance, can rely on qualitative representations and processes. Such qualitative accounts are the topic of **section 11**. In chapter 11.1, *John Horty* and *Olivier Roy* present the current state of the art in deontic logic (i.e., the logic of obligations and permissions). Then, in chapter 11.2, *Shira Elqayam* deals with the psychological theories of deontic reasoning. Legal reasoning is sufficiently different from deontic logic to deserve a separate treatment. This is given in chapter 11.3 by *Eric Hilgendorf* and in chapter 11.4 by *Henry Prakken*. A surplus of these chapters is that they represent different legal cultures, which heavily determine the field: chapter 11.3 is embedded in German law, while chapter 11.4 is based on the Anglo-Saxon legal system.

Whenever we are engaged in practical reasoning and decision making, moral issues come into play. In fact, they often are more important than issues of rationality. However, morality is a different topic, which we do not tackle in this handbook. We only wanted to address a few connections to rationality. These connections are the topic of **section 12**, which consists of two philosophical chapters and a psychological one. Chapter 12.1 by *Christoph Fehige* and *Ulla Wessels* describes the relation between rationality and morality in philosophy, while chapter 12.2 by *Michael Smith* deals with moral reasons, also from a philosophical point of view. But moral judgments are also a big topic in psychology. In chapter 12.3, *Alexander Wiegmann* and *Hanno Sauer* describe the psychological view on moral judgments and their relation to rationality. This concludes our chapters on practical rationality.

Part IV: Facets of Rationality

Sections 13–15 deal with further facets of rationality. Their topics are so distinct that they could not be subsumed under the headings of theoretical and practical rationality. In fact, most of the questions arise in *both* areas of human reasoning. For example, **section 13** is concerned with visual and spatial thinking. Such cognitive processes are an important complement to the many chapters in which reasoning is conceptualized, more or less explicitly, in linguistic or propositional terms. An important question in this context is the connection between logical and diagrammatic reasoning. In chapter 13.1, *Mateja Jamnik* takes the perspective of a computer scientist and argues that diagrams can help to reason logically. Another question is how humans and artificial intelligence systems reason about space and time. In chapter 13.2, *Marco Ragni* describes some differences and explains why people sometimes commit errors in spatial and temporal reasoning. In chapter 13.3, *Markus Knauff* explores whether visualization supports or hampers human reasoning. The empirical results show that it can indeed sometimes impede reasoning.

Another distinct topic is scientific rationality, which has theoretical as well as practical aspects, and which is, of course, of essential concern to us as scientists. The three chapters in **section 14** deal with different aspects of scientific rationality. One idea is that the sciences exercise a particularly sophisticated form of epistemic rationality, which may have stronger claims on objectivity. This idea is pursued in chapter 14.1 by *Line Edslev Andersen* and *Hanne Andersen*. Another topic is the value-freedom of science. This is the topic of chapter 14.2 by *Anke Bueter*. Finally, it is also important how scientific results are communicated to and perceived by the public. In chapter 14.3, *Rainer Bromme* and *Lukas Gierth* discuss the public understanding of science and its relation to scientific rationality.

The final **section 15** consists of still more chapters of psychological and even political interest. Chapter 15.1 by *Henry Markovits* explains how children learn to reason rationally. Chapter 15.2 by *Keith Stanovich, Maggie Toplak*, and *Richard West* presents the psychological findings concerning the connection between intelligence and rationality—it is not quite as close as it may seem. Chapter 15.3 by *Stephanie de Oliveira* and *Richard Nisbett*, finally, reports how training can improve people's rational thinking and that such training does not even need to be intensive or time-consuming.

This is also an optimistic message at the end of this handbook, which offers a big journey through the state of the art in rationality research. We can only recommend again to start this journey with our introductory chapter. It gives a systematic structure to the field and should help our readers to make cross-connections between the philosophical and psychological methods, theories, and ideas represented in this handbook. Research in these (and some other) areas has been conducted separately for too long, an anachronism that this handbook aims to overcome.

12.1 Rationality and Morality

Christoph Fehige and Ulla Wessels

Summary

Is practical rationality on the side of morality? Is it even the benchmark or the ground of morality? Why do what morality requires you to do? The diversity of answers that are still in the running and of considerations for and against them is astonishing. Our aim is to delineate the structure of the debate and to locate and clarify some major questions, options, and moves. We organize the presentation around a pair of prominent sample views, linking the rationality of an action to the agent's desires but its morality to the general welfare. We expound how different the matter looks for other views of rationality or of morality. On the whole, thoughts about each of the two normative domains and about conflicting norms in general suggest that, even regarding well-informed agents, rationality and morality cannot be fully harmonized. To some extent, convergence will remain gappy and contingent.

1. The Question

How do the judgments of practical rationality relate to those of morality? Let us call that question the RM question, with "R" for practical rationality and "M" for morality. The RM question is complex because each of the two relata is controversial in its own right—what *is* the rational thing to do, and what *is* the morally right thing to do?—and so is the pecking order among them: does an action have to be morally right in order to qualify as rational, or vice versa, or is there no such connection?

Judgments of the two kinds appear at variance in many cases that have existential weight. It may well happen, for example, that morality appears to require an affluent person to donate eighty percent of her income to a charity that saves lives efficiently, whereas practical rationality appears to require her to use the same money for completing her beloved collection of abstract paintings. We would expect a satisfactory answer to the RM question to get some kind of grip on such constellations. The answer should either show that R and M are in concord after all or tell us, insofar as they are not, what follows from the discord for theories of normativity and for thoughtful agents.

2. Actions and Requirements

As usual, it makes sense to restrict the inquiry. The plan is to look at rationality and morality in relation to actions. There are other items that beckon for our attention—think of the rationality and morality of beliefs, decisions, desires, emotions, intentions, maxims, or ways of life—but we will not extend the discussion to them. One consequence of the focus on actions is that the rationality that pertains is *practical* rationality by definition, which enables us to omit the adjective "practical" most of the time.

We will restrict the scope further by looking at only one kind of assessment by R and M, that of actions as (rationally or morally) required, permitted, optional, or forbidden. These are known as the "deontic" assessments, and it is the "all things considered" versions of them that we will be concerned with. We present the logical relations among the deontic terms in figure 12.1.1. There is one couple of terms, not in the diagram, that we will reserve for use in the domain R: the terms "rational" and "irrational" for actions that are rationally permitted and rationally forbidden, respectively. There is another couple, in the diagram,

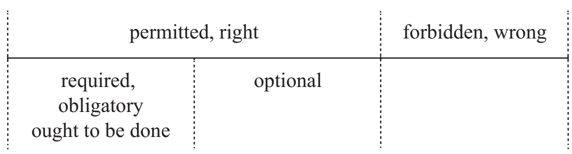

Figure 12.1.1
The main deontic terms and the logical relations among them. Terms written in the same field are roughly synonymous. Each term can occur in discourse about R and in discourse about M ("rationally permitted" vs. "morally permitted," etc.).

that today we will reserve for use in the domain M: in this chapter, the terms “right” and “wrong” will be used for moral assessments.

3. Two Sample Doctrines: Instrumentalism and Utilitarianism

It will help to look at the relations between one common criterion of R and one common criterion of M and to widen the view from there. As to the rationality of actions, many criteria that have been proposed are variations of one simple tenet: that it is rational for a person to try to get what she wants. There are competing ways of refining that outlook, and here is the version that will serve as our sample of a criterion: *An action that the agent can perform is rational if and only if the agent believes that no other action that she can perform brings about more fulfillment of her intrinsic desires*.

The performing, believing, and desiring in the criterion should all be understood as happening or obtaining at the same moment, but the believing and desiring don’t have to occur in the agent’s consciousness; they may be purely implicit. An intrinsic desire should be understood as a desire of something for its own sake, not of it as a means to something else; henceforth in this chapter, whenever we write “desire,” we will mean “intrinsic desire.” The quantitative notion of fulfillment takes into account both the number and the strengths of desires.

We will treat the criterion as the defining feature of “instrumentalism.”[1] The label is apt because the criterion codifies a view of actions as tools, assessing them with respect to their putative efficiency in achieving the agent’s ends. Instrumentalism is a close relative of that theory of rational decision-making that puts the maximization of expected utility (MEU) center stage and plays a large role in the behavioral sciences. The MEU theorist’s probabilities and amounts of utility correspond, by and large, to the instrumentalist’s beliefs and amounts of desire fulfillment.

In the moral domain, our sample criterion will be utilitarian: *An action is morally right if and only if no other action that the agent can perform brings about more welfare, worldwide*. The utilitarian message is that welfare counts, no matter whose welfare it is, and that nothing else does. What is that all-important stuff called “welfare”? One widely held view, and one that we will assume here, is that a person’s welfare is the fulfillment of her desires and that in consequence, due to conceptual connections between pleasure and desire fulfillment, pleasure is an important part of welfare.[2]

4. Convergence

To what extent do our sample criteria of R and M move in sync? One source of hitches can be an agent’s beliefs, which play a role in one criterion but not in the other. For example, even the most fervent utilitarian can have erroneous beliefs about the impact of her actions on the amount of general welfare, and those beliefs can make it rational for her to do what is morally wrong.

If we leave aside the threat posed by deficient beliefs, we reach more significant notions of convergence and divergence. Let us call an agent well-informed if her beliefs are in such a good state that the rational thing for her to do would not change if we improved them further (that is, if we corrected false beliefs or added true ones). Using well-informedness as a stepping-stone, let us understand “convergence” and “divergence” as follows: R and M converge insofar as the actions of well-informed agents that are morally required are also rational; R and M diverge insofar as the opposite is the case.

4.1 Convergence through Moral Desires

If instrumentalism is on the right track, the principal forces of convergence will have to be desires that point in the right direction. Less metaphorically speaking, they will have to be desires that, provided the agent is well-informed, have a propensity to make it rational for her to perform an action that is morally right. Let us call them RM desires. Life abounds with such desires, and one challenge is to produce a helpful classification.

Some RM desires aim directly at something that would be morally positive in itself. Figure 12.1.2 presents them as the “moral desires.” Some of those attitudes even “bring up” the topic of morality. Examples are desires that the world be a better place, to do the right thing, or to be a virtuous person. Other moral desires do not invoke morality as such but exhibit the relevant directness all the same. Depending on what is morally positive in itself, examples might be desires that some specific people be happy, that there be a lot of happiness, that everybody be treated equally—or desires to keep promises and to refrain from telling lies. The propensity that defines RM desires is present in either case. Applied to our sample moral doctrine, utilitarianism: both if Mary desires to do the right thing and if Mary desires to maximize welfare worldwide, it holds true that, provided she is well-informed, the desire will tend to make it rational for her to do the right thing.

Sympathy and moral sentiment, widespread as they are, serve as high-yield sources of those moral desires that do not invoke morality.[3] If Mary sympathizes with others,

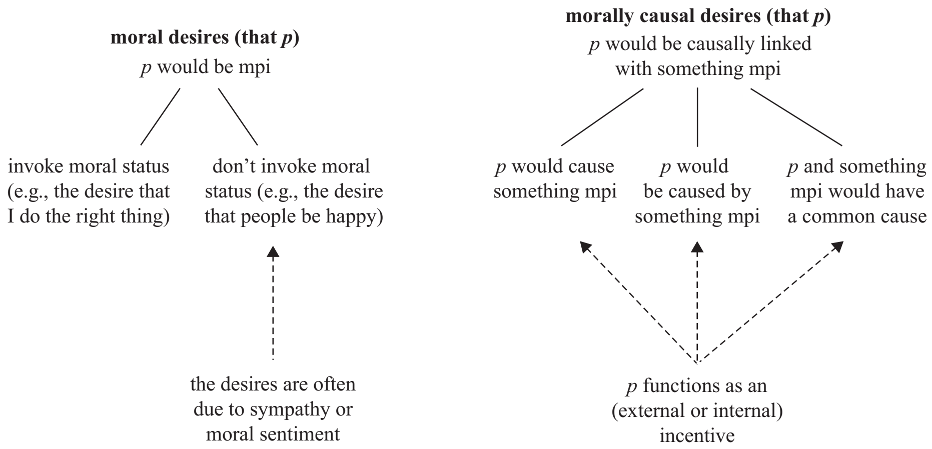

Figure 12.1.2
Some important kinds of RM desires. The abbreviation "mpi" stands for "morally positive in itself." One and the same desire can belong to more than one kind, even on the same level.

her sympathy is likely to give rise to or even to constitute a desire that others fare well. And if Mary has moral sentiments (for example, sentiments of moral admiration, indignation, or satisfaction) with regard to acts of a certain kind, the sentiments are likely to give rise to or even to constitute desires for or against performing acts of that kind. The two sources are so rich that there have been proposals to feed morality from them alone, giving us an "ethics of sympathy" or a "sentimentalist ethics."[4]

Are all moral desires contingent? Is it just as possible for a person to have them as to lack them? The rationality of morality would be a more robust affair if there were numerous strong moral desires that we cannot fail to have. An argument has been devised that purports to establish the existence of those resources. Everybody, the argument aims to establish, necessarily desires that other people have pleasure, with the strength of those desires proportional to that of the pleasure at issue. The idea is that, if you fully represented to yourself that another person experiences a specific pleasure, you would (this being entailed by full representing) experience that very pleasure yourself—and would thus yourself be pleased while representing. And since a disposition to be pleased when fully representing a state of affairs to oneself is a desire that the state of affairs obtain, you desire that the other experience the pleasure.[5] If the argument works, those desires—which on conceptual grounds everybody holds regarding everybody else all the time—will do a sizable part of the work that we are anxious to see done.

4.2 Convergence and Rationality without Egoism

Our glance at other-regarding moral desires, no matter whether they are necessary or contingent, is a good occasion for setting aside egoism. One might think on the following grounds that the relations between rationality and morality are particularly strained: (i) they would be strained if rationality required us to act egoistically, and (ii) according to instrumentalism rationality requires exactly that, because the word "egoistic" stands for the property of acting in the light of one's own desires.

The train of thought is misguided. The notion of "egoism" that is advanced in claim (ii) is both unusual and apt to weaken claim (i). The usual understanding is that acting on one's own desires is not a sufficient condition for being egoistic but that it also matters what those desires are. The usual understanding is that, if a good-hearted person strongly desires that other people fare well and acts on *such* desires, she is not an egoist but an altruist, while an egoist is a person who lacks or fails to put into action *such* desires. To be sure, somebody might go along with that understanding itself but still link instrumentalism to egoism by adding the claim that such desires do not exist. However, we know of no sound argument for the additional claim.[6]

The misclassification of rationality as egoistic can also result from sloppy thinking about the *combination* of instrumentalism and the theory of welfare as desire fulfillment. It is true that the combination entails some connection between rational action and the agent's welfare: in acting rationally at point of time t, the well-informed person-stage Mary-at-point-of-time-t maximizes the fulfillment of its desires and cannot help thereby maximizing its own welfare. We should beware of letting that connection blur the picture. In the first place, the rational person (or person-stage) at issue need not be concerned with her (or its) own welfare and can act, due to the nature of her (or its) desires, highly altruistically. Secondly, the person-stage that acts can still fail to maximize the welfare of the entire person, who is extended over time and may have different desires later. Thirdly, even the limited connection we are looking at requires as one ingredient a certain view of welfare; if we adopt a hedonistic view instead (seeing the welfare of an entity as the pleasure that the entity has and not as the fulfillment of her desires) and keep instrumentalism, the connection vanishes.

According to most conceptions of rationality, instrumentalism included, rationality does not require us to act egoistically. We need to distinguish the question how rationality relates to morality from the questions how egoism, how prudence, and how an agent's self-interest, self-love, or welfare relate to morality.[7] Many well-known discussions from the history of philosophy are largely about questions of the second kind. Plato, for example, assigns a key role in the *Republic* to the case of Gyges, who is ruthless in using for his own advantage his power to become invisible; Aristotle writes about virtue as a constituent of an agent's flourishing in the *Nicomachean Ethics* and Henry Sidgwick about individual versus universal happiness in *The Methods of Ethics*. Those discussions apply to our question at best partly or indirectly.

4.3 Convergence through Morally Causal Desires

Some RM desires, also indicated in figure 12.1.2, are related to the good or the right in a different way. If a person desires *p* and *p* would be causally linked to something that is morally positive in itself, we will speak of a "morally causal desire." We distinguish subgroups of such desires by distinguishing three kinds of the linkedness.

A desire that *p* is in the first subgroup if *p* would cause something morally positive. For example, a person desires to cultivate a garden, to keep the kitchen clean, or to mend broken bones, and her doing these things would cause pleasure in the beholders, eaters, and patients and would thus cause something that is morally positive in itself. A desire that *p* is in the second subgroup if, conversely, something that is morally positive in itself would cause *p*. For example, Mary desires that her parents treat her to cake, which the parents do only when, and in that case because, they are happy. The treat is desired and is the effect of the happiness, with the happiness being morally positive in itself. A desire that *p* is in the third subgroup if *p* and something that is morally positive in itself have a common cause. Consider, for example, heartless Mary, who does not care about the victims of malaria but who desires to be praised at a charity ball for donating to the fight against malaria. If Mary donates, neither does the praise cause the thing that is morally positive in itself (the praise does not causally affect the saving of the lives) nor vice versa, but the two have a common cause: Mary's donation. In such cases, too, the one comes with the other, and that matters for RM purposes.

The key feature, wherever morally causal desires are in play, is the indirectness. A desire can fail to be directed at anything that would be morally positive in itself and still tend to make actions rational that are (or that cause things that are) morally positive in themselves. Links of the relevant kind are legion: you do the right thing and some "other" desire of yours is fulfilled. In that sense, large chunks of morality are connected to rationality through carrots and sticks.

The incentives can be external—think of fellow-citizens who in response to your morally positive behavior honor and help you and refrain from ignoring, deriding, jailing, or lynching you. The incentives can also be internal—think of the joys of believing to have done what was good or right and of the absence of pangs and remorse. Even when the incentives are internal, the constellation differs from the one that characterizes the first main group of RM desires, the moral desires. Two distinctions apply. In the first place, when you desire *your joy* of doing good and when you desire *to do good*, those are two different desires.[8] Secondly, we should in some cases distinguish even for one and the same desire between the reason to put it into one group and the reason to put it into another. Consider, for example, your desiring your own joy of having done good. The reason to count that desire as a morally causal one (your joy is caused by something morally positive) is different from the reason to count it as a moral one (your joy, too, is something morally positive in itself). One desire can be both.

Incentives are studied by game theorists in particular. We understand quite well by now, with regard to several moral standards, how even agents who have no moral desires and are at the same time without ifs and buts instrumentally rational find themselves drawn into actions and outcomes that are morally positive. Under various conditions, repeated encounters in the same group of such agents become, not least because rewards and punishments can emerge, a breeding ground for various amounts of cooperation, equality, justice, solidarity, trust, and public good. While significant general results, most prominently a family known as the "folk theorems" of game theory, have been established by mathematical methods of the more traditional kind, there is also evidence from agent-based computer simulations, sometimes involving entire artificial societies: the rational thing to do for the virtual agents, it turns out, is often to spare or even to assist each other.[9]

So much for the many devices in human minds and societies that make it rational for people to do the right thing a lot of the time. On the other hand, given criteria of R or M resembling our two samples, there is not much hope of full convergence. In the example from our opening paragraphs, the agent might favor abstract paintings over human lives, which in the absence of competing considerations would make it the rational thing for her to violate her moral obligation. When incentives and the agent's moral desires do not add up, rationality will take a stand against morality.

5. Resisting Divergence by Aligning Rationality with Morality

The dominant impetus among philosophers is to keep the divergence of R and M in check. A proof that there is no divergence is the Holy Grail of practical philosophy, yearned for but hard to attain. Some philosophers go as far as to let their thinking on R itself or on M itself be governed by the premiss of "moral rationalism"—the claim that every action that is morally required is rational.[10] With that sweeping premiss or without, answers to questions like "Why be moral?" or "Why *act* morally?" are sought after, and the risk that, even with regard to well-informed persons, sometimes no good answer emerges is often perceived as an invitation to rethink R or M. Those who want to rethink in order to reduce or even eliminate divergence have essentially two options: they assimilate rationality to morality or vice versa. They could also combine the two moves, moralizing rationality *and* rationalizing morality, but for the bigger picture it will suffice if we treat each of the two in turn.

5.1 Smaller Departures from Instrumentalism

As to alternative views of practical rationality, modest deviations from instrumentalism make some difference regarding divergences. One example is the temporal extension of the conative basis. We could modify the instrumentalist criterion so that it covers not just the desires that the agent has at the time of acting but all those that she had, has, or will have. The modification would prevent the rational agent from behaving ruthlessly toward her past self or her future self. Morally speaking, it would be a step in the right direction—but only a small, intrapersonal one. Anchoring in a rational agent regard for the welfare of all her own person-stages is a far cry from anchoring in that agent regard for everybody's welfare.

A second example originates in the observation that, if instrumentalism is correct, rational agents can get caught in traps of practical rationality. Such traps are situations in which, if every agent acts rationally, each of them receives, predictably, less fulfillment of her desires than she would if everyone acted irrationally. General compliance with the requirements of rationality thus leads to the opposite of the moral goal, which is general welfare.

Many real-life situations appear to be traps of practical rationality. Consider societies that become much nastier because people carry weapons. Each individual reasons that to carry a weapon is likely to be more conducive to the fulfillment of her desires either way—that is, both if people whom she encounters carry a weapon and if they don't. Because of the individual decisions for weapons, the considerable advantages of a weaponless society and of avoiding an arms race remain out of reach for all. Even if a weaponless society could still be achieved by changing the individual decisions through the imposing of sanctions, the sanctioning itself would gobble resources and thus welfare. The ubiquity of such traps in human interaction and the unequivocalness of the moral setbacks—less welfare for one and all—have incited a vast complex of research in ethics, economics, and psychology.[11]

Can we devise a criterion of rationality that preserves the spirit of instrumentalism but spares us the traps? David Gauthier is one of the theoreticians who have tried.[12] Gauthier suggests, controversially, that a rational agent can choose, "on utility-maximizing grounds, not to make further choices on those grounds," but to adopt a certain stance that will determine her behavior. She will then be a "constrained maximizer." The stance that Gauthier has in mind is roughly this: I will do, provided that so will the others who are involved, my part in making possible an outcome that is better for everybody than the outcome that unconstrained maximizers could achieve. Since an agent who has recognizably adopted such a stance will sometimes produce and reap fruits of cooperation that are not available to one who hasn't, it is rational to become such an agent. Individual instrumental rationality is declared to be a tad more collective than we thought.

If Gauthier's claims are sound and help us make headway with the RM question, they do so within limits. Gauthier's ambition is restricted to establishing rational support for a morality of *mutual* benefiting. He has no ambition to provide support for moral obligations toward beings who are in need but have nothing to offer.

5.2 Larger Departures from Instrumentalism

We find more radical consequences for the shape and extent of divergence when we turn our attention to more radically different pictures of practical rationality. The main move is the dethroning of desire.

Suppose that we break with instrumentalism by replacing the appeal to desire fulfillment with an appeal to conformity to reasons. Let us agree that facts can be reasons and that constellations of facts (for example, regarding pains and cures) can make it the case that there is, altogether, for a person more reason to perform one action (say, to see her cardiologist) than another (say, to see her homeopath). The new criterion of rationality could then read as follows: *An action that the agent can perform is rational if and only if she believes that there is not more reason for her to perform another action instead.*[13] Suppose further that we take many reasons to be "worldly" in that they do not involve the agent's desires. Perhaps, for example, *that Peter's education would be finished if he received a donation* is a reason to donate, even if the agent desires no such finishing or donating and desires nothing that comes with it; or perhaps *that serenading the moon would increase the glory of the moon* is a reason to serenade the moon.[14]

The impact on the RM nexus would be momentous. The task of showing that it is rational for a well-informed agent to do what morally she ought to do no longer has as its central component the task to track down a fitting constellation of desires of hers but the task to identify a fitting constellation of reasons for her to do things. In order for the entire scheme to be successful, worldly reasons for actions need to exist *and* to have a considerable affinity to morality *and* to be connected to rationality as stated by the new criterion. Whether all or at least some of the three conditions are fulfilled is controversial.

Which morality you take to be connectible to rationality-in-the-new-spirit will depend on your views, possibly your intuitions, about the realm of reasons. If morality is concerned with promise-keeping, loyalty, or the increase of human knowledge, and so are reasons to do things and in the same proportion, then every well-informed rational agent will do the right thing: keep her promises and so on. Our sample moral doctrine, utilitarianism, is no exception. If there is always most reason to maximize the amount of the fulfillment of everybody's desires, then every well-informed rational agent will do the right thing and maximize that amount.

Following a markedly different path, Immanuel Kant, too, ends up advocating some variety of rational benevolence, maximization included. We read that a rational being would try, "as far as he can, to advance the ends of others" and would come to the conclusion: "the ends of a subject that is an end in itself must, as much as possible, also be *my* ends."[15] In which way according to Kant rationality secures so much morality is extraordinarily contentious, even among his followers. He draws on the claim that a rational being is autonomous in the literal sense of giving herself a law. To Kant, that feature appears imbued with moral significance. Being autonomous, the agent's rational will is not pushed around by anything, not even by the agent's own inclinations; it finds itself with nothing left to be constituted by than the respect for rationality itself and for people who have it and for their ends; and, such being the character of laws, that will is general, not concerned with one person or group in particular. Sound statements of those connections remain a desideratum.

As expected, the picture of the relation between R and M changes when that of R changes. Our brief encounters with the "worldly reasons, not desires" approach and the "autonomy, not desires" approach have illustrated tectonic movements and the hopes of convergence that can be associated with them.

6. Resisting Divergence by Aligning Morality with Rationality

If divergence is to be avoided, what about getting to work at the other end? We could truncate morality. The fewer actions are morally obligatory in the first place, the smaller the risk that an action is morally obligatory but irrational.

The demands of some moralities are so removed from most agents' conative constitution that cutting down on the demanding looks particularly promising. According to utilitarianism, for example, an agent ought to give the same weight to her own welfare and to that of her nearest and dearest as to everybody else's. Since hardly an agent's desires manifest such impartiality, utilitarian obligations that it is irrational for the obligated agent to comply with are thick on the ground.

Various moral prerogatives for agents have been suggested. We could permit the agent, for some factor $k > 1$, to attach up to k times as much weight in her decisions and actions to her own welfare as to the welfare of everybody else; or permit her, for some $l > 0$, never to give up more than l units of her own welfare; or permit her, for some threshold value m, never to make her own welfare fall below *m*.[16] We could think of other permissions, too, concerning her projects rather than her welfare or concerning the welfare of those she is close to rather than just her own. Every such prerogative is a loosening of utilitarianism; it would allow the agent to be partial in the sense of granting in her decisions extra force to this or that, even if the amount of worldwide welfare suffers.

Prerogatives are apt to narrow the gap between R and M but will not close it. A morality that involves a prerogative will still demand *something* (for instance, that the interests of others be respected to *some* extent), and there is bound to be some well-informed agent who in some situation has desires that make it irrational for her to comply even with those moderate demands.

Structurally, contract theories in the Hobbesian tradition inhabit the same middle ground. Those theories are reciprocitarian. They say, with various qualifications, that what a person morally ought to do in relation to another person is to play by rules that satisfy the following condition: the fulfillment of the desires of each of the persons would increase if each of them, rather than none, played by those rules. For example, if both refrain from insulting each other, that will save both of them some distress.

Once again, we find the curtailing of moral obligation. If contractarians are right, there are fewer moral obligations than we thought and thus fewer that it might be irrational to meet. Most notably, about all persons (and other beings) who are not in a position to increase the fulfillment of her desires, the contractarian will say that she owes them nothing. And once again we also find that the curtailing does not suffice to provide full alignment with rationality. Sometimes a well-informed agent will have and see the possibility of maximizing the fulfillment of her desires by breaking even the few rules of that "minimal morality" and getting away with it. And no earthly regime of sanctions will eliminate such possibilities, since all such regimes are gappy.[17]

Some thinkers have suggested that we go one step further and prevent all divergence by fully rationalizing morality. The proposal is to combine moral rationalism with the claim that instrumentalism is by and large on the right track. Gilbert Harman endorses the combination. He writes: "If *S* says that (morally) *A* ought to do *D*, *S* implies that *A* has reasons to do *D*." And he continues: "I assume that the possession of rationality is not sufficient to provide a source for relevant reasons, that certain desires, goals, or intentions are also necessary." Since agents might lack the relevant attitudes, Harman infers "that there might be no reasons at all for a being from outer space to avoid harm to us" and "that, for Hitler, there might have been no reason at all not to order the extermination of the Jews."[18]

The moral consequences are remarkable. If the moral "ought" requires reasons, and reasons require desires, but the desires aren't in place and thus neither are the reasons, then neither is the moral "ought." While Harman invites us to say other things about Hitler (for example, that Hitler is evil or our enemy), he claims that the moral "ought" is out of place. It is "odd to say," Harman asserts, that "it was wrong of Hitler to have acted as he did" or that "Hitler ought morally not to have ordered the extermination of the Jews."

And so the rationalization of the moral "ought" would be completed. The approach is Procrustean. There is no mismatch between the moral "ought" and the practical rationality of its addressees because every instance of a moral "ought" that would not conform is discarded.

7. Accepting Divergence

Should we accept that R and M diverge? We should if it seems to us that they do and we see no reason to resist the claim. The two parts of that condition deserve separate treatments.

7.1 Finding Some Divergence in the First Place

Not many of us enter the inquiry holding fully convergent views of R and of M, each of which they deem plausible in its own right, independently of any pressure to see the two in line. It is true that there is full convergence according to some views of R and of M that we have encountered, but the independent plausibility of those views is the crux.

On the rational side, how plausible is the claim that something other than the agent's aims and projects rules the roost? The subjective picture exerts a considerable pull: if you've set your heart on something, it is rational for you to go after it, and if you haven't, it is not. On the moral side, the questions are inverse. How plausible is the claim that, provided you are indifferent to other people's welfare and have to fear no backlash from wrecking it, you are morally permitted to wreck it? On either side, the claims that would secure full convergence do not have the ring of truth.

The problem does not just arise when, at least regarding well-informed agents, one of the two domains is claimed to fully look the way we always took the other one to look, with the implausibility due to the fact that *one* view does all the reaching out. If a view of R and a view of M met halfway, the implausibility would be distributed evenly but not lessened. Although we pointed out that some independently plausible components with a conciliatory effect may well be missing from our two sample doctrines (which we didn't call sample doctrines for nothing), no such components are in sight that would happen to dovetail, resulting in a perfect fit.

7.2 Not Theorizing Away the Divergence We Find

Matters would look different if our thinking were driven by the will to rule out divergence. Maybe there are general considerations that speak against divergence and that should be taken into account when we form our views of R and M? It is surprisingly difficult to find distinct statements of such considerations in the literature, but here is a brief attempt to articulate and assess some candidates.

The argument from the *negativity of non-compliance* says: "Since people by and large act rationally, every morally required act that is irrational is a morally required act that is unlikely to be performed—which is a bad thing." We respond that any such badness would be a sad truth and that good theorizing should acknowledge truths, sad ones included. The badness at issue does not provide a right kind of reason for changing our views of R or M.

A follow-up argument adduces the *pointlessness of moral judgment due to the negativity of non-compliance*: "The point of engaging in moral judgment is to avoid the said negativity. Why bother if the project does not boost the right and the good through compliance?" Part of the answer is that not only are there many different functions that moral thinking, judging, and speaking have, but also many different paths, including indirect ones, on which the function of boosting the good and the right can be fulfilled. Some of the moral point of considering or making or uttering the moral judgment that, say, Mary ought to spend half of her spare time fighting for a certain cause may well be independent of the factual question whether Mary ends up conforming to the judgment. Sorting out and signaling our moral view of the matter can help shaping decisions, education, outlooks, politics, relationships, and sanctions in a myriad of ways, many of which do not even relate to Mary in particular.

Next, there is the threat of the *unjustifiedness of moral judgments*: "To 'provide morality with a foundation' or to 'justify moral judgments' is or includes showing that it would be rational to act, if the opportunity arose, in line with the moral judgments at issue. Moral judgments that are subject to divergences are therefore unfounded and unjustified." The complaint is worth pondering. Still, you justify something *to somebody*, and so the premiss of the complaint licenses at best the conclusion that, in cases of divergence, the "ought" judgment has not been justified to *all* persons who, according to the judgment, ought to do a certain thing under certain circumstances. But maybe to some of them. Secondly, the premiss of the complaint is controversial in that there are other conceptions of what it is to justify a moral judgment. Thirdly, let's not forget that generally speaking, since justification stops somewhere, the use of unjustified items might be respectable.

A final argument asserts the *inacceptability of normative impasses*: "When an agent grasps the fact of the situation and the morally required action is irrational, what is there left for us to tell her? We can tell her that two kinds of norms that apply to her impending action point in opposite directions, irreconcilably so, and that we have no third kind of norm that would adjudicate between them. We can wish her good luck at the normative crossroads and move on. None of that is of any help to her, and theorizing about normativity should do better than that. It should avoid divergences."

The most general reply to the objection is that theorizing about normativity should "do better" only if there is independent evidence that it got the lay of the land wrong. That evidence would need to be produced—and will hardly consist in the fact that something is or feels awkward for a well-informed agent. Moreover, the objection misses its target, divergence, because divergence does not entail the alleged source of awkwardness, the absence of normative guidance. Divergence allows for the possibility that there is a boss: one of the divergers or some adjudicator. The possibility is very much alive in the literature on normative pluralism, where the view is not rare that divergences between kinds of norms coexist with unequivocal overall norms.[19]

When it comes to R and M, the observation that divergence and guidance can coexist gains in stature, since one of the two divergers is practical reason itself. There is a fairly straightforward sense in which practical reason is always in charge. Practical reason deals with practical reasons—with all of them. The fascinating question whether all of them involve agents' desires makes no difference for the following consideration, which is quite general. Neither does it make a difference whether some kind of incommensurability threatens to hold *among* an agent's practical reasons. We may ignore that possibility here because we're grappling with the specter of lacks of guidance that originate *between* R and M. Our question is whether the risk of such lacks exists and justifies the axiomatic excluding of a divergence of R and M. Surely a risk of such lacks could not justify such an excluding if there is lack of guidance even *within* R. Thus, what remains to be looked at is only the other case, in which all is well within R: there is some balance of all practical reasons that an agent has and thus something that the agent has most reason to do.

And now to the question where the balance leaves *moral* practical reasons. They can relate to the "most reason" verdict in two different ways but can escape neither the verdict nor the guidance it gives. We can understand moral practical reasons either as being practical reasons of a certain kind (think of yellow bicycles, which are bicycles), in which case the balance of all practical reasons will have taken them into account—or as *not* being practical reasons (think of "root beer," which is not beer, and of "toy money," which is not money), in which case the balance of all practical reasons will not have taken them into account. There is guidance by the balance either way. It is guidance on the level "most practical reason," a level on which all practical reasons—that is, all reasons to do things—have been taken into account. That seems guidance enough.

From the list of objections against divergence, not much is left. It seems that we should bemoan, but not deny, the existence of divergence.

8. Conclusion

We have explored in which sense and why it is in many cases rational for people to do what morally they ought to do but also why with respect to many cases, even of well-informed agents, the diagnosis is controversial. We have sided with the common-sense view that performing actions that are stupid (to use the laity's term for "irrational") and performing actions that are morally wrong are two very different kinds of shortcomings. The action that is stupid can be altruistic, benevolent, beneficent, and morally right, and the action that is not stupid can be egoistic, malevolent, maleficent, and morally wrong. Given the duality, theoreticians of the rational and the moral will keep or turn their spotlights on the kinds, mechanics, and extents of convergence and divergence, including the metanormative challenges posed by competing norms. In practice, both desires and ways of fulfilling them ought to be shaped—and that "ought" is a moral one—so that divergence is reduced.

Acknowledgments

More people have given us a hand regarding some aspect or other of this chapter than we can list here, and we are grateful to every one of them. Special thanks go to Kevin Baum, Inga Lassen, Susanne Mantel, Helge Rückert, Rudolf Schüßler, Stephan Schweitzer, and Christian Wendelborn.

Notes

1. More on doctrines like instrumentalism and on their rivals in Hutcheson (1728, section 1 of treatise 2), Millgram (2001), Schroeder (2007), and Schmidt (2016). For further references, see Fehige (2001, note 1).

2. For the equating of welfare with desire fulfillment, see von Wright (1963, esp. sections 5.9 and 5.11) and Carson (2000, chapter 3); see also Wessels (2011), with numerous sources. For the claim that pleasure is one kind of desire fulfillment, see Fehige (2004, esp. pp. 143–145) and Heathwood (2007), both of which give further references. As to normative ethics: utilitarianism and its rivals are sketched in Vaughn (2010/2013, chapter 2) and treated more thoroughly in Copp (2006, part 2); a helpful introduction to utilitarianism is Bykvist (2010).

3. Hutcheson (1728) covers in some detail the desires associated with the "publick sense," which is the disposition "to be pleased with the *Happiness* of others, and to be uneasy at their *Misery*" (art. 1.1 of treatise 1), and with the "moral sense," which is the disposition to have moral sentiments. More on the connections between morality, desire, sentiment, and sympathy in Bricke (1996, esp. chapter 6), Fehige (2004), and Fehige and Frank (2010).

4. Single-source approaches to morality have been championed, for example, by Arthur Schopenhauer (1841, esp. section 16), who counts on sympathy, and by Francis Hutcheson (1725, preface and treatise 2) and David Hume (1751, appendix 1), who count on the moral sentiments.

5. The argument is put forth in Fehige (2004); the book includes a discussion, in chapter 6, of the limits of even that connection between R and M.

6. Lucid treatments include Hutcheson (1728, esp. art. 1.3 of treatise 1 and the beginning of treatise 2) and Sharp (1923, section 2).

7. Gregory Kavka captures the difference with maximum clarity when he distinguishes the "Wider Reconciliation Project" (1985, section 5) from a narrower one. More on handling egoism and its relation to morality can be found in Cholbi (2011).

8. More on the important distinction, for example, in Hutcheson (1728, art. 1.4 of treatise 1), Rashdall (1907, esp. pp. 17–18, 28–32), and Schlick (1930, sections 2.6–2.8).

9. For the first kind of approach, see Fudenberg and Tirole (1991, chapter 5, esp. section 5.1.2) and Maschler, Solan, and Zamir (2008/2013, chapter 13, esp. sections 13.5 and 13.6). The agent-based simulation approach is surveyed by Gotts, Polhill, and Law (2003); a telling example is Hegselmann (1998).

10. Moral rationalism is also known as the “claim of overridingness” and is highly controversial; contributions to the dispute include Scheffler (1992a, chapter 4), Cholbi (2011, esp. section 1), Portmore (2011), and Dorsey (2012). An illuminating critical study of ways of engineering convergence is Brink (1992).

11. Peterson (2015) can serve as a gateway to the area; see also Dawes (1980) and Binmore (1994).

12. A good starting point is Gauthier (1986, esp. chapter 6, which has the upcoming quotation on p. 158). Appeals to “resolute choice” (McClennen, 1985) bear some resemblance to Gauthier’s “constrained maximization.” Gauthier later espouses a conception of rationality that he acknowledges is morally charged to begin with (2013, p. 624); for a similar step, see McClennen (2012).

Other devices, too complex to sketch here, have been invoked against the traps: the claim that people who are in similar circumstances are bound to act similarly (see Davis, 1985, and, on “mirror strategies,” J. V. Howard, 1988), the proposal to shift our attention from Nash equilibria to “dependency equilibria” (Spohn, 2009, foreshadowed by Aumann, 1987), and, although with an emphasis on the explanatory rather than the normative dimension, the conceptualization of a relevant decision situation as a “metagame” that involves partial commitments (N. Howard, 1971, esp. sections 2.5, 3.1, 3.2) or of the decision-makers as members of a group (Bacharach, 2006, esp. section 4 of the “Conclusion,” and Butler, 2012). For critical thoughts on some of the approaches, see, e.g., Binmore (1994, chapter 3).

13. The view that rationality is the corresponding of actions to beliefs about reasons can be spelt out in quite different ways; see, for example, Scanlon (1998, sections 1.1.3–1.1.5) and Parfit (2011, sections 1 and 17).

14. The claim that there are worldly normative reasons for actions has acquired quite a few supporters. Among them are Thomas Nagel (1970, esp. chapter 10, and 1986, esp. sections 8.4, 8.5, 9.2, 9.3), Thomas Scanlon (1998, sections 1.9–1.11), Jonathan Dancy (2000, chapters 2 and 5), Philippa Foot (2001, chapter 4), Frederick Stoutland (2001, sections 3.2 and 3.3), and Derek Parfit (2011, sections 1–15).

15. The quotations are from Kant (1785/1903, p. 430); for the moral impact of autonomy, see, for example, pp. 405, 428–434, 444, 447–452. Kant later invokes the moral law as a “fact of reason” (1788/1908, pp. 31–32, 42–43); whether in doing so he renounces, summarizes, or supplements his justificatory efforts is controversial. After Kant, attempts to ground morality in the respect for rationality itself have been numerous and varied; Smith (2011) is one case in point.

16. The three kinds of prerogatives are considered, one each, in Scheffler (1992b, section 1), Mulgan (2001, section 5.5), and Nagel (1986, p. 202). For further reflections on possible kinds and shapes of prerogatives, see Wessels (2002) and Stroud (2010). Some prerogativists make it clear that anti-divergence is where they come from. It is for the sake of keeping M in the orbit of R that they conceive of M as partial and as in that respect non-utilitarian (see Nagel, 1986, pp. 202–203, and Portmore, 2011).

17. A particularly acute analysis of the connections between R and M in the contractarian project is provided by Rainer Trapp (1998), who also explains (pp. 339, 356–359) that it will not always be rational for a person to do what by contractarian lights she morally ought to do. Contractarians who acknowledge the divergence include Peter Stemmer (2017, pp. 646–648, esp. note 34) and Gregory Kavka (1985, pp. 305–308 and, most clearly in terms of rationality, section 5). That a contractarian like Gauthier might avoid the divergence by operating with a modified conception of rationality (Gauthier, 1986, chapter 6) is a different matter.

18. The preceding quotations are from Harman (1975, p. 9); the subsequent ones are from Harman (1977, p. 107). Stemmer has retracted the crucial claims (2017, note 34) but used to travel a very similar path. For a while, he saw norms geared so radically to the addressee’s wanting that even an accidental hole in the sanctioning was considered to constitute a hole in the norm itself. Insofar as an individual action would not be followed by a sanction that the agent herself wants to avoid, so the retracted claims run, the norm not to perform that action “does not exist,” and the action “is not really forbidden” (Stemmer, 2008, p. 181).

19. That domain-specific “ought” judgments relate to overarching ones, which have the last word, is argued by McLeod (2001) and Woods (2018, section 10.2.2). In a similar spirit, Case (2016) provides a powerful argument for the conditional claim that, if you accept “source pluralism” and “conflict,” you are committed to accepting “authoritative adjudication.” However, support for the claim that “ought” judgments from different domains diverge is wider and comes also from authors who deny that there is an adjudicative level—see Baker (2018) and the sources given there.

References

Aumann, R. J. (1987). Correlated equilibrium as an expression of Bayesian rationality. *Econometrica*, *55*, 1–18.

Bacharach, M. (2006). *Beyond individual choice*. Princeton, NJ: Princeton University Press.

Baker, D. (2018). Skepticism about ought *simpliciter*. *Oxford Studies in Metaethics*, *13*, 230–252.

Binmore, K. (1994). *Playing fair*. Cambridge, MA: MIT Press.

Bricke, J. (1996). *Mind and morality*. Oxford, England: Oxford University Press.

Brink, D. O. (1992). A puzzle about the rational authority of morality. *Philosophical Perspectives*, *6*, 1–26.

Butler, D. J. (2012). A choice for 'me' or for 'us'? Using we-reasoning to predict cooperation and coordination in games. *Theory and Decision*, *73*, 53–76.

Bykvist, K. (2010). *Utilitarianism*. London, England: Continuum International.

Carson, T. L. (2000). *Value and the good life*. Notre Dame, IN: University of Notre Dame Press.

Case, S. (2016). Normative pluralism worthy of the name is false. *Journal of Ethics and Social Philosophy*, *11*, 1–19.

Cholbi, M. (2011). The moral conversion of rational egoists. *Social Theory and Practice*, *37*, 533–556.

Copp, D. (Ed.). (2006). *The Oxford handbook of ethical theory*. Oxford, England: Oxford University Press.

Dancy, J. (2000). *Practical reality*. Oxford, England: Oxford University Press.

Davis, L. H. (1985). Is the symmetry argument valid? In R. Campbell & L. Sowden (Eds.), *Paradoxes of rationality and cooperation* (pp. 255–263). Vancouver, Canada: University of British Columbia Press.

Dawes, R. M. (1980). Social dilemmas. *Annual Review of Psychology*, *31*, 169–193.

Dorsey, D. (2012). Weak anti-rationalism and the demands of morality. *Noûs*, *46*, 1–23.

Fehige, C. (2001). Instrumentalism. In E. Millgram (Ed.), *Varieties of practical reasoning* (pp. 49–76). Cambridge, MA: MIT Press.

Fehige, C. (2004). *Soll ich?* [Should I?]. Stuttgart, Germany: Reclam.

Fehige, C., & Frank, R. H. (2010). Feeling our way to the common good. *The Monist*, *9*, 141–165.

Foot, P. (2001). *Natural goodness*. Oxford, England: Oxford University Press.

Fudenberg, D., & Tirole, J. (1991). *Game theory*. Cambridge, MA: MIT Press.

Gauthier, D. (1986). *Morals by agreement*. Oxford, England: Oxford University Press.

Gauthier, D. (2013). Twenty-five on. *Ethics*, *123*, 601–624.

Gotts, N. M., Polhill, J. G., & Law, A. N. R. (2003). Agent-based simulation in the study of social dilemmas. *Artificial Intelligence Review*, *19*, 3–92.

Harman, G. (1975). Moral relativism defended. *Philosophical Review*, *84*, 3–22.

Harman, G. (1977). *The nature of morality*. New York, NY: Oxford University Press.

Heathwood, C. (2007). The reduction of sensory pleasure to desire. *Philosophical Studies*, *133*, 23–44.

Hegselmann, R. (1998). Experimental ethics. In C. Fehige & U. Wessels (Eds.), *Preferences* (pp. 298–320). Berlin, Germany: De Gruyter.

Howard, J. V. (1988). Cooperation in the prisoner's dilemma. *Theory and Decision*, *24*, 203–213.

Howard, N. (1971). *Paradoxes of rationality*. Cambridge, MA: MIT Press.

Hume, D. (1751). *An enquiry concerning the principles of morals*. London, England: Millar.

Hutcheson, F. (1725). *An inquiry into the original of our ideas of beauty and virtue*. London, England: Darby.

Hutcheson, F. (1728). *An essay on the nature and conduct of the passions and affections: With illustrations on the moral sense*. London, England: Darby & Browne.

Kant, I. (1903). *Grundlegung zur Metaphysik der Sitten* [Groundwork of the metaphysics of morals]. In *Kant's gesammelte Schriften* (Vol. 4, pp. 385–463). Berlin, Germany: Reimer. (Original work published 1785. English quotations from *Groundwork of the Metaphysics of Morals*, trans. M. Gregor & J. Timmermann, 2011, Cambridge, England: Cambridge University Press.)

Kant, I. (1908). *Kritik der praktischen Vernunft* [Critique of practical reason]. In *Kant's gesammelte Schriften* (Vol. 5, pp. 1–163). Berlin, Germany: Reimer. (Original work published 1788)

Kavka, G. S. (1985). The reconciliation project. In D. Copp & D. Zimmerman (Eds.), *Morality, reason and truth* (pp. 297–319). Totowa, NJ: Rowman & Allanheld.

Maschler, M., Solan, E., & Zamir, S. (2013). *Game theory* (Z. Hellman, Trans.). Cambridge, England: Cambridge University Press. (Original work published in Hebrew 2008)

McClennen, E. F. (1985). Prisoner's dilemma and resolute choice. In R. Campbell & L. Sowden (Eds.), *Paradoxes of rationality and cooperation* (pp. 94–104). Vancouver, Canada: University of British Columbia Press.

McClennen, E. F. (2012). Rational cooperation. *Synthese*, *187*, 65–93.

McLeod, O. (2001). Just plain "ought." *Journal of Ethics*, *5*, 269–291.

Millgram, E. (Ed.). (2001). *Varieties of practical reasoning*. Cambridge, MA: MIT Press.

Mulgan, T. (2001). *The demands of consequentialism*. Oxford, England: Oxford University Press.

Nagel, T. (1970). *The possibility of altruism*. Oxford, England: Oxford University Press.

Nagel, T. (1986). *The view from nowhere*. Oxford, England: Oxford University Press.

Parfit, D. (2011). *On what matters*. Oxford, England: Oxford University Press.

Peterson, M. (Ed.). (2015). *The prisoner's dilemma*. Cambridge, England: Cambridge University Press.

Portmore, D. W. (2011). Consequentialism and moral rationalism. *Oxford Studies in Normative Ethics*, *1*, 120–142.

Rashdall, H. (1907). *The theory of good and evil* (Vol. 1). Oxford, England: Clarendon Press.

Scanlon, T. M. (1998). *What we owe to each other*. Cambridge, MA: Harvard University Press.

Scheffler, S. (1992a). *Human morality*. Oxford, England: Oxford University Press.

Scheffler, S. (1992b). Prerogatives without restrictions. *Philosophical Perspectives*, *6*, 377–397.

Schlick, M. (1930). *Fragen der Ethik* [Problems of ethics]. Vienna, Austria: Springer.

Schmidt, T. (2016). Instrumentalism about practical reason: Not by default. *Philosophical Explorations*, *19*, 17–27.

Schopenhauer, A. (1841). Preisschrift über die Grundlage der Moral [Prize essay on the basis of morals]. In *Die beiden Grundprobleme der Ethik* (pp. 101–280). Frankfurt/Main, Germany: Joh. Christ. Herrmannsche Buchhandlung.

Schroeder, M. (2007). *Slaves of the passions*. Oxford, England: Oxford University Press.

Sharp, F. C. (1923). Some problems in the psychology of egoism and altruism. *Journal of Philosophy*, *20*, 85–104.

Smith, M. (2011). Deontological moral obligations and non-welfarist agent-relative values. *Ratio*, *24*, 351–363.

Spohn, W. (2009). Wider Nash-Gleichgewichte [Against Nash equilibria]. In C. Fehige, C. Lumer, & U. Wessels (Eds.), *Handeln mit Bedeutung und Handeln mit Gewalt* (pp. 131–149). Paderborn, Germany: Mentis.

Stemmer, P. (2008). *Normativität* [Normativity]. Berlin, Germany: De Gruyter.

Stemmer, P. (2017). Moral, moralisches Müssen und Sanktionen [Morality, the moral "must," and sanctions]. *Deutsche Zeitschrift für Philosophie*, *65*, 621–656.

Stoutland, F. (2001). Responsive action and the belief–desire model. *Grazer Philosophische Studien*, *61*, 83–106.

Stroud, S. (2010). Permissible partiality, projects, and plural agency. In B. Feltham & J. Cottingham (Eds.), *Partiality and impartiality* (pp. 131–149). Oxford, England: Oxford University Press.

Trapp, R. W. (1998). The potentialities and limits of a rational justification of ethical norms. In C. Fehige & U. Wessels (Eds.), *Preferences* (pp. 327–360). Berlin, Germany: De Gruyter.

Vaughn, L. (Ed.). (2013). *Contemporary moral arguments* (2nd ed.). Oxford, England: Oxford University Press. (Original work published 2010)

von Wright, G. H. (1963). *The varieties of goodness*. London, England: Routledge.

Wessels, U. (2002). *Die gute Samariterin* [The good Samaritan]. Berlin, Germany: De Gruyter.

Wessels, U. (2011). *Das Gute* [The good]. Frankfurt/Main, Germany: Klostermann.

Woods, J. (2018). The authority of formality. *Oxford Studies in Metaethics*, *13*, 207–229.

Contributors

Rakefet Ackerman, Faculty of Industrial Engineering and Management, Technion—Israel Institute of Technology, Technion City, Haifa, Israel

Max Albert, Department of Economics, Justus Liebig University Gießen, Germany

J. McKenzie Alexander, Department of Philosophy, Logic and Scientific Method, London School of Economics, London, UK

Ali al-Nowaihi, Division of Economics, School of Business, University of Leicester, UK

Hanne Andersen, Department of Science Education, University of Copenhagen, Denmark

Line Edslev Andersen, Department of Mathematics—Science Studies, Aarhus University, Denmark

Johan van Benthem, Department of Philosophy, Stanford University, Stanford, USA; Department of Philosophy, Tsinghua University, Beijing, China; Institute for Logic, Language and Computation, University of Amsterdam, the Netherlands

Jean-François Bonnefon, Toulouse School of Economics & Centre National de Recherche Scientifique, France

Rainer Bromme, Institute of Psychology, University of Münster, Germany

John Broome, Faculty of Philosophy, University of Oxford, UK; School of Philosophy, Australian National University, Canberra, Australia

Anke Bueter, Department of Philosophy and History of Ideas, Aarhus University, Denmark

Ruth M. J. Byrne, School of Psychology and Institute of Neuroscience, Trinity College Dublin, University of Dublin, Ireland

Nick Chater, Behavioural Science Group, Warwick Business School, University of Warwick, UK

Peter Collins, School of Human Sciences, University of Greenwich, UK; Munich Center for Mathematical Philosophy, Ludwig Maximilian University Munich, Germany

Leda Cosmides, Department of Psychological & Brain Sciences, Center for Evolutionary Psychology, University of California, Santa Barbara, USA

Nicole Cruz, Department of Psychological Sciences, Birkbeck College, University of London, UK

Stephanie de Oliveira, Decision Lab, Department of Psychology, University of Michigan at Ann Arbor, USA

Sanjit Dhami, Division of Economics, School of Business, University of Leicester, UK

Franz Dietrich, Paris School of Economics & Centre National de Recherche Scientifique, France

Hans van Ditmarsch, Centre National de Recherche Scientifique, Laboratoire Lorrain de Recherche en Informatique et ses Applications, University of Lorraine, France

Didier Dubois, Emeritus Research Director, Centre National de Recherche Scientifique and Université Paul Sabatier, Institut de Recherche en Informatique de Toulouse, France

Shira Elqayam, School of Applied Social Sciences, Faculty of Health and Life Sciences, De Montfort University, Leicester, UK

Orlando Espino, Department of Cognitive, Social and Organizational Psychology, University of La Laguna, Tenerife, Spain

Jonathan St. B. T. Evans, Emeritus Professor of Psychology, University of Plymouth, UK

Christoph Fehige, Department of Philosophy, Saarland University, Saarbrücken, Germany

Klaus Fiedler, Department of Psychology, Heidelberg University, Germany

Lupita Estefania Gazzo Castañeda, Department of Psychology, Justus Liebig University Gießen, Germany

Lukas Gierth, Institute of Psychology, University of Münster, Germany

Andreas Glöckner, Department of Psychology, University of Cologne, Germany; Max Planck Institute for Research on Collective Goods, Bonn, Germany

Vinod Goel, Department of Psychology, York University, Toronto, Canada; Department of Psychology, Capital Normal University, Beijing, China

Till Grüne-Yanoff, Division of Philosophy, Royal Institute of Technology, Stockholm, Sweden

Rebecca Gutwald, Munich School of Philosophy, Munich, Germany

Ulrike Hahn, Birkbeck College, University of London, UK, Munich Center for Mathematical Philosophy, Ludwig Maximilian University Munich, Germany

Alan Hájek, School of Philosophy, Australian National University, Canberra, Australia

Stephan Hartmann, Munich Center for Mathematical Philosophy, Faculty of Philosophy, Philosophy of Science, and the Study of Religion, Ludwig Maximilian University Munich, Germany

Ralph Hertwig, Center for Adaptive Rationality, Max Planck Institute for Human Development, Berlin, Germany

Eric Hilgendorf, Department of Jurisprudence, Julius Maximilian University Würzburg, Germany

Brian Hill, Department of Economics and Decision Sciences, HEC Paris & Centre National de Recherche Scientifique, France

John Horty, Department of Philosophy, University of Maryland, College Park, USA

Mateja Jamnik, Department of Computer Science and Technology (Computer Laboratory), University of Cambridge, UK

P. N. Johnson-Laird, Department of Psychology, Princeton University, Princeton, NJ; Department of Psychology, New York University, USA

Gabriele Kern-Isberner, Department of Computer Science, Technical University Dortmund, Germany

Sangeet Khemlani, Navy Center for Applied Research in Artificial Intelligence, Naval Research Laboratory, Washington, DC, USA

Karl Christoph Klauer, Institute of Psychology, Albert Ludwig University Freiburg, Germany

Hartmut Kliemt, Department of Economics, Justus Liebig University Gießen, Germany

Markus Knauff, Department of Psychology, Justus Liebig University Gießen, Germany

Anastasia Kozyreva, Center for Adaptive Rationality, Max Planck Institute for Human Development, Berlin, Germany

Fenrong Liu, Department of Philosophy, Tsinghua University, Beijing, China; Institute for Logic, Language and Computation, University of Amsterdam, the Netherlands

Henry Markovits, Department of Psychology, University of Québec at Montréal, Canada

Ralf Mayrhofer, Department of Psychology, University of Göttingen, Germany

Linda McCaughey, Department of Psychology, Heidelberg University, Germany

Björn Meder, Health and Medical University Potsdam, Germany; iSearch and Center for Adaptive Rationality, Max Planck Institute for Human Development, Berlin, Germany

Georg Meggle, Institute for Philosophy, University of Leipzig, Germany

Arthur Merin †, Department of Philosophy, Philosophy of Science, and the Study of Religion, Munich Center for Mathematical Philosophy, Ludwig Maximilian University Munich, Germany

Julian Nida-Rümelin, Department of Philosophy, Philosophy of Science, and the Study of Religion, Ludwig Maximilian University Munich, Germany

Richard Nisbett, Institute for Social Research, Department of Psychology, University of Michigan at Ann Arbor, USA

Mike Oaksford, Department of Psychological Sciences, Birkbeck College, University of London, UK

Klaus Oberauer, Department of Psychology, University of Zurich, Switzerland

David P. O'Brien, Department of Psychology, Baruch College & the Graduate Center of the City University of New York, USA

David E. Over, Department of Psychology, Durham University, UK

Judea Pearl, Computer Science Department, University of California, Los Angeles, USA

Andrés Perea, Department of Quantitative Economics, Maastricht University, the Netherlands

Danielle Pessach, Institute for Psychology, Albert Ludwig University Freiburg, Germany

Martin Peterson, Department of Philosophy, Texas A&M University, College Station, USA

Niki Pfeifer, Department of Philosophy, University of Regensburg, Germany

Henri Prade, Emeritus Research Director, Centre National de Recherche Scientifique and Institut de Recherche en Informatique de Toulouse, Université Paul Sabatier, France

Johannes Prager, Department of Psychology, Heidelberg University, Germany

Henry Prakken, Department of Information and Computing Science, Utrecht University; Law Faculty, University of Groningen, the Netherlands

Marco Ragni, Department of Computer Science, Albert Ludwig University Freiburg, Germany; Department of Psychology, Justus Liebig University Gießen, Germany

Werner Raub, Department of Sociology/ICS, Utrecht University, the Netherlands

Hans Rott, Department of Philosophy, University of Regensburg, Germany

Olivier Roy, Faculty of Humanities and Social Sciences, University of Bayreuth, Germany; Munich Center for Mathematical Philosophy, Germany

Hanno Sauer, Department of Philosophy, Utrecht University, the Netherlands

Hans Bernhard Schmid, Department of Philosophy, University of Vienna, Austria

Gerhard Schurz, Düsseldorf Center for Logic and Philosophy of Science, Heinrich Heine University Düsseldorf, Germany

Niels Skovgaard-Olsen, Department of Psychology, University of Göttingen, Germany

Sonja Smets, Institute for Logic, Language and Computation, University of Amsterdam, the Netherlands; Department of Information Science and Media Studies, University of Bergen, Norway

Michael Smith, Department of Philosophy, Princeton University, NJ, USA

Kai Spiekermann, Department of Government, London School of Economics, UK

Wolfgang Spohn, Department of Philosophy, University of Konstanz, Germany; Eberhard Karl University, Tübingen, Germany

Julia Staffel, Department of Philosophy, University of Colorado at Boulder, USA

Keith E. Stanovich, Department of Applied Psychology and Human Development, University of Toronto, Canada

William B. Starr, Sage School of Philosophy, Cornell University, Ithaca, NY, USA

Florian Steinberger, Department of Philosophy, Birkbeck College, University of London, UK

Thomas Sturm, Catalan Institute for Research and Advanced Studies (ICREA), and Department of Philosophy, Autonomous University of Barcelona, Spain; Kantian Rationality Lab, Immanuel Kant Baltic Federal University, Kaliningrad, Russia

Valerie A. Thompson, College of Arts and Science, University of Saskatchewan, Canada

John Tooby, Department of Anthropology, Center for Evolutionary Psychology, University of California, Santa Barbara, USA

Maggie E. Toplak, Department of Psychology, Faculty of Health, LaMarsh Centre for Child and Youth Research, York University, Canada

Michael R. Waldmann, Department of Psychology, University of Göttingen, Germany

Ralph Wedgwood, School of Philosophy, University of Southern California, Los Angeles, USA

Ulla Wessels, Department of Philosophy, Saarland University, Saarbrücken, Germany

Richard F. West, Department of Graduate Psychology, James Madison University, Harrisonburg, VA, USA

Alex Wiegmann, Department of Psychology, University of Göttingen, Germany; Institute for Philosophy II, Ruhr University Bochum, Germany

John Woods, Department of Philosophy, University of British Columbia, Vancouver, Canada

Niina Zuber, Bavarian Research Institute for Digital Transformation, Munich, Germany

Name Index

Italic numbers indicate that the author is denoted by “et al.” or a similar phrase.

D

H

I

J

K

L

M

N

O

Q

R

T

U

V

Subject Index

Italic page numbers refer to definitions or explanations.

F

G

T

W

Z