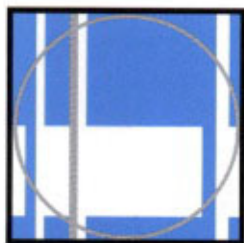




CAMBRIDGE
STUDIES IN
PHILOSOPHY



MAKING MORAL SENSE

BEYOND HABERMAS
AND GAUTHIER

LOGI GUNNARSSON

This page intentionally left blank

Is it rational to be moral? Can moral disputes be settled rationally? Which criteria determine what we have a good reason to do? In this innovative book, Logi Gunnarsson takes issue with the assumption made by many philosophers faced with the problem of reconciling moral norms with a scientific world view, namely that morality must be offered a non-moral justification based on a formal concept of rationality. He argues that the criteria for the rationality of an action are irreducibly substantive, rather than purely formal, and that assuming that morality must be given a non-moral justification amounts to a distortion of both rationality and morality. His discussion includes substantial critical engagement with major thinkers from two very different philosophical traditions, and is notable for its clear and succinct account of Habermas' discourse ethics. It will appeal to anyone interested in practical reason and the rational credentials of morality.

LOGI GUNNARSSON teaches philosophy at the Humboldt-Universität zu Berlin. Among his publications are *Wittgensteins Leiter: Betrachtungen zum Tractatus*, and a number of articles in journals including *Journal of Philosophical Research*, *Deutsche Zeitschrift für Philosophie* and *Dialektik*.

CAMBRIDGE STUDIES IN PHILOSOPHY

General editor ERNEST SOSA

Advisory editors

JONATHAN DANCY *University of Reading*
JOHN HALDANE *University of St. Andrews*
GILBERT HARMAN *Princeton University*
FRANK JACKSON *Australian National University*
WILLIAM G. LYCAN *University of North Carolina, Chapel Hill*
SYDNEY SHOEMAKER *Cornell University*
JUDITH J. THOMSON *Massachusetts Institute of Technology*

RECENT TITLES

PAUL HELM *Belief policies*
NOAH LEMOS *Intrinsic value*
LYNNE RUDDER BAKER *Explaining attitudes*
HENRY S. RICHARDSON *Practical reasoning about final ends*
ROBERT A. WILSON *Cartesian psychology and physical minds*
BARRY MAUND *Colours*
MICHAEL DEVITT *Coming to our senses*
SYDNEY SHOEMAKER *The first-person perspective and other essays*
MICHAEL STOCKER *Valuing emotions*
ARDA DENKEL *Object and property*
E. J. LOWE *Subjects of experience*
NORTON NELKIN *Consciousness and the origins of thought*
PIERRE JACOB *What minds can do*
ANDRE GALLOIS *The world without, the mind within*
D. M. ARMSTRONG *A world of states of affairs*
DAVID COCKBURN *Other times*
MARK LANCE & JOHN O'LEARY-HAWTHORNE *The grammar of meaning*
ANNETTE BARNES *Seeing through self-deception*
DAVID LEWIS *Papers in metaphysics and epistemology*
MICHAEL BRATMAN *Faces of intention*
DAVID LEWIS *Papers in ethics and social philosophy*
MARK ROWLANDS *The body in mind*

MAKING MORAL SENSE

Beyond Habermas and Gauthier

LOGI GUNNARSSON
Humboldt-Universität zu Berlin



CAMBRIDGE
UNIVERSITY PRESS

PUBLISHED BY THE PRESS SYNDICATE OF THE UNIVERSITY OF CAMBRIDGE
The Pitt Building, Trumpington Street, Cambridge, United Kingdom

CAMBRIDGE UNIVERSITY PRESS

The Edinburgh Building, Cambridge CB2 2RU, UK
40 West 20th Street, New York, NY 10011-4211, USA
477 Williamstown Road, Port Melbourne, VIC 3207, Australia
Ruiz de Alarcón 13, 28014 Madrid, Spain
Dock House, The Waterfront, Cape Town 8001, South Africa

<http://www.cambridge.org>

© Logi Gunnarsson 2004

First published in printed format 2000

ISBN 0-511-03481-4 eBook (Adobe Reader)

ISBN 0-521-78023-3 hardback

*For my parents,
Gunnar Ólafs and Ingibjörg Lúðvíksdóttir*

Contents

<i>Preface</i>	<i>page xi</i>
PART I: PROBLEMS	1
1 The justificatory crisis of morality	3
2 Alternative resolutions of the justificatory crisis	9
3 Subjective reasons	23
4 Substantive reasons	33
5 Overcoming rationalism	43
PART II: AGAINST RATIONALISM	51
6 Gauthier's contractarianism	53
7 From here to pre-social agreement	71
8 Habermas' discourse ethics	86
9 Discoursing about discourse	109
PART III: FOR THE SUBSTANTIVE APPROACH	127
10 Self-understanding and self-assessment	129
11 The possibility of progress	152
12 Practical arguments vs. impossibility arguments	171
13 Evaluation of others	197
14 Universality without neutrality	215

Contents

PART IV: FOR PARTICULARIST SUBSTANTIVISM	227
15 Against formalism	229
16 Particularist substantivism vs. list-substantivism	247
Appendix Transcendental vs. universal pragmatics	262
<i>Bibliography</i>	266
<i>Index</i>	279

Preface

Das Schwere ist hier, nicht bis auf den Grund zu graben, sondern den Grund, der vor uns liegt, als Grund zu erkennen.

The difficult thing here is not to dig down to the ground; no, it is to recognize the ground that lies before us as the ground.

(Wittgenstein, *Bemerkungen über die Grundlagen der Mathematik*, VI.31. Trans. G. E. M. Anscombe.)

When it comes to the justification of moral views, philosophers tend to think that they need to “dig down to the ground”, to find a secure non-moral foundation for morality to stand on. This is a fundamental mistake. The substantive reasons we have regarding morality lie before us; the difficulty is to recognize them as reasons which do not need a non-moral foundation. The task of this book is to help us recognize the reasons that lie before us as reasons.

I have worked on the issues in this book in two main stages. My attempts to formulate my ideas found a preliminary ending with the submission of my dissertation at the University of Pittsburgh in 1995. Since then I have been thoroughly rethinking my theses and arguments and this work presents my current thoughts.

In writing this book, I have benefited from the generous advice of many people. My greatest debt is to John McDowell. Not only is this work deeply influenced by his writings, but as my dissertation advisor he was an invaluable source of inspiration and criticism in our numerous enjoyable philosophical conversations. Annette Baier, Robert Brandom, David Gauthier and Nicholas Rescher were also members of my dissertation committee and I am grateful to all of them for reading my work so carefully and for the many helpful discussions we had. Since Gauthier’s work is under criticism in this book, he deserves special credit for invariably receiving my arguments

in the spirit that philosophy lives from criticism. Although James Conant was not on my committee, we discussed my work on many occasions, and I owe him gratitude for those fruitful conversations.

Jürgen Habermas I thank for responding fairly and vigorously to my criticisms of his theory in the discussions we had during my stays at the J.W. Goethe-Universität in Frankfurt (1989–1990 and the winter term of 1992–1993). He also kindly invited me to join his doctoral colloquium and sponsored my research fellowship from Deutscher Akademischer Austauschdienst, which I thank for its support.

Many other people read my work during the years 1992–1995. I have benefited not only from extensive discussions with Joel Anderson, Felmon Davis and Matthias Kettner but also from the comments of these people: Bruce Basara, Donald Bruckner, Raymond Geuss, Bennett Helm, Friedrich Kambartel, Angelika Krebs, Hans-Peter Krüger, Jonathan Mandle, Eric Marcus, Jennifer Whiting, Lutz Wingert, and Iris Young.

The second stage in the development of the ideas in this book began in 1996 with a grant I received from the Icelandic Research Council to work on this project. This grant was renewed for 1997 and I am very grateful to the Council for its support. Since 1996, I have profited from trying my ideas on audiences in many different places. In particular, I want to mention the insightful comments given by Neera K. Badhwar and Bernard Gert, who were my commentators at the 1997 Pacific and Eastern meetings of the American Philosophical Association, respectively.

Of the people who have read, criticized and commented on my work in the last few years, I owe special gratitude to two persons. From the time I joined the Humboldt-Universität in 1997, Jay Wallace has been an especially inspiring and challenging philosophical interlocutor. His penetrating criticisms and resourceful constructive comments have led me to reformulate my arguments and ideas in many places. Mikael M. Karlsson was my teacher at the University of Iceland and has been a philosophical companion since then. Here I not only want to thank him for being an endless source of ideas and objections over the years, but also for his acute and rich constructive criticism of recent drafts of this book.

Since 1996, I have also been helped greatly by comments from the following persons: Röbert Haraldsson, Ulrike Heuer, Jeffrey

Preface

Honnold, Kristján Kristjánsson, Martin Löw-Beer, Katja Vogt, and two anonymous readers for Cambridge University Press. My editor at Cambridge, Hilary Gaskin, I thank for showing interest in the project in the first place and for her professional and competent advice along the way. Likewise, I am grateful to the copy-editor, Leigh Mueller, and the production controller, Caroline Murray, for their careful editorial work.

Material in chapters 8–9 is based on Gunnarsson 1995a and material in chapter 12 on Gunnarsson 1997a. It appears here with the publishers' kind permission.

Although this book is done, I am afraid that I won't be able to stop working on the issues in it. One might say that the book is an attempt to employ practical reasons to recommend a certain way of thinking. About this pragmatic bend and other aspects of this work, I have – ever since I started talking about these topics – been engaged in a continuing dialogue with Eva Klingenstein and I do not expect it to stop. I thank her for reading my material and for her skeptical enthusiasm.

PART I

Problems

1

The justificatory crisis of morality

We are honest and truthful, we pay our debts and keep our promises. We are caring and concerned, yet impartial and just. We are sensitive, friendly, merciful, forgiving, generous, thankful, loyal, and self-sacrificing. We are politically conscious and active, and we are respectful of people's rights whatever their gender, race, or sexual orientation. And lately we have even started recycling. In short, we are just great.

Unfortunately, accompanying this feeling of greatness is the nagging worry that we are simply being stupid. The fear is that the very source of our pride is actually a sign of our stupidity: that being moral is, in the final analysis, fundamentally irrational.

There are plenty of reasons to suspect that we are indeed being irrational. I will mention three. The most obvious reason is that being moral often requires us to sacrifice our interests or to act against our desires. We keep our promise to meet somebody for dinner even though we would much rather do something else. We divide the cake fairly though we want all of it, and we even save our enemies while rather wanting to see them dead. Now if being moral requires us systematically to act against our desires in this way, how can it be rational?

The second reason for being suspicious about morality does not as such have anything to do with a possible conflict with the satisfaction of desire. It depends on the obvious fact that the morally evaluative vocabularies which we use to guide our lives represent only one possible way of evaluating. Other moral or non-moral evaluative vocabularies would lead us to evaluate our lives quite differently. For example, instead of striving to treat others fairly and congenially, a person could set it as her ideal to treat them ruthlessly or indifferently, or she could give herself high marks for being cool rather than concerned, or original and independent rather than loyal and

thankful. Given the obvious possibility of conflict between these different ways of evaluating, it is by no means clear that it is rational to let our current moral evaluative scheme dominate our lives or to use it at all.

The third doubt concerning our self-satisfaction about our moral virtues is rather different. Here the question is not whether in acting morally we are doing what is rational *for us*. We do not employ a moral vocabulary only to guide our own lives, but also to criticize others. Here the worry is that our criticism of others does not amount to rational criticism but that it is rather a way of exercising power over others under the guise of moral comment. In other words, in morally criticizing others, we are not interacting with them rationally but rather abusing them. Underlying this worry is the question whether it can ever be rationally settled who is right: we or they. If it cannot be rationally settled, then our criticism can only be abuse in disguise.

This problem becomes particularly pressing when the criticizer and the criticized are members of two radically different cultural communities. It could be argued that the critic inevitably relies upon the practices of her community and that she can only be shown to be right if these practices are rationally superior to the practices of the other community. However, the argument goes, it is impossible to show the practices of one of two radically different cultural communities to be more rational than the other. I do not think that this argument is good or that the problem is insoluble, but it is a problem which needs to be resolved before we may assume that our criticism of other cultural practices can be a piece of *rational* criticism.

These worries all present a problem about *the rationality of morality*. For the sake of convenience, this problem may be divided into two fundamental subproblems: (1) *The basic choice problem*: is it rational to be guided by moral considerations at all? (2) *The moral alternatives problem*: is it rational to be guided by one particular moral view as opposed to others? These are the two main problems which I shall discuss in this work.¹ Notice that it is certainly possible to answer

¹ Although I shall also discuss other problems, I use the distinction between these two problems to structure my discussion. A third subproblem should be mentioned here. This is *the problem of priority*: is it rational to give moral reasons priority over other reasons? (Cf. Scanlon 1998, 148.) This problem must be distinguished from the basic choice problem. Even if it is rational to take moral considerations into account in rational deliberation, it still needs to be asked whether moral reasons can be overridden

only the first question positively. In that case, one would suppose that it is rational to be guided by some moral view or other, but think that the choice among different moral perspectives is not a matter of rationality.

It is extremely tempting to think that the only possible solution of these problems is to offer a *non-moral* justification of morality. Such a justification would demonstrate the rationality of morality on entirely non-moral premises. This is tempting because it seems that any other kind of justification would be question-begging and would not have the necessary independence from morality to provide criteria for deciding which moral view is the most rational.

One central thesis of this work is that it is entirely misguided to think that morality needs a non-moral justification. This thesis distinguishes the work from the writings of both the friends and the foes of non-moral justifications of morality. The former are busy constructing such justifications, whereas the latter occupy themselves with tearing them down or with giving *a priori* arguments to the effect that such justifications are bound to fail. Thus, even the foes of non-moral justifications seldom call into question the assumption that morality would be unjustified if such a justification cannot be given. This, however, is precisely the assumption which I want to call into question. I shall argue that *even if* there are flawless non-moral justifications of morality, it is a mistake to think that morality *needs* such a justification. In fact, I argue that to proceed on the assumption that morality needs such a justification distorts our view of rationality, morality, and the relationship between the two. Thus, it is not my aim to argue that non-moral justifications are impossible, but rather that – even if possible – they are not an ideal against which the success of justifications of morality and moral views should be measured.²

One powerful motivation for non-moral justifications of morality is at the same time a reason for thinking that these justifications must be purely *formal*. The thought here is that doubts about the rationality of morality arise precisely because moral thinking relies heavily upon *substantive* intuitions. For example, actions are taken to be morally

by other reasons, and if they can, how it is to be decided when they are overridden. This issue of overridingness is the problem of priority.

² This means that *rational* justifications of morality and moral views neither are to be equated with nor need to be supported by *non-moral* justifications.

wrong because they are *cruel* or right because they are *considerate*. However, so the argument goes, it can always be asked whether it is rational to guide one's life by such substantive considerations. And in order to show this to be rational, it won't help to appeal to other substantive considerations. The problem is not that these considerations are moral but that they are substantive. No actions are rational or irrational on account of some substantive features but rather on account of formal ones. Thus, in order to solve the justificatory crisis of morality, it is not enough to offer a non-moral justification of morality. The justification must also be purely formal.

According to this view, morality needs a formal non-moral justification. I call a theory "*rationalistic*" if it aims to deliver such a justification. In this book, rationalism will be my main target of criticism. As an alternative to it, I present another justificatory ideal which violates not only the rationalistic requirement that the justification of moral views must be formal but also that it must be non-moral: I argue that a justification of moral outlooks based on *substantive* reasons which *cannot be purified of moral content* is an adequate justification and is preferable to a rationalistic justification.³

Although it is widely assumed that morality needs a rationalistic justification, only a few philosophers actually offer a *purely* rationalistic justification. The works of these philosophers will be the focus of the argument that my substantive approach should be favored over rationalism. If I did not undermine the actually existing rationalistic positions, my argument would remain unconvincing. After criticizing these few, selected positions, I then go on to explain why I think that other rationalisms have the same flaw. In this way, I hope to deliver arguments which are convincing in their specificity while at the same time indicating how they have a general application.

There are two basic kinds of rationalism, depending upon whether the concept of rationality employed is "Hobbesian" or "Kantian". I

³ It should be noted that the alternative to rationalism that I offer is also to be contrasted with theories which attempt to justify morality from a *substantive, non-moral* starting point. In favoring my alternative, I shall be defending the idea of giving justifications which have neither a non-moral nor a formal starting point. In other words, my approach is to be contrasted with the idea of giving a justification of the ethical life from an Archimedean point outside it, whether that point is understood in terms of a substantive notion of well-being or a formal notion of practical reason (see Williams 1985, chs. 2–4). In chapter 5, section 2 (hereafter 5.2), I explain how I plan to deal with theories which offer substantive non-moral justifications.

will focus on the work of the two contemporary philosophers who have perhaps done the most in recent years to develop these two conceptions of reason: David Gauthier and Jürgen Habermas. Gauthier's contractarianism is an impressive attempt to provide a rigorous Hobbesian justification of morality with the help of the tools of rational choice theory; while with his theory of communicative reason, Habermas has surely made one of the most important contributions to the development of a Kantian concept of reason in recent decades.⁴

Because of the deep differences between these two thinkers, and because Gauthier tends to be studied by "analytic" philosophers and Habermas by "Continental" thinkers, the common rationalistic core of their theories has been overlooked. It is sometimes noted in the literature that they are both, broadly speaking, contractarians.⁵ However, this book does not criticize them as contractarians. For this reason, I shall not discuss at any length the theory of the other, perhaps most prominent, contemporary defender of a Kantian approach to moral and political theory – namely John Rawls. Rawls is a contractarian and a Kantian, but he is not, in my sense, a rationalist.

To see that Rawls – as opposed to Gauthier and Habermas – is not a rationalist, we need only to consider briefly the attitude of these thinkers to "reflective equilibrium" justifications. Contrary to Rawls, Gauthier and Habermas both distance themselves from the idea of a reflective equilibrium as the ultimate justification of moral and political norms.⁶ Roughly speaking, a moral judgment has been given a reflective equilibrium justification if it has been shown that this judgment is in reflective equilibrium with our moral principles and considered moral judgments. A state of reflective equilibrium has been reached if the process of modifying our moral principles in the

⁴ In this work, I shall only be concerned with contemporary versions of rationalism. Another recent work which explicitly defends a Hobbesian rationalism is Danielson 1992. Different kinds of Kantian rationalism are offered in Apel 1973; 1988c; Gewirth 1977; Kuhlmann 1985; and Korsgaard 1996.

⁵ See Heath 1995, 80–82.

⁶ Habermas 1988d, 89 [78–79]; 1988e, 127 [116]; Gauthier 1986, 5, 269. Rawls, in contrast, is happy to see it as the ultimate justification; see Rawls 1993, 28, 51–53. I briefly compare Rawls and Gauthier in 6.2, and Rawls and Habermas in 8.2. (In citing texts which appear in my bibliography under their original German title, I first give the reference to the German text and then, in square brackets, to an English translation.)

light of our considered moral judgments and vice versa has been completed in the sense that no further adjustments seem proper.⁷ The point to notice here is that reflective equilibrium justifications remain firmly *within* morality: moral principles are justified in terms of other *moral* principles and considered *moral* judgments. This immediately raises doubts as to whether such justifications can meet skeptical worries about the rationality of morality. According to rationalism, in order to dissolve these skeptical worries, one must give a justification of morality which – contrary to reflective equilibrium justifications – does not rely on any moral intuitions.

The appeal of rationalism is obvious. To appeal to moral intuitions to demonstrate the rationality of morality seems viciously circular. To rely on other substantive intuitions seems just as hopeless, since it seems that the rationality of following such intuitions can always be called into question. And, in contrast to scientific theses, there seems to exist no empirical confirmation of moral principles.⁸ Thus, it seems that the only possible savior of morality would be a *formal non-moral* justification. It is the task of this work to undermine this rationalistic justificatory ideal and to replace it by my substantive approach.

In the next chapter, I shall give a fuller and more precise account of rationalism and sketch my own alternative to it.

⁷ For a more detailed discussion of reflective equilibrium justifications, see chapter 15.

⁸ According to Alan Gewirth, empirical facts serve to test the correctness of the factual statements of natural science. These empirical facts are an “independent variable” that serves to determine the correctness of factual statements. Gewirth believes that such an “independent variable” seems – on the face of it – to be missing in the case of moral statements and that in the absence of such an “independent variable” no answer can be given to moral skepticism. His rationalism is supposed to solve this problem by demonstrating the existence of an “independent variable” for the case of morality (without assuming any metaphysically suspect moral facts or assimilating morality to natural science) (Gewirth 1977, 4–9, 78, 175–177, 365).

2

Alternative resolutions of the justificatory crisis

I SUBJECTIVIST RATIONALISM

It is Gauthier's declared aim in *Morals by Agreement* to argue that "[m]orality . . . can be generated as a rational constraint from the non-moral premisses of rational choice."¹ One way of interpreting the project of starting from non-moral premises – and this is indeed how Gauthier understood it in this work – is that the goal is to show that "agents lacking all moral concerns . . . would rationally introduce morality into their interactions in order better to achieve their nonmoral ends."² In his more recent article "Value, Reasons, and the Sense of Justice," Gauthier has outlined another justification that can also be understood as relying only on non-moral premises. There, the idea is not to show that moral sensibility – or, more specifically, the sense of justice which is the focus of Gauthier's discussion in this article – is a "mere instrument for our nonmoral gratification."³ Rather, the aim is to show that the sense of justice is of value to agents "whatever their particular aims and concerns."⁴ It is on account of this idea, as will be explained, that I take Gauthier to be a rationalist. This idea can be captured by saying that "justice is a necessary instrumental value."⁵ To show justice to be a *necessary instrumental* value is, in my terminology, to give a *subjectivist rationalistic*

¹ Gauthier 1986, 4.

² Gauthier 1993a, 201.

³ Gauthier 1993a, 201.

⁴ Gauthier 1993a, 199. This claim must be qualified. For example, it does not hold for "an agent whose life-plan is focused on the destruction of his fellows, who lives to kill." Strictly speaking, it holds only for "those persons whose overarching life-plans make them welcome participants in society" (Gauthier 1993b, 188, 189) (see 6.3).

⁵ Gauthier 1993a, 199.

justification of justice.⁶ The subjectivism is reflected in the instrumentality of the value and the rationalism in the necessity.

What does it mean to say that something is of necessary instrumental value? To say that something is of *instrumental* value is to say that it is valuable as a means to something else that is valuable. This is where Gauthier's subjectivism surfaces. Practical reason is strictly instrumental: it is silent on which ends we should have and can only tell us how best to pursue our ends, where these ends are taken as subjectively given.⁷ To show that something is of *necessary* instrumental value is to show that it is valuable – in the instrumental sense – whatever our ends may happen to be.⁸

Gauthier's justification of morality is thus *formal* in two senses. First, reason is understood instrumentally and it is thus silent on which ends we should pursue. Second, morality is supposed to be rational for the agent no matter what the substantive contents of her goals are.

By showing that morality is of necessary instrumental value, Gauthier wants to solve two problems he sees morality confronted with. The first problem is a variation on the problem of the *rationality of morality* which I mentioned in the last chapter. For Gauthier, this problem takes the following form: since for him instrumental rationality is the only notion of practical reason there is, morality cannot survive a conflict with the deliverances of instrumental reason. However, according to Gauthier, in order to show that it is rational for a person to be moral, it is not enough to show that she must be moral in order to achieve the (moral or non-moral) ends that she happens to have. Gauthier wants to be able to say that actions may be irrational even if they are the best fulfillment of the ends that the agent *happens to have*. Having those ends – for example, to be kind to one's fellow humans – may stand in the way of the person's reaping some benefits which she might otherwise be able to enjoy. Now since instrumental reason is incapable of evaluating the agent's ends

⁶ I do not assume that Gauthier thinks that we can, or need to, show that all of what we ordinarily think of as morality can be shown to be of necessary instrumental value. In showing in *Morals by Agreement* that the rational constraints on actions are moral constraints, his concern is really with showing that these constraints are *just*. The principle of interaction justified in that work is a principle of justice (Gauthier 1986, 6, 150–156, 208–223, 233–267).

⁷ Gauthier 1986, 24–26, 46–55.

⁸ Gauthier 1993a, 198–199.

directly, morality cannot be shown to be rational by establishing that it helps the agent to fulfill certain rationally privileged ends. Thus, the only possible way of demonstrating the rationality of morality consists in establishing that it is instrumentally rational to be moral *whatever* the agent's ends are, i.e., in showing that morality is of *necessary* instrumental value.⁹

The second problem Gauthier wants to solve concerns the “*categorical force*” or “*unconditionality*” of morality. He takes morality as presenting us with unconditional demands because “[f]rom the standpoint of the agent, moral considerations present themselves as constraining his choices and actions, in ways independent of his desires, aims, and interests.”¹⁰ This does not just mean that moral requirements sometimes conflict with our self-interest. According to Gauthier, morality has a “*prescriptive grip*” which cannot be explained entirely (as a Humean might think) in terms of our sympathetic feelings, since morality speaks to those “*hard cases*” where even our sympathetic feelings would not move us to act in accordance with what morality demands of us. Morality operates somehow independently of our affections, including our sympathetic concern for the well-being of our fellows.¹¹ The problem is that instrumental reason seems – at first sight – to be unable to deliver morality's unconditional demands.¹² Since what is instrumentally rational for an agent depends on her contingently given ends, it seems that unconditional demands can never be shown to be instrumentally justified. By showing that morality is of necessary instrumental value, Gauthier would solve this problem: if morality is indeed of necessary instrumental value, it is rational to be moral not just if one happens to have certain goals but whatever one's goals are.¹³

Before defining rationalism, a misunderstanding of Gauthier's claim that morality is of necessary instrumental value must be dismissed. It might be thought that Gauthier's point is simply that it is in the long-term interest of the straightforward instrumental reasoner

⁹ Gauthier 1986, 11; 1988b, 386–389; 1991a, 18–25; 1993a, 180–183, 189, 197–204. For a more elaborate discussion of this point, see 6.2.

¹⁰ Gauthier 1991a, 16.

¹¹ Gauthier 1991a, 17–18.

¹² This problem does not coincide with the problem of the rationality of morality. One surely does not need to assume that morality speaks to us in unconditional demands in order to question the rationality of morality.

¹³ Gauthier 1991a, 20–25, 29–30.

to behave morally. This would not mean that the instrumental reasoner reasons morally. Rather, each time this instrumental reasoner has to decide what to do, the question comes up whether she should follow the moral course of action, and each time her reasoning will be instrumental: even though it is generally in her interest to behave morally, it may sometimes be in her long-term interest to act immorally, and then she has a good instrumental reason to do so. In other words, in the case of such a conflict between morality and instrumental reasoning, instrumental reasoning is overriding.

Gauthier, however, does not want to show that morality is of necessary instrumental value only in the way just described. Rather, he wants to show that there is an instrumental rationale for ceasing to reason exclusively instrumentally and starting to reason *morally* as well. The agent should cease to be a straightforward instrumental reasoner and become a *constrained* instrumental reasoner: more specifically, an instrumental reasoner constrained by morality. This means that in order to show that one should do something it is not enough to show that instrumental reasoning would tell one that it is in one's (long-term) interest to do that. There exists an instrumental rationale for reasoning morally – as opposed to exclusively instrumentally – and if moral reasoning were to tell one not to do a particular action, it might be that one would have an overriding reason not to act in this way. Morality (that is, the morality for which there is an instrumental rationale) is a direct source of reasons for the morally constrained instrumental reasoner, whereas this is not so for the straightforward instrumental reasoner.¹⁴

What exactly makes Gauthier's theory rationalistic? I understand rationalism as a theory which addresses itself to a moral skeptic and aims to refute the skeptic on the skeptic's own terms. This moral skeptic has a purely formal understanding of rationality and wants to be rational in this sense. She thinks, however, that being rational in

¹⁴ Gauthier 1986, 167–170; 1993a, 197–199; 1993b, 185–191; 1996, 20–28. There is a second way in which Gauthier's claim could be misinterpreted. One might suppose that he wants to show that there is an instrumental rationale for becoming a morally constrained instrumental reasoner, where this is understood as an instrumental rationale for becoming irrational and making irrational choices. This is not Gauthier's understanding of it. He thinks that the morally constrained instrumental reasoner is rational and that the choices that she makes based on morally constrained instrumental reasoning are themselves rational. See Gauthier 1986, 184–187; 1994.

this way does not require her to accept any moral norms.¹⁵ In Gauthier's case, this means that the skeptic is a person who aims to be a fully rational instrumental reasoner and thinks that being rational in this way does not require her to start to reason morally. Gauthier's theory is rationalistic because it is meant to establish that the formal rationality which the skeptic herself accepts requires her to respect certain moral principles. I accordingly define rationalism as follows:

A theory is rationalistic if and only if it (1) addresses itself to a moral skeptic who has a formal understanding of rationality and (2) aims to show that this moral skeptic cannot be rational in this sense unless she respects certain specific moral norms.

As I already mentioned, in trying to show that the skeptic cannot be formally rational unless she accepts certain moral norms, rationalism is aiming to refute the skeptic *on the skeptic's own terms*. A distinction of Robert Nozick's can be usefully employed to illustrate this point. He distinguishes between the foreign and domestic relations of a subject's system of beliefs. According to Nozick, one way in which a subject could respond to a skeptic who questions the possibility of knowledge is to try to *convince* the skeptic by showing that in the light of some of the skeptic's other beliefs the skeptic herself must think that knowledge is possible. This is a task of the foreign relations department of the subject's belief system, since the topic is how *the skeptic's* beliefs fit together. The bureau of internal affairs has another job. The subject has accepted some of the things that the skeptic points out and this presents the subject with a problem as to how knowledge is possible. The goal is not to convince the skeptic, but rather to explain how knowledge is possible in light of the problem presented. Since the skeptic may not accept all the statements on which the explanation depends, the explanation is a matter for the subject's domestic relations department.¹⁶ Applying this terminology to the present case, we can say that rationalism is engaged in foreign relations. Its goal is not simply to show that the

¹⁵ It should be noted that this is a very special form of moral skepticism. For example, my skeptic has a purely formal understanding of rationality. Other moral skeptics might not. When I speak of "*the moral skeptic*" in this work, I have in mind the kind of skeptic described here, without assuming that there are no other forms of moral skepticism. I focus on this kind of moral skepticism because it is the form which rationalism speaks to.

¹⁶ Nozick 1981, 15–17.

moral skeptic is irrational, but to show that she is irrational even by her own standards.

Attention should be drawn to the fact that rationalism, as I have defined it, is committed to showing that it is rational to respect *certain specific* moral norms. This does not mean that rationalists such as Gauthier and Habermas agree on what norms it is rational to accept. It just means that all rationalists aim to justify *some* specific moral norms. If rationalism were to show successfully that a rational skeptic must indeed accept certain specific norms, then rationalism would have solved the problem of the rationality of morality in both of its versions: both the *moral alternatives problem* (the problem *which* morality it is rational to accept) and the *basic choice problem* (the problem whether it is rational to be moral *at all*). My aim is to show that it is not necessary to refute the moral skeptic on her own terms to solve either of these problems (see 11.4).

2 INTER-SUBJECTIVIST RATIONALISM

I understand Habermas as a rationalist because he offers a transcendental-pragmatic or universal-pragmatic justification of certain moral norms.¹⁷ Let me thus start by explaining briefly what it means to give such a justification.

Transcendental-pragmatic or universal-pragmatic justifications (for short, “up-justifications”) are *transcendental* in the sense that they work by asking what makes a certain activity, e.g. doubt, possible. The idea is that by answering this question we will discover what makes doubt possible in the first place and will thereby discover what cannot be doubted itself. The sense in which an up-justification is

¹⁷ Habermas is not the only philosopher to offer such justifications. For example, Karl-Otto Apel and Wolfgang Kuhlmann also do. Habermas prefers to call his justifications “universal-pragmatic” rather than “transcendental-pragmatic” in order to stress that he – contrary to Apel and Kuhlmann – does not understand them to be delivering *a priori* knowledge but rather empirical-philosophical knowledge of the presupposition of our current concept of reason, a concept that might itself change (Habermas 1984c, 379–385 [21–25]; 1988d, 104–108 [94–98]; 1991f, 190–195 [80–84]). Despite this difference, the justifications offered by all three philosophers are rationalistic: they all offer a non-moral formal justification of morality. Thus, though I will not discuss Apel and Kuhlmann explicitly in the text, in the discussion of the views which I think they share with Habermas I also include footnote-references to their work. For a longer discussion of the central differences between them relevant to our topic, see the Appendix.

pragmatic is best explained by considering a notion central to it, the notion of *pragmatic (or performative) self-contradiction*. Consider an example of a person who asserts “I do not exist.” One could argue that this person has by the very act of asserting this involved herself in a pragmatic self-contradiction. If her assertion is to be successful, certain presuppositions must be fulfilled. If these presuppositions are not fulfilled she will not have succeeded in asserting anything. Presumably, one of the presuppositions is that she exists. Now, this presupposition of the assertion contradicts the propositional content of the assertion (“I do not exist”). In other words, this is a case of a *pragmatic self-contradiction* for the following reason: the contradiction is between a presupposition of an assertion as a performance or action (the assertion in its pragmatic sense) and the propositional content of the assertion. This pragmatic self-contradiction shows that one cannot coherently doubt one’s own existence.¹⁸

Up-justifications of moral norms are just one instance of this general form of justification. The goal is to show that a radical moral skeptic – a skeptic who asserts “There is nothing irrational about rejecting moral norms” (or something equally radical) – gets caught in a pragmatic contradiction, since there is a contradiction between the propositional content of this assertion and the conditions of possibility of the assertion considered as a contribution to a rational argumentation.

No contribution to rational argumentation may contradict the conditions of the possibility of rational argumentation. According to Habermas, one such condition is that nobody be excluded from the argumentation. If somebody were excluded from participating, then the form of interaction between the participants would not be rational argumentation, properly speaking. In fact, it would not be argumentation at all. The implicit end of argumentation is to resolve rationally some question which is under discussion. To exclude somebody from the discussion is a threat to the rationality of the answer. Thus, the assertion “Smith should be excluded from the argumentation” cannot count as a contribution to a rational argumentation, since it serves to undermine the very conditions that make rational argumentation possible (the presuppositions of rational argumentation): there is a contradiction between the propositional content of this assertion and

¹⁸ Habermas 1988d, 90–91 [79–80]; Apel 1975, 262–269; Kuhlmann 1985, 88–89. It should be noted that all of these authors give credit to Hintikka 1962.

the conditions that must be fulfilled if it (as performance) is to count as a contribution to rational argumentation.

Habermas would argue that the moral skeptic's assertion of "There is nothing irrational about rejecting moral norms" is similarly problematic. He believes that among the presuppositions of rational argumentation are some norms with moral content. Thus, as a participant in a rational argumentation about the question "Is there anything irrational about rejecting moral norms?" the moral skeptic (who thinks that the answer is negative) must presuppose that certain moral norms are being respected. If they are not respected – given that they are presuppositions of rational argumentation – the discussion of this question cannot count as rational argumentation. This means that these moral norms cannot be rationally rejected, since they must be respected if there is to be any such thing as rational argumentation. The impossibility of rationally rejecting these norms is crystallized in the pragmatic contradiction between the propositional content of the skeptical assertion "There is nothing irrational about rejecting moral norms" and the presuppositions of this assertion considered as a contribution to rational argumentation (where these presuppositions include the one that certain moral norms cannot be rationally rejected).¹⁹

If this kind of up-justification of moral norms is successful, it seems to solve the problem of the *rationality* of morality.²⁰ It seems that since it is the rationality of morality itself that is at stake, appeal to moral intuitions will beg the question against the moral skeptic. Up-justifications are constructed so as not to depend on the acceptance of any moral intuitions. Up-justifications of moral norms amount to an analysis of the presuppositions of *rational argumentation as such* rather than the presuppositions of moral argumentation: the skeptic does not engage in moral argumentation but only in "meta-moral" argumentation about the question whether it is irrational to reject moral norms. Since the up-justifications rely on an analysis of the presuppo-

¹⁹ Habermas 1988d, 96–103 [86–93]; Apel 1988c, 352–357; Kuhlmann 1985, 181–215.

²⁰ I mention here only the problem of the rationality of morality, since it seems to me that Habermas does not want to account for the unconditionality of morality in a rationalistic way (Habermas 1991f, 132–137, 186–192 [31–35, 77–81]). Apel and Kuhlmann, in contrast, do want to give a rationalistic account of the unconditionality of morality; see Apel 1973, 415–417 [270–271]; Kuhlmann 1992b, 154–157; 1985, 227–239.