Preferences: An Introduction and a Short Bibliography
Preferences

Edited by
Christoph Fehige and Ulla Wessels

Walter de Gruyter · Berlin · New York
1998
All I want is to sit on my arse and fart and think of Dante.

Samuel Beckett

Desires and wants, however intense, are not by themselves reasons in matters of justice. The fact that we have a compelling desire does not argue for the propriety of its satisfaction any more than the strength of a conviction argues for its truth.

John Rawls

Take any demand, however slight, which any creature, however weak, may make. Ought it not, for its own sole sake, to be satisfied? If not, prove why not. The only possible kind of proof you could adduce would be the exhibition of another creature who should make a demand that ran the other way. The only possible reason there can be why any phenomenon ought to exist is that such a phenomenon actually is desired.

William James
Preface

*Preferences* is a collection of essays on the concept and the role of preferences (desires, and the like) in practical reasoning. Ground covered includes welfare, prudence, rational decision making, and all areas of moral philosophy: ethics (applied and not so applied), metaethics, and deontic logic. A special symposium looks at possible preferences and their significance in matters of life and death, including the notoriously thorny question how many people there should be. All the essays are published here for the first time.

The book is not just for specialists. We have given it an introduction that, though it may move swiftly, at least starts from scratch; a selected bibliography is also provided.

Most of the authors were able to meet in advance, and to present, discuss, and then revise their contributions. But the line has to be drawn somewhere, and authors who receive a reply in this volume were not permitted to adjust their papers in the light of the final version of the reply. The initial exchange took place in Saarbrücken and Saarlouis in June 1992.

*Everybody has been very kind to us. Georg Meggle — selfless and cheerful as usual — co-designed the project and supported it from beginning to end. When we proposed the meeting, we were backed up by Franz von Kutschera and Wolfgang Lenzen. Barbara Schumacher helped prepare and run it.

The editors of *Perspectives in Analytical Philosophy*, Georg Meggle and Julian Nida-Rümelin, have welcomed the book in precisely the form we suggested. The authors have been co-operative and patient throughout. Christopher Abbey and Seán Matthews have given valuable advice, linguistic and otherwise, to many of us. Kornelius Bamberger was able, and kind enough, to convert most of the data that the contributors sent us. Thomas Fehige gave these data a neat, uniform lay-out. Patrick Agsten, Morika Claßen, Franziska Muschol, Ulf Schwarz, and Valentin Wagner have assisted us, efficiently and in numerous respects; the same holds true of Karin Thom. With this list in chronological order, one important acknowledgement comes last: de Gruyter publishers. Working with Hans-Robert Cram was a pleasure; ditto, at the technical end, with Grit Müller.*
The conference that gave rise to this book was made possible by the financial assistance of: Deutsche Forschungsgemeinschaft (DFG), Fritz Thyssen Stiftung, Ministerium für Wissenschaft und Kultur des Saarlandes, Universität des Saarlandes, and Vereinigung der Freunde der Universität des Saarlandes. The DFG (research project "Was zählt?") has also funded our own work on this volume.

We thank all these persons and institutions for their support.

* 

We share the belief, now regarded in some quarters as both unsound and old-fashioned, that, in essence, morality is all about welfare, and welfare all about preferences. Some of the contributors to this volume would agree, some would not. With luck, this collection will help advance matters a little.

Christoph Fehige and Ulla Wessels
Leipzig, January 1998
Contents

Abstracts ........................................... xiii

Introduction

CHRISTOPH FEHIGE AND ULLA WESSELS
Preferences – an Introduction ..................... xx
Preferences – a Short Bibliography ................ xlv

Part I: Preference and Decision

ELIJAH MILLGRAM
Deciding to Desire .................................. 3
SYDNEY SHOEMAKER
Desiring at Will (and at Pill): A Reply to Millgram .... 26

CHRISTOPH LUMER
Which Preferences Shall Be the Basis of Rational Decision? .... 33
ANTONELLA CORRADINI
Intrinsic Desirabilities: A Reply to Lumer ............... 57

RICHARD B. BRANDT
The Rational Criticism of Preferences .................. 63
ANNA KUSSE
Rational by Shock: A Reply to Brandt .................. 78

J. DAVID VELLEMAN
Is Motivation Internal to Value? ..................... 88
GEORG MEGGLE
Motivation and Value: A Reply to Velleman ............. 103

SVEN DANIELSSON
Numerical Representations of Value-Orderings: Some Basic Problems 114
ULRICH NORTMANN
Interval Orders Defended: A Reply to Danielsson ........ 123
Contents

Part II: Preference and Metaethics

LENNART AQVIST
Prima Facie Obligations in Deontic Logic: A Chisholmian Analysis
Based on Normative Preference Structures .......................... 135

UWE BOMBOSCH
The Meaning of “Ought, Prima Facie” and Decision Situations:
A Reply to Aqvist ..................................................... 156

FRANZ VON KUTSCHERA
Values and Duties .................................................... 163

WILFRIED HINSCH
Beyond Duty: A Reply to von Kutschera .......................... 172

MARCUS WILLASCHEK
Agency, Autonomy and Moral Obligation ......................... 176

HILARY BOK
Autonomy and Morality: A Reply to Willaschek ................ 204

DAVID WIGGINS
In a Subjectivist Framework, Categorical Requirements and Real
Practical Reasons .................................................... 212

DAVID GAUTHIER
Subjective Obligation: A Reply to Wiggins ....................... 233

ALLAN GIBBARD
Preferences and Preferability ..................................... 239

JULIAN NIDA-ROMLIN
Goodness and Rational Preferability: A Reply to Gibbard ...... 260

Part III: Preference and Ethics

JOHN BROOME
Extended Preferences ............................................. 271

RUDOLF SCHOSSLER
Wish You Were Me: A Reply to Broome and a Comment on Harsanyi’s
Extended Preference Theory ..................................... 288
Contents

RAINER HEGSELMANN
Experimental Ethics: A Computer Simulation of Classes, Cliques, and Solidarity ................................ 298

ULRICH KRAUSE
Solidarity among Rational Egoists: A Reply to Hegselmann ........ 321

RAINER WERNER TRAPP
The Potentialities and Limits of a Rational Justification of Ethical Norms, or: What Precisely is Minimal Morality? ........ 327

ANTHONY SIMON LADEN
A Hobbesian Choice: A Reply to Trapp ............................ 361

CHRISTOPH FEHIGE, RICHARD M. HARE, WOLFGANG LENZEN, JEFF McMAHAN, PETER SINGER, THOMAS SPITZLEY, AND ULLA WESSELS
Symposium on Possible Preferences ................................. 367

CHRISTOPH FEHIGE AND ULLA WESSELS
Introduction to Possible Preferences ............................... 367

PETER SINGER
Possible Preferences .................................................. 383

RICHARD M. HARE
Preferences of Possible People ...................................... 399

WOLFGANG LENZEN
Who Counts? ......................................................... 406

ULLA WESSELS
Procreation ............................................................ 429

JEFF McMAHAN
Preferences, Death, and the Ethics of Killing .................... 471

THOMAS SPITZLEY
McMahan on Psychological Continuity and the Value of Future Goods ......................................................... 503

CHRISTOPH FEHIGE
A Pareto Principle for Possible People ........................... 508
Contents

Notes on Contributors ............................................. 544

Name Index .......................................................... 550

Subject Index ....................................................... 555
Abstracts

These are abstracts of the papers that receive a reply, not of the replies themselves. The abstracts appear in the alphabetical order of the authors' names; for the contributions to the symposium on possible preferences, see the final abstract.

LENNART AQVIST
Prima Facie Obligations in Deontic Logic: A Chisholmian Analysis Based on Normative Preference Structures
(page 135)

The paper argues for an analysis of the W. D. Ross notion of prima facie obligation which results from adding a certain Chisholm-style definition to the system $G$ of Dyadic Deontic Logic, supplemented with so-called propositional quantifiers. In the semantics for that system a von-Kutschera-inspired conception of normative preference structures turns out to be of vital importance.

UWE BOMBOSCH's comment, "The Meaning of 'Ought, Prima Facie' and Decision Situations", begins on p. 156.

RICHARD B. BRANDT
The Rational Criticism of Preferences
(page 63)

Preferences are rationally criticized if vivid representation of confirmed beliefs will result in a reversal or strengthening. It is universally agreed that plans can be so criticized but not basic preferences for types of events. Define "preference" as "desiring more". Psychologists agree that desire for an event-type is increased if an event-type has been associated with pleasant events in the past — conditioning by contiguity. (The status of bodily needs — like thirst and hunger — is different; such needs are fixed by chemical imbalances in the body.) But many events are pleasant for evolutionary reasons; if they weren't pleasant and hence the pleasant type of event wanted, the individuals would not survive. This connection — pleasant event, being wanted from classical conditioning, and hence preference — opens the way to rational criticism. For reflection on facts can alter preferences when the preference is seen to be a result (1) of inadequate representation of facts, or (2) of influence by temporary motivational states, or (3) of stimulus generalization from abnormal cases, or (4) of overlooking unpleasant facts about the object, or (5) of failure of making discriminations, or (6) as a result of suggestions by teachers, or (7) as a result of false or unjustified factual beliefs. The author suggests we
say a preference has been rationally criticized if reflection on these defects results in a modification of the preference.

Anna Kusser's comment, "Rational by Shock", begins on p. 78.

John Broome
Extended Preferences
(page 271)
Ordinalism is generally taken to imply that interpersonal comparisons of good are impossible. But some ordinalists have argued that these comparisons can be made in a way that is consistent with ordinalism, on the basis of extended preferences. This paper shows that this argument is mistaken, and ordinalism is indeed incompatible with interpersonal comparisons of good.


Sven Danielsson
Numerical Representations of Value-Orderings: Some Basic Problems
(page 114)
Measures of value or preference usually presuppose value or preference relations which are weak orders. Numerical representations of semiorders and of interval orders have to some extent also been considered. It is fairly obvious, however, that value- and preference-orderings often are not, and should not be expected to be, even interval orders. A way of representing partial orders is suggested.

Ulrich Nortmann's comment, "Interval Orders Defended", begins on p. 123.

Christoph Fehige and Ulla Wessels
Preferences – an Introduction
(page xx)
In theories of practical reasoning, we can encounter preferences (desires, and the like) in five places. Two of them are the form and the content of rationality; the other three are the form, the content, and the foundation of morality. This introduction presents the terrain and explains its overall structure; it also pays a brief visit to each of the locations and points out some of the disputes surrounding them. The doctrine of preferentialism and its problems will be a convenient leitmotiv, since it is widely held and employs preferences, and preferences only, on all the five levels. The tour is structured as follows. After a prologue that sketches preferentialism, we will consider the very
concept of a preference (section 1). We will then look at the possible roles of preferences in rationality (section 2), and at the triad of roles they might play in morality (section 3). Finally, there is a selected bibliography.

**ALLAN GIBBARD**

Preference and Preferability

(page 239)

What does "good" mean? The paper starts with two vague truisms: That goodness is a matter of preferability, and that of two things, the preferable one is the one it is rational to prefer. In his book *Wise Choices, Apt Feelings* (1990), the author had proposed a theory of what "rational" means; this paper faces two questions: (1) What concept of preference, if any, fits the formula that the preferable of two things is the one it is rational to prefer? (2) How should this formula be filled out: Rational for whom to prefer, when? Classical decision theory treats preference as consisting in one's disposition to choose. Such revealable preference will probably not serve as a good explanatory concept in a scientific psychology, it is suggested, but it may be much the concept that is needed for purposes of defining preferability. Roughly, the preferable of two things is the one it is rational to choose. This needs to be refined, though: Talk of what is preferable to what purports to be neutral among parties to the conversation. Indeed one use of the term "good", prominent among philosophers, treats all humanity as our conversational group. If rational intrinsic preferences need not be impartial, then not all considerations that bear on rational choice need be matters of goodness so understood. Good-making considerations will be those considerations that bear on choices consequentially and neutrally, and goodness will be a matter of how these good-making considerations sum up.

Julian Nida-Romelín's comment, "Goodness and Rational Preferability", begins on p. 250.

**RAINER HEGSELMANN**

Experimental Ethics:
A Computer Simulation of Classes, Cliques, and Solidarity

(page 298)

The article deals with two questions: (a) Can relations and networks of solidarity emerge in a world exclusively inhabited by rational egoists, who are unequal and choose their partners opportunistically? (b) If networks of solidarity do emerge in such a world, what do they look like? By means of computer simulations it is shown that networks of solidarity can emerge in such a world. But the networks will show quite distinct features of some class segregation.

It is argued that both deontological and consequentialist principles have their legitimate place in ethics, and that neither kind is reducible to the other. The problem, then, is how to integrate them into a unified system. A simple solution would be to have duties override value considerations, and the discussion centers on the merits and shortcomings of this proposal.

Wilfried Hinsch's comment, "Beyond Duty", begins on p. 172.

Theories of rational decision normally distinguish basic and other preferences, using only the former for calculating an agent's utility function. The idea behind the distinction is that, on the one hand, a theory of rational decision must allow criticism of at least a part of the agent's actual preferences; on the other hand, so as not to lose touch with the agent's real interests, it must rely on his factual preferences. Different decision theories have declared as basic various sets of preferences, thereby arriving at very different utility functions. Therefore, the question of which preferences shall be basic is of large practical importance. Nonetheless, it has rarely been discussed.

The article criticises some standard approaches, but mainly develops criteria for the selection of basic preferences. One of the principles for the selection of basic preferences, for example, is epistemic rationalisation. From these principles, then, 12 conditions of adequacy for the selection of the preferential basis are derived, e.g. taking over only intrinsic preferences, and of these not the single preferences but their underlying criteria.

Antonella Corradini's comment, "Intrinsic Desirabilities", begins on p. 57.

We do not, and cannot, normally come to have desires by simply deciding to have them. It is argued that this is not a contingent fact, and that the explanation for this fact shows a widely held view of practical reasoning to be false.

Sydney Shoemaker's comment, "Desiring at Will (and at Pill)", begins on p. 26.
Abstracts

Rainer Werner Trapp
The Potentialities and Limits of a Rational Justification of Ethical Norms, or: What Precisely is Minimal Morality?
(page 327)

Starting from the insight that, due to certain epistemological peculiarities of 'normative truth', normative statements cannot claim to be objectively (= O) true, the paper systematically works out the idea of basing the O-validity of general moral norms on their O-utility rather than on their O-truth. According to this idea any restriction of choice, in an n-person-conflict of interests S, qualifies as O-valid if it fulfills one of the two following criteria: Either compliance to it by at least a specifiable number k of the n individuals in S would make everybody already in each instance of S better off than norm-free anarchy (= criterion (I), which establishes two classes of unconditionally O-valid norms each avoiding a corresponding type of trap of prudence), or it would, under certain assumptions of the interacting individuals on the probabilities of the roles taken in their respective lifetime-sequences of situations of type S, increase everybody's utility payoff in the long run (= criterion (II), which establishes three classes of only conditionally O-valid norms). Thus even 'non-veiled' rational egoists refusing to initially concede any rationally unfoundable moral protonorm whatsoever, one that demands some (Harsanyian, Rawlsian, ... ) impartial standpoint in considering an agreement on mutual restrictions of behaviour, will -- so it is argued -- have to contract on at least these norms in a fictitious original agreement. The latter's extension defines the system $M_{\text{min}}$ of minimal morality. Though being far more comprehensive than related approaches to 'morals by agreement' (notably Gauthier's), $M_{\text{min}}$ will finally be assessed as morally insufficient due to its not containing any compensatory norms. Since some of the latter, according to widespread convictions, are indispensable and since these, at the same time, are not justifiable as O-valid on the basis of whatever brand of veil-free contractarianism, any programme of founding a satisfactory moral system on mere collective rationality is considered as doomed to fail eventually.


J. David Velleman
Is Motivation Internal to Value?
(page 88)

The view that something's being good for a person depends on his capacity to care about it -- sometimes called internalism about a person's good -- is here derived from the principle that 'ought' implies 'can'. In the course of this derivation, the limits of internalism are discussed, and a distinction is drawn between two senses of the phrase "a person's good".

Georg Meggle's comment, "Motivation and Value", begins on p. 103.
In this paper, the author tries to show that Hume, interpreted as a genealogist of morals—not as empiricist, prescriptivist, projectivist, expressivist or error theorist—can do justice to the moral phenomena that moral philosophers discuss under the heading of the categorical imperative. His position on this matter is compared and contrasted with that of Kant. It is claimed that Hume discusses the real reasons, such as they are, why, regardless of inclination, we should heed the categorical requirements of morality.


The paper proposes and, in part, defends an understanding of human agency, autonomy, and moral obligation as integral parts of our concept of a person. Specifically, the first part (sects. 1–12) argues for a causal theory of action in which the acting person plays a central role in the causal history of her actions. The person exercises her causal influence according to normative principles of rationality. That presupposes some independence from her own motivation including the ability to acknowledge or reject parts of it as a basis of her rational decisions. This ability is constitutive of the autonomy of the person. The second part (sects. 13–29) presents an argument to the effect that the concept of autonomy presupposes a general universalist principle of morality. Autonomy involves a distinction between motives that are ‘authentic’ and motives that are not. This distinction does not rest on a substantive idea of what autonomous action is, but rather on a formal or procedural notion. Nevertheless, it presupposes a normative standard which is different from and largely independent of the motives a person in fact has. This standard can be found in the ideas of impartial benevolence and universal rational consent which inform universalist conceptions of morality.

Hilary Bok’s comment, “Autonomy and Morality”, begins on p. 204.

Sometimes our actions make a difference not just to the frustration or satisfaction of preferences that exist (have existed, or will exist), but to the very question which preferences will exist; so they require us to look not only at actual, but also at possible pref-
ferences. These actions, their morality and their rationality, are the topic of the present symposium.

Most choices concerning a preferrer's life or death are dramatic and obvious examples of such actions (no life, no preference), and they have come to dominate the discussion of possible preferences, and this symposium as well. Thus, on the more applied level, this is a symposium about the morality of conception and contraception, abortion, population policy and killing, about the value of life and the badness of death.

For a guide to this web of issues, see the "Introduction to Possible Preferences" at the beginning of the symposium (p. 367); more information on the various contributions, and on how they relate to each other, is given in the last section (pp. 379–81) of that introduction.
In theories of practical reasoning, we can encounter preferences (desires, and the like) in five places. Two of them are the form and the content of rationality; the other three are the form, the content, and the foundation of morality. This introduction presents the terrain and explains its overall structure; it also pays a brief visit to each of the locations and points out some of the disputes surrounding them. The doctrine of preferentialism and its problems will be a convenient leitmotiv, since it is widely held and employs preferences, and preferences only, on all the five levels. The tour is structured as follows. After a prologue that sketches preferentialism, we will consider the very concept of a preference (section 1). We will then look at the possible roles of preferences in rationality (section 2), and at the triad of roles they might play in morality (section 3). Finally, there is a selected bibliography.

1. The Concept of Preference
   1.1. General Problems
   1.2. The Taxonomy of Preferences

2. Preference and Rationality
   2.1. General Problems
   2.2. Rational Decision Theory

3. Preference and Morality
   3.1. Normative Preferences and the Format of Morality
       3.1.1. Consequentialism
       3.1.2. Deontic Logic
   3.2. Preference Satisfaction and the Content of Morality
       3.2.1. Welfare and Preference Satisfaction
       3.2.2. Distributing Preference Satisfaction
   3.3. Shared Preferences and the Foundations of Morality

Preferences, so the received opinion, are the alpha and omega of practical reasoning: people are rational if they do what they believe will best satisfy their own preferences (more on this in section 2 of this introduction); and people

* To find writings about issues raised in the introduction, consult the similarly structured bibliography that follows it. We are grateful to Christopher Abbey, Krister Bykvist, Georg Meggle, and Elijah Millgram for helpful comments. We thank the Deutsche Forschungsgemeinschaft (DFG) for supporting the research project "Was zählt?"; work on this survey was part of the project.
are moral if they do what will best satisfy everybody’s preferences (more on that in section 3). Call this type of position preferentialism.\footnote{Unable to spot an existing label, we used “preferentialism” when we wrote the drafts of our papers for this volume, which have been circulating widely since 1992. But the word is an obvious choice, and it’s quite possible that other people had the same idea; information about earlier occurrences would be welcome.}

Typically, preferentialism is inspired by an enlightened respect for people’s autonomy: prejudice aside, it is their own wishes that count. Chacun à sa façon.

Popeye, for example, likes eating spinach, and eating spinach conflicts with none of his other wishes. Surely this gives him a reason to do what he believes will provide him with spinach. It would be puzzling for a concept of rationality to deny this; reason would lack sensitivity, and nobody would want to apply it to their own affairs. Say that Mary, however, does not like spinach, and does not believe it to be conducive to any other end of hers. Then how could there be a reason for her to eat it? It would be puzzling if a concept of rationality required her to eat spinach no matter what she wanted; reason would be dictatorial, and again nobody would want to apply it to their own affairs.

Same thing with morals. Given that Popeye wants spinach, the world would, ceteris paribus, be a little better if we gave him some; so that is what we ought to do. And given that Mary does not want spinach, the world would, ceteris paribus, not be better if we gave her any; so we have no obligation to do so. This is how, if we believe preferentialism, both reason and morality correlate with preferences.

1. The Concept of Preference

1.1. General Problems

Precisely what it is that preferentialists say, and their opponents deny, will depend on what they mean by “preference”, or – for some versions of the theory – by words such as “aim”, “desire”, “end”, “project”, “want”, “wish”, or “liking” (all of which we will use interchangeably). Concepts of preference can vary in at least three basic dimensions: logic, representation, and charge.

- Logic. Does the verb “prefer” take two objects, as in “John prefers bananas to apples”? Though this is what its etymology suggests, it is now often used, like “wanting”, with just one object. And then what sort of object(s)? To get a logically well-behaved concept of preference, philosophers tend to
restrict the choice essentially either to propositions (states of affairs, etc.)
or to possible worlds (outcomes, scenarios, etc.).

What about axioms? Which of the following candidates, for instance,
should qualify as conceptual truths?

- Asymmetry: if (at point of time \( t \), all things told) you prefer world \( \alpha \)
to world \( \beta \), then you do not (at point of time \( t \), all things told) prefer
world \( \beta \) to world \( \alpha \).

(Time indices and the like will be omitted in the following ex­
amples.)

- Connexity of weak preference: for every two outcomes \( \alpha \) and \( \beta \), either
\( \alpha \) is weakly preferred to \( \beta \) or vice versa.

("Weakly to prefer \( \alpha \) to \( \beta \" \) means preferring \( \alpha \) to \( \beta \) or being indif­
ferent between the two.)

- Reflexivity of weak preference: every outcome is weakly preferred to
itself.

- Transitivity: for every three outcomes \( \alpha \), \( \beta \), and \( \gamma \), it holds true that,
if \( \alpha \) is (weakly) preferred to \( \beta \), and \( \beta \) (weakly) preferred to \( \gamma \), then \( \alpha \) is
(weakly) preferred to \( \gamma \).

- If you want it to be the case that \( p \), you do not want it to be the case
that non-\( p \).

- You want \( p \land q \) if (and even only if?) you want \( p \) and you want \( q \).

- You want \( p \) if and only if you prefer \( p \) to non-\( p \).

- If an agent prefers outcome \( \alpha \) to outcome \( \beta \), then, for every \( p \) with
\( 1 \geq p \geq 0.5 \), a lottery that will give him outcome \( \alpha \) with probability \( p \)
and outcome \( \beta \) with probability \( 1 - p \) will also be preferred by him to
outcome \( \beta \).

A set of answers to formal questions like these will characterize a logic of
preference – a prohairetic logic, as it is sometimes called. The status of its
axioms, too, will require reflection: is compliance with them necessary for
something to qualify as a system of preferences at all, or just to qualify as
a respectable (for instance, consistent or rational) one?

- Representation. Desiring, it would seem, has something to do with grasping
or representing the object(s) of desire. By itself, no reaction to the sentence
"I shall drive a Rolls Royce" will qualify as a wish to drive a Rolls Royce.
The preferrer must have got the semantics right, too. It has to be the case,
say, that, if she fully represented to herself the possibility of driving a Rolls
Royce, she would attempt to make it come true, or would feel good (or
whatever). But precisely what does preferring have to do with represent­
ing? And what is it for somebody to represent, or grasp, something? Philo­
sophers of mind have been wondering for a long time; no agreement is in
sight, and some of them want to do without representation altogether.

- **Charge.** If a preference is, at a first approximation, a pro- or an anti-attitude towards something, then what do the "pro" and the "anti" consist in? What, as we can put it, is the charge of a preference? Some theories turn to feelings: you want it to be the case that $p$ if, other things being equal, the thought of $p$ would make you happy (or enthusiastic, or yearning); or if, other things being equal, you'd be more pleased believing that $p$ than you would be believing that non-$p$. Other theories turn to actions: you want something if, in appropriate circumstances, you would try to get it. Some people claim that a desire is a belief; for them, the task of explicating charge is not the task of saying what distinguishes desires from beliefs, but what distinguishes desires from other beliefs. Belief or not, once we know what to mean by "charge", we will also want to define a measure for it (a little more about that on p. xxxi below). For, presumably, preferences can vary in strength.

The three dimensions — logic, representation, charge — characterize an array of conceptual options that we cannot explore here in any detail. Neither will we rehearse how the choice would affect every philosophical evergreen that it does affect — such as: is the will free? How do the mind in general and the will in particular relate to the body? What do we mean by "weakness" of the will? One thing we can do, however, is glance at a question of method. When constructing a notion of preference, how would we decide which way to go? Here comes a simple but important part of the answer: there is nothing wrong in the preferentialist using her creed itself as a criterion. To a significant degree, she can tailor concepts of preference for the sole purpose of making the preferentialist doctrine true, or plausible. Suppose, for example, she believes hedonic happiness (feeling good, that is) to be of moral importance; to make sure that preferentialism respects this claim, she could, when defining "preference", simply stipulate (or rig up her concepts of representation and charge to entail) that people prefer, other things being equal, to feel good.

It is therefore unfair of the anti-preferentialist to 'refute' preferentialism by pouring just any old concept of preferring into it, and pointing out that that combination has crazy implications. The manoeuvre is deplorable, and too common in moral philosophy. But it is equally unfair, and no less frequent, for a preferentialist not to tell us what is meant by "preference". As long as she doesn't, her creed says practically nothing; it could be free of errors by being free of content.

This sheds some light on the question whether it should not be the *philosophy of mind* that guides us through the conceptual maze. On the one hand,
it might be able to help, and today preferentialists are far from exhausting that possibility. When mongering, for the purposes of rational decision making and morals, their own concept of desire, they would be wise to exploit the extensive general debate on propositional attitudes, mental representation, and qualia.

On the other hand, the transfer will be limited. Suppose we ask philosophers of mind where they get their criteria of adequacy from. For many of them, the final touchstone for an explication of beliefs and desires is whether it helps empirical psychology meet the usual standards for a scientific theory: it should be elegant, have predictive success, dovetail nicely with the rest of science (say with biology or computer science), etc. Normative contexts, like rationality (in the not merely explanatory sense) and morality, make no explicit appearance on this list; and for them, some of the items that do appear on it are at best of minor importance. Thus, the concept of preference favoured by the scientific criteria might diverge radically from that favoured by normative criteria.

Take, for example, what we called the charge of a preference — i.e. the property in virtue of which an attitude towards something is an attitude for or against it. The dovetail requirement from the scientific list might well suggest a definition in terms of brain activity, neurons, and so forth (which fits in nicely with chemistry, which in turn fits in tolerably well with physics); but a normative requirement — for instance, that we are out to capture something that would make a moral difference — might well point towards a more phenomenal definition. (“Neurons or not,” we might want our preferentialism to say, “where the thought of $p$ would arouse no longing or no pleasure, there is no moral reason to bring $p$ about.”) What we cannot do is first say that preferences are what matters and then leave it to philosophy of mind to tell us what we mean by “preference”. Not, that is, if philosophy of mind is philosophy of science for empirical psychology.

Making up our mind about logic, representation, and charge respectively will be difficult enough. Even then, the combination of our favourite three answers could still be dissonant. Say, for example, that in order to get the nicest logic of preference we should work with preferences between possible worlds. This is a plea that experts from the department of mental representation might well want to oppose — for how could real-life people ever react towards an entire possible world (unless it happens to be a particularly small and dull one)? Our mind’s eye has a limited field of vision, and maybe most worlds are simply too large for it. If such clashes among the conceptual requirements were unavoidable, they would provide a very strong argument against preferentialists: they cannot find a coherent concept of preference at all, or at least none that fits their bill.
1.2. The Taxonomy of Preferences

From the most general questions, we turn to taxonomical ones. Specific problems in the theory of practical reasoning will point us to specific types of preferences worth capturing and baptizing. Here are sketches of a few important examples:

- Wanting something for its own sake (an *intrinsic preference*) as opposed to wanting it for the sake of something else (an *extrinsic preference*). Typically, preferentialist moralities want to count intrinsic preferences only.
- Wanting something if all other things were equal (ceteris paribus preference, *pro tanto preference*) as opposed to wanting it even in view of the costs, or putative costs, of getting it (preference all things told, *final preference*). The distinction is closely related to the previous one.
- Preferences the preferrer is aware of (explicit preferences) as opposed to those that he isn't (implicit preferences).
- Preferences that are compatible with rationality (rational preferences) as opposed to those that aren't (irrational preferences). The rational preferences include those that are neither dictated nor ruled out by rationality (arational preferences); a preference for strawberries over blueberries seems to be, where all other things are equal, a safe example. Typically, “rationality” is meant to refer either to the full representation of the content of the preference or to certain requirements of consistency among a person's preferences (for both these aspects, see sect. 1.1 above); and typically, only rational preferences are supposed to count in morals.
- Preferences with a fishy history (heteronomous preferences) as opposed to others (autonomous preferences, authentic preferences). Standards of fishiness are of course debatable (and some would say fishy); preferences that would not survive psychotherapy and preferences caused by a drug, a brainwashing, a disease, or an unfair biography are standard examples. A fishy history is, as its name suggests, frequently taken to undermine a preference's moral, or rational, standing. Some people seem to call a preference heteronomous in so far as it merely happens to the preferrer, rather than being created, or at least bossed around, by his rational self (sometimes believed to be, ipso facto, his moral self). These distinctions and their point are not always clear (to us).
- Preferences that could exist (possible preferences), dividing up into those that could but don't exist (merely possible preferences) and those that have existed, exist, or will exist (actual preferences). Among preferentialists, the question whether there is any moral or prudential point in actualizing and
satisfying possible preferences (as opposed to just satisfying actual preferences) is controversial.

- Preferences somebody had (past preferences), has (present preferences), and will have (future preferences). Some people think that an agent may rationally discount, or even ignore, her own past preferences, and some think the same holds for her future preferences; it is also held sometimes that for moral purposes we can ignore all or some past preferences (for example those of the dead, or those revised by the preferrer).

- Preferences whose entire content would, if true, be simultaneous with the preferring itself (synchronic preferences) as opposed to others (asynchronous preferences, for example now-for-then preferences, then-for-now preferences, ex-post preferences).

- Preferences that concern just the preferrer’s state of mind. If you restrict preferentialism to them, as some people suggest we should, then you get (one version of) hedonism.

- Preferences that concern a substantial part of one’s own life (global preferences) as opposed to those that concern only a small part of one’s life (local preferences).

- Extended preferences, i.e. preferences as to what should be the case were one in somebody else’s shoes.

- Metapreferences, i.e. preferences whose content is, or entails, the existence or non-existence of certain of the preferrer’s preferences.

- External preferences, i.e. preferences whose content entails the existence or non-existence, or the satisfaction or frustration, of other people’s preferences; preferences that are not external are sometimes called personal preferences. External preferences can be moral, benevolent, altruistic preferences, or immoral, malevolent, anti-social, evil, sadistic preferences, or mixtures of these. Some preferentialists suggest we give no weight to external preferences, or at least not to evil preferences, in moral reasoning.

Both the types and the details of such distinctions vary from problem to problem and from author to author; one factor they will tend to correlate with are the author’s favourite objections to, or versions of, preferentialism.

2. Preference and Rationality

2.1. General Problems

Having glanced at the conceptual tasks that preferences are likely to confront us with, we can now turn to their roles in practical reasoning. We start with
rationality (this section) and will move on to morality (section 3). Since some important issues concern, mutatis mutandis, both these areas, our decision to discuss a certain point in this rather than the next section, or vice versa, may be debatable.

Recall the first part of the preferentialist creed. Somebody is rational, it says, if he does what he expects will best satisfy his own preferences.

In a sense, this part of the creed contains two appeals to preferences. The first of them is rather structural. It is hiding behind the word “best”, and concerns the relation “at least as good as” that is presupposed by “best”. A relation, it is claimed, that has the formal properties typical of a preference relation is constitutive for, or induced by, rationality, and plays a central role for or in it. Such a relation is something very much like an ordering, or quasi-ordering, defined, say, over actions, lotteries, or outcomes. It can go by any name that conveys some impression of comparativity, for instance “at least as good as” (perhaps with the insert “for the agent”), or, for whatever $X$, “meeting criterion $X$ at least as well as”, or “producing at least as much $X$ as”. If the claim sounds terribly abstract, never mind – it is terribly abstract.

We come to the second claim about preference and rationality, which is less formal. Rationality, it says, has a lot to do with the agent’s preferences or desires in some not too far-fetched psychological sense of those terms.

The two options can diverge. Somebody could hold, for example, that the rationality of an action is, ceteris paribus, proportional to the amount of knowledge that it will help the agent to acquire, or that the agent believes it will help him to acquire. If we assume, for the sake of the argument, that knowledge can be measured, then this concept of rationality would involve a preference in the purely formal sense (viz. a quasi-ordering of actions), but none in any serious psychological sense (since maybe the agent doesn’t care tuppence for knowledge).

The difference - preferences in the form versus preferences in the content - is important. (It will be dealt with again in the sections on ethics, see pp. xxxi–xxxix.) Given a critique of preferentialist concepts of rationality or morality, one should ask whether it applies to every theory of the same form or just to the sub-class of theories that use preferences-in-the psychological-sense as the substance to go into the form. For reasons of space, we shan’t always be able to pursue the distinction in this introduction.

We return, then, to the preferentialist double bill of form and content: to the claim that it is rational for somebody to do what he expects will best fulfil his own wishes. If we ask why this is supposed to be so, we shall approach deep waters – how does one argue for a theory of rationality? Where reason itself is to be measured, what can serve as the yardstick?
Since every theory of rationality might lack a glorious response to this type of challenge (the ultimate challenge), perhaps the preferentialist, too, may restrict herself to non-fundamental forms of justification. She could point to our pre-theoretical beliefs, and claim that her theory accommodates them best. And indeed, choosing what looks like the best means to one’s ends is at least one central part of our everyday notion of rationality, or even of sanity. If somebody told us that he wants to get rid of a headache; that he could take aspirin; that this, but nothing else, would help; that taking it wouldn’t conflict with any other projects of his – and that he is not going to take it, then we would start wondering.

However, is means-ends rationality the only part of our pre-theoretical notion? What about the ends themselves? Can’t they be irrational? Preferentialism, say some of its critics, doesn’t take this possibility seriously enough; the doctrine, they say, is too instrumentalist.

Typically, the preferentialist, too, will accept certain requirements of rationality for the preferences themselves. She is, for instance, likely to demand some sort of consistency among them (for examples of such constraints, see pp. xxi ff. above). And she is likely to insist that, for a preference to be a preference in the sense she’s after, it has to be accompanied by, or would have to survive, a full representation – whatever that may be – of the desideratum (see pp. xxii f. above).

But do criteria like consistency and semantical competence go far enough? It seems that some preferences which, pre-theoretically, we would have considered as irrational (or, to use the pre-theoretical term, crazy) could meet the preferentialist standards: e.g. a preference for eating mud, or for exterminating a racial minority. If that is so, then maybe preferentialism does not accommodate our initial beliefs about rationality best; maybe it generalizes from too narrow a choice of examples, all misleadingly similar to the aspirin case.

The preferentialist will reply that, as so often in good theory making, her doctrine, though based on preconceptions, will also force us to revise some of them. For instance, we should indeed stop calling people irrational who really (i.e. consistently and competently) want to eat mud. Furthermore, she would ask what the stronger criteria that an opponent might suggest would look like, and why. Should preferences count as irrational just because they are rare, or ‘unnatural’? Could, requirements of consistency and representation aside, a judgement on the rationality of ends fail to be anything but arbitrary?

One project for which the question of instrumentalism could make a drastic difference is the rational foundation of morals (see also pp. xxxix–xli below). If, say, the preference for exterminating a racial minority can be rational, but the racial minority ought to flourish, then we have a gap between
reason and morality; if such preferences can't be rational, we might be spared the gaps.

Preferentialist rationality has also been criticized for being consequentialist— it focuses, that is, on the possible consequences of an action. Since looking at a possible consequence is not always in the agent's own best interest (some unpleasant consequences, for example, can be caused by contemplating them), preferentialist rationality is a questionable method. Maybe, however, no other method is superior to it—and if it were, preferentialism might still offer the best criterion of rationality. (More about consequentialism when we come to ethics, see pp. xxxii–xxxiv.)

Sages warn us that preference satisfaction is a Sisyphean task. We are all like the fisherman's wife. Once a wish is granted, a new one pops up and takes its place; we shall have gained nothing. The preferentialist will reply that, firstly, this situation may not be inescapable; we can form our characters and try to develop a reasonably satisfiable amount and type of wishes. Preference satisfaction, and the sages' warning, could even be our rational guides for that enterprise itself. Secondly, suppose the situation were indeed inescapable. Our desires are a Hydra, and there is nothing we can do about it. It is not clear whether this would refute the preferentialist criterion. If it were part of the human condition that, by preferentialist standards, many actions just don't make a difference, then why should this argue against the standards, rather than against the human condition? (Compare rationality to a metal detector: if there is no metal we do not blame the detector for being silent.)

The sages also warn us that preference satisfaction and hedonic happiness might diverge. Suppose that you vehemently desire a job, but that soon after you get it your days are no brighter than before—you might even feel disappointed. Again, preferentialists would claim that their theory can do justice to this problem. They ask whether in such a case the relevant preference would really have survived their requirement of full representation (see pp. xxii f. above). And if the answer is yes, they could still point to people's preferences for spending their conscious time as pleasantly as possible (see p. xxiii above). Then your appointment has satisfied one preference (the one for the job), but not the other (the one for feeling good).

The objections to, and problems within, preferentialist rationality are too numerous for us even to list them here. To mention just a few more: does practical reason process beliefs and desires deductively or defeasibly? And instead of encouraging the agent to maximize something (say her preference satisfaction), should we perhaps advise her to satisifice (to get a certain, but not necessarily maximum, amount of goods or satisfaction)? And does an agent, in order to be rational at a certain time, have to take all her present and future
preferences equally seriously? And even her past ones? And what if her choice is also a choice between preferences? Would she, on pain of irrationality, have to choose to acquire the set of preferences that she expects to generate most satisfaction? If we answered in the affirmative, would we get into trouble? Would it follow that all the rational preferrer can ever prefer for its own sake is preference satisfaction itself? And that she would have to welcome the prospect of a brainwashing that would leave her in love with every single feature of the world as it is (murder and hamburgers included)?

2.2. Rational Decision Theory

Attempts to render preferentialist rationality more precise have led to a formal field of inquiry known as rational decision theory. Typically, the initial question is what the rational thing to do would be if you weren't certain which of your actions would have which consequences.

In the worst case, you have no idea even how likely your actions are to cause one outcome rather than another. Various general decision rules for this case have been discussed, but none has established itself as canonical. One famous candidate with obvious limitations is 'maximin': "Take an action whose worst possible outcome is not worse than the worst possible outcome of any alternative action."

If we decide to leave the most dramatic cases of ignorance aside, then, for every action and each of its possible outcomes, the agent can assign at least a probability to the claim that the action would have that outcome. For this case, the basic idea, known as Bayesian decision theory, is to go by the actions' expected utility. To each of an action's possible outcomes we assign a numerical value that represents (something like) 'how strongly' the outcome is preferred by the agent; for each outcome, we multiply its value by the probability of the proposition that the action would have that outcome; when we have all these products (one for every outcome the action might have), we sum them up, call the sum the action's expected utility, and advise the agent to choose the action (or one of the actions) with the highest expected utility.

What happens here is that we weight utility with probability. Given a possible outcome, the higher its value, the higher the expected utility of the respective action; if the value is positive, then the lower the probability of achieving the outcome, the lower the expected utility; etc. This is the type of reasoning we all use when, for the possibility of a gain, we would have to risk a loss. We run the risk if the probability for the gain is high and the loss small; we refuse when the probability is too small or the loss too dramatic.
One vital (or lethal) question for any such theory is where it gets the numbers from that are supposed to represent what the outcomes are worth. After all, numbers do not generally pop up in an agent’s mind when he contemplates a possible outcome. Ideally, a decision theorist should be able to show that, if the agent is rational, then ipso facto the numbers would be sufficiently determined. Attempts to prove this sort of uniqueness have to spell out what they mean by “rational”, and usually do so by requiring the agent’s preferences to satisfy certain formal conditions of ‘consistency’ (see sect. 1.1 above). In most of these attempts, the bulk of the numerical work is performed by axioms that, like the last example from the list on p. xxii, involve probabilities. Roughly speaking, we watch the agent gamble. His willingness to risk a good is supposed to show how good the good is for him.

Rational decision theory has grown into a vast area of research. Practically all the ‘rationality axioms’ used for uniqueness proofs are open to doubt and continue to provoke heated discussion. By dropping, adding, or modifying an axiom here or an axiom there we can generate, and then study, different theories that more or less capture or vary the Bayesian idea. Situations in which several rational decision makers interact raise numerous special problems, and their study has become known as game theory. The theories are widely applied, too – in biology, for instance, as well as in economics, moral philosophy, action-theoretic semantics, military strategy, and technology assessment.

The spirit of Bayesian decision theory and some of the usual ‘rationality axioms’ give rise to a number of paradoxes that have received considerable attention: paradoxes of transitivity, Allais’s paradox, Newcomb’s paradox, and others. Best known, and perhaps most important, is the Prisoner’s Dilemma – a case in which preferentialist rationality seems to require each of two agents to do what he knows will, if they both do it, leave each of them worse off than if they had both performed an alternative action. Whether preferentialist rationality does indeed require this, and to what extent it would thereby discredit itself, are matters of contention.

3. Preference and Morality

We can now turn to the role of preferences in ethics. If preferentialism is right, there are even three such roles: the form, the content, and the foundation of morality. The form, since a morality is supposed to be a normative preference relation – see section 3.1 below; the content, since what counts in morals are the preferences of the affected parties – see section 3.2; and the foundation, since to tie morality to rationality we would have to show that people want to
do what they ought to do – see section 3.3. Let us have a brief look at each of these roles in turn.

3.1. **Normative Preferences and the Format of Morality**

Recall the second part of the preferentialist creed: people are moral if they do what will best satisfy everybody's preferences. Just like its rational cousin (see p. xxvii above), this moral claim has a structural in addition to a substantial aspect. Apart from its moral content, that is, it implies a certain conception of the format of morality: a morality ranks possible worlds (or outcomes or scenarios), saying, for any two of them, which is at least as good as the other; and what you ought to do is bring about what is, according to the ranking, the best world you can.

Such a ranking of outcomes – "morally at least as good as" – is sometimes called a normative preference relation. Its middle name is no coincidence. The formal requirements one might want the relation to obey are the same as those for the (weak) preferences of a decently rational individual: connexity, reflexivity, transivity (to quote just some of the examples from p. xxii). And the debates surrounding these requirements involve the same type of considerations (similar 'counterexamples' etc.) for the normative as for the individual case.

The picture of morality as a normative preference relation should hardly surprise us. Betterness looms large in pre-theoretical moral discourse anyway. Just think of questions like "Wouldn't it be better to let the woman decide by herself?"; and of the fact that "You ought to do what's best" is unlikely to count as a helpful piece of moral advice – it's just too close to a deontic tautology. Furthermore, given that morality ought to play a role in our rational decision making, the two of them had better speak the same language: if what a rational moral person goes by is her (or has the form of) preferences, then these must be moral preferences, and morality should specify them.

3.1.1. **Consequentialism.** The view that we ought to bring about the best world we can, frequently called (moral) consequentialism, has met with several objections. Many of them resemble objections to the type of rational consequentialism that we met in section 2.1 (especially on pp. xxix f.).

One complaint is that moral consequentialism, by declaring the optimum to be obligatory, makes morality too demanding. (This is a moral analogue of the claim that rational agents would satisfice rather than maximize.) Surely, the consequentialist will reply, morality can be demanding – but in what sense could it be too demanding? If a 'morality' were too demanding by moral standards (say, it requires me to give up my life for a goldfish), then we have got the
normative preference relation wrong, and all we have to do is work with the right one. And if it is too demanding just by the agent's standards, should we not call the agent immoral rather than withdraw the moral judgement?

To those who, nevertheless, uphold the complaint against demandingness, several remedies have been suggested. If you want to go on using normative preference relations, you could buffer their demands with some sort of threshold. You could, for instance, pick out one outcome and say that all the agent ought to do is realize an outcome at least as good as that one; it would be better of him, but not obligatory for him, to bring about one that is even better than that. Thus, the obligatory might fall short of the optimum. The difference, the good that is not obligatory, is the supererogatory. Concepts of supererogation can be quite sophisticated. One could, for example, ask how much is at issue for the agent and how much for morals, and then define the decisive threshold in terms of these two factors: what personal sacrifice can be demanded for what moral achievement?

Consequentialism has also been criticized for being self-defeating as a moral method. Having your goals in mind may put them out of reach. (How good, for instance, would a world be in which lovers were nice to each other because, and just as long as, they thought morality required it? And would "lovers" still be the right word?) More specifically, the fact that consequentialism requires us all to maximize according to the same normative preference relation has been suggested to cause at least part of the trouble. Such a universal preference relation would seem to hang like a sword of Damocles over our personal relationships. Support for your friend (or spouse, or child) would always be conditional upon the fact that you cannot, by redirecting your resources, do more good to somebody else.

Just like rational consequentialists, moral consequentialists will reply that in the end they know of no better method, and certainly of no better criterion. And what, they will ask, if we had a morality in which normative preference relations differed from agent to agent? Then sometimes people's correct moral preferences would clash – there would have to be fist fights even between saints. But a major point of having a morality is to render battles, at least among its followers, unnecessary.

Among the candidates for the formal constraints on a normative preference relation, connexity (or, as some people call it, completeness, or linearity) has received special attention: why should it be the case that every two possible worlds are comparable by moral standards?

Another important question is whether consequentialism might not be empty: is there any moral theory at all that could not be couched in consequentialist jargon? Take, as a simple example, a 'rule-oriented' theory consisting of
the one and only rule that you ought not to tell lies. Wouldn’t it be equivalent
to saying that the world in which you don’t tell a lie is better than that in which
you do, or that the more lies are told in a world the worse the world (or some­
thing along these lines)? It is true of every rule and every action that in the
possible world which the action will bring about the rule is either broken or
not; so why cannot whatever is important about rules be captured in compar­
ning possible worlds? Not all consequentialists would have to regret the empti­
ness of consequentialism; some of them would welcome the moral neutrality
of a claim they had always intended to be no real claim at all, but a formal
framework that could accommodate all sorts of claims.

3.1.2. Deontic Logic. Having presented normative preference relations,
we should add that they have come to dominate the semantics of deontic logic.
(Deontic logic is the logic of moral judgements.) The basic idea is that a model
of deontic logic is a set of possible worlds (to be looked at as the feasible set,
i.e. as the set of worlds that can be brought about in a given situation) with
an at-least-as-good-as relation ≥☐ over it; for any sentence A, “It ought to be
the case that A” is defined to be true in the model if, and only if, A is true in
all the ≥☐-maximal worlds (in the best feasible worlds, that is).

This said, it should not surprise us that many authors conceive of deontic
logic as identical with the logic of preference, viz. as the logic of normative
preference. What is different is not the logic as such, but, firstly, the intended
interpretation of its central operator: “It ought to be the case that” in the one
case, “Individual a wants it to be the case that” in the other. Secondly and
accordingly, the choice of preference relations in particular models: normative
relations in the one case, individual or subjective ones in the other.

Preference-based deontic logic generates the usual amount of real or al­
leged paradoxes, accompanied by a debate as to what modifications, if any,
they suggest. (Moral relatives of the so-called paradox of material implication,
for instance, have given rise to dyadic systems of deontic logic.) The formal
properties that the relation ≥☐ should be required to meet are controversial.
There is also the problem what the semantics should say if there are no ≥☐-maximal worlds in a model because it is infinite and the worlds keep getting
better and better. The paradigm is still struggling to capture notions like that
of a prima facie (or ceteris paribus, or intrinsic) obligation. Multi-modality is
another challenge; modal logics (e.g. of time, action, causation, or necessity)
have begun to interact on a large scale, and there is general consensus that de­
ontic logic, too, should be made to tie in with most of them.
3.2. Preference Satisfaction and the Content of Morality

From the format of morality, we will now turn to its content: what is it that makes actions right or wrong, or some worlds better than others? It is the welfare of the affected parties, many theories answer, and welfare (well-being, utility, and the like) should be spelled out in terms of preference satisfaction.

Not everybody believes in this answer. Some people would even deny its first half—the position, called welfarism, that welfare is all that counts. A beautiful landscape, some of them think, would make the world a better place even if no sentient being ever set eyes on it or were any better off for it. The value of life falls, according to some, in the same category; a life, they would say, ought to be continued even if nobody, not even the person who lives it, could be made to enjoy or prefer it. Immanuel Kant thought that criminals ought to be punished even if this didn't help anybody. Some theories tell us that while looking at people's welfare we should also ask whether they deserve it. Et cetera.

As usual with taxonomies, some theories will be borderline cases. What counts, it is sometimes held, is meeting people's needs, or respecting their rights. These positions can be seen either as versions of or as alternatives to welfarism; the question will depend, obviously, both on one's concept of welfare and on the explication of needs and rights one has in mind. A proponent of the latter two will be a special brand of welfarist, if, as has been suggested, they can be explicated in terms that are themselves components of the notion of welfare— for example, in terms of preference satisfaction.

3.2.1. Welfare and Preference Satisfaction. Suppose that we are indeed welfarists. We shall then have to explicate the somewhat hazy concept of welfare a little further. (In fact, this is something we'd have to do even if we thought that welfare was just one of several things that matter.) Either the concept is tied to people's states of mind, or it isn't. Since states of mind are difficult to define and to detect, let alone to measure, it is sometimes suggested that we leave them out of the definition. The idea is to work with 'objective' lists instead, lists of things that, if we are very lucky, can be characterized without direct reference to mental states; your welfare is a matter, say, of having a car, a house, a job, the right to vote, and so on.

It seems, however, that such a list is an adequate moral guide only since, and if, people enjoy or want the things, or the effects of the things, that are on it. (Why should we say that a job, for instance, would benefit somebody who, even if he fully represented to himself the option and its consequences, were indifferent to it?) But then all the real work is done by states of mind after all,
and we can as well say so. Only they can turn things into goods. The moral currency, then, is not jobs, but what people, in some sense of the expression, “get out of them”. There is no point in objective lists making measurement easier by measuring the wrong thing.

Suppose next, therefore, that we do want to link welfare to states of mind. The two big options are pleasure and preferences. A welfarist who limits welfare to pleasure— a moral hedonist, that is— must subscribe to some types of moral conclusions that many of us find embarrassing. According to hedonism, we would be doing people a favour if, without telling them in advance, we connected them to a pleasure machine for the rest of their lives. And somebody’s wishes would matter only to the extent that their satisfaction would make him feel good; his preferences, e.g., for what should happen after his death would have no moral weight whatsoever.

Those are among the reasons why many welfarists turn from pleasure to preferences. So is the idea (sometimes purely intuitive, sometimes grounded in further arguments) that rational decision making and morals should be concerned with the same substance. After all, prudence is about what is good for somebody, and what is good tout court is made up from what is good for people. If we believe in this sort of correspondence, then hedonism in morality would mean hedonism in rationality; we would have to say that it is, ceteris paribus, irrational to invest effort in the satisfaction of a desire if it is known that the desideratum will not be experienced. This sounds implausible. If you want the un-experienced desideratum to be the case, then why should rationality force you to ignore your preference?

Yet another argument for preferences is that from comparability and measurability. A welfarist will, in some sense or other, want to compare or quantify welfare. “Causing people great pain is worse than causing them mild pain”, is one of the things she will want to say. The most promising road to comparability seems to lead, even in the case of pleasures and pains, via the agent’s preferences: “Your possible headache $X$ is milder than your possible toothache $Y$”, we could say, “if, other things being equal, you would prefer having $X$ to having $Y$.” And remember from section 2.2 that rational decision theory might even allow us to proceed from somebody’s preferences between outcomes to numerical values representing how good the various outcomes are for him.

But the argument from comparability and measurability is problematic. Could not a person’s preference between certain pleasures or pains drastically fail to mirror what, by any reasonable standard, we would be prepared to call their comparative value? If this is so, then the value should not be defined in terms of the preferences. And as to the numbers that preference-based decision theory gives us, the ‘rationality axioms’ needed to generate them are disputable
Preferences – an Introduction

Even if they weren’t – we are still miles away from showing how these numbers could have any interpersonal significance. (How do we get numbers such that, if the value of outcome $a$ for you is 5, for me 3, we can be sure that we have captured a morally relevant sense in which you would get more out of $a$ than I would?) And since we want to be able to comment on choices between, say, giving the only coconut there is either to John or to Jim, interpersonal significance is crucial.

Now suppose that for an account of what welfare is the welfarist does indeed employ preferences. Should it be preferences only, or should she say that both pleasure and preferences count? Looking at the moral substance, she might lean towards the dual conception. Even then she could be well advised to unify her terminology. When the time for theory-making and comparing and measuring comes, monist accounts are easier to handle. So she could define pleasure into preference, saying that people want, other things being equal, their conscious lifetime to be as pleasant an experience as possible (see p. xxiii above). One thing this stipulation would entail is that genuine masochism is impossible; a “masochistic preference” (if we want to go on using this expression at all) could at best be a preference for a certain type of pleasure, but not for the opposite of pleasure.

Suppose, finally, that we emerge from all these controversies believing that what counts is preference satisfaction, and preference satisfaction only. This makes us moral preferentialists, or, as we could also put it, preference-welfarists. We still have to face the difficult questions raised in section 1: for precisely what concept of a preference do we claim that, ceteris paribus, the frustration of a preference would be a bad thing? And do we really want to claim it for all types of preferences, including, for example, the preferences of the dead, and malevolent preferences?

Furthermore, if we want to talk about quantities of preference satisfaction, what measure do we have in mind? And do we think that a satisfied preference is better than one that doesn’t exist in the first place? (The answer will inform our position on advertising, procreating, and the like.)

3.2.2. Distributing Preference Satisfaction. There is also the question what we ought to do with preference satisfaction: just maximize the amount of it? That is what preference utilitarianism bids us to do. On the one hand, there are good reasons for being a utilitarian. One of them is the idea (widely held, in different versions and for different reasons) that morality is rational empathy: qua being empathic, the moral person will acquire everybody’s preferences; qua being rational, she will try to maximize the satisfaction of her preferences, in-
cluding the acquired ones; therefore, the rational moral person will end up trying to maximize the satisfaction of everybody's preferences.

On the other hand, it is well known that utilitarianism can be cruel. Increasing the overall amount of satisfaction may require us to impose vast frustrations on each member of a minority (say by torturing them), because it might be the only way of making everybody from the majority a little better off. This may even be required if the minority is much worse off to begin with than the majority; keeping up a 'decent' minimum level of welfare for everybody is of no concern for the utilitarian. For him, the distribution of welfare is simply irrelevant.

Today, many welfarist alternatives to utilitarianism are trying to avoid these 'counter-intuitive' implications. Raising the total amount of welfare, they say, is at best one goal among others: providing minimum levels of welfare for everybody, helping those who are worst off, or distributing welfare equally.

At the same time, utilitarians continue to defend their doctrine. They claim, on various grounds, that the right social choice should resemble the rational individual choice, and that, since the latter is a matter of maximizing, so is the former. They also point out that there are infinitely many non-utilitarian welfarist formulae and that it is hard to see how we can make a non-arbitrary choice between them. Furthermore, quite a few of these formulae have their own share of counter-intuitiveness in that they violate certain attractive moral principles; it is very hard, for example, to reconcile egalitarian constraints with Pareto's principle (the claim that, if something is worse for nobody and better for at least one, then it is better). Since the rejection of utilitarianism is frequently based on moral intuitions, we are faced with the general question, too, how much moral authority we are willing to grant them. It is sometimes argued that we should not trust their verdict on science-fiction cases; and that for realistic cases a careful study (that takes into account all the risks, consequences, and alternatives) would show that utilitarianism simply fails to entail the counter-intuitive judgements it is usually accused of.

If we have a doctrine both about what welfare is and what should be done about it in principle, then applying it to real-life issues - such as starting wars, telling lies, or eating animals - is another piece of hard work. Omitting this step, as academic philosophy seems sometimes inclined to, would render the whole enterprise of ethics pointless. Furthermore, applications can shape theories just as much as vice versa. They can do so in one of two ways. If we take moral intuitions seriously, then a particular intuition about a real-life problem will have at least some say in the choice of our general moral principles.
And even if we take intuitions less seriously, applications can reveal blindspots in our theories. Recently, for example, debates on death, abortion, and procreation have sent moral philosophers back to the drawing board; discussion of these issues had revealed that the concept of welfare or of justice, or both, needed a major rethink.

3.3. Shared Preferences and the Foundations of Morality

So much for the form and the content of morality. Even if they had nothing to do with preferences, preferentialist rationality would still post preferences at the foundations of morals. This is so because a foundation for a morality is an argument to the effect that it is rational to do what the morality says one ought to do; thus, if rationality is a matter of preferences, so will be the foundations of morals: founding morals is showing that agents want to do what they ought to do (see also pp. xxviii f. above).

The task can be specified further if we also buy the preferentialist story as to what it is that agents ought to do: satisfy, as best as they can, everybody's preferences. Then to show of an agent that she wants to do what she ought to do is to show that she wants to satisfy everybody's preferences. So if we are preferentialists all the way, then to found morality is to show that an agent wants what others want.

Here is a selection of prominent approaches to the task:

- A few philosophers seem to believe that preferences are essentially universal. It is supposed to be true in virtue of the meaning of "preference" or "rational preference" that, if somebody (rationally) prefers something, you share his preference – at least if you know about it and are rational yourself.

- Others focus their attention on the type of agent who already knows that she wants to make moral judgements; the very meaning of these judgements, they say, links her preferences to those of others. One problem with this approach is why an agent should be ready to make moral judgements at all, and, more specifically, moral judgements in precisely the sense of the metaethical premisses that the argument leans on. If the answer needs to adduce empirical claims (like the one that, in the world as it is, the agent would be better off if she did), then we are left with a considerable loss of generality, and the approach collapses into a version of minimal morals (see the final item on this list).

- It is also common to point to what are called people's moral intuitions, or moral sentiments. If, for example, you 'perceive', or feel, starvation to be unjust, then this will coincide (perhaps by definition, perhaps just in the world as it is) with a wish of yours (at least a ceteris paribus wish) that
there be no starvation; and this wish of yours is one for the satisfaction of other people's wishes, viz. their wishes to survive. Assume that we can find a moral theory that embodies all such intuitions or sentiments of yours, and that we are sufficiently careful to use only data that are backed up by your desires. Then you will indeed want to do what the theory says you ought to do. Unfortunately, intuitions or sentiments differ from one person to the next. What we are likely to end up doing this way is to found myriads of moralities, incompatible with each other.

- Next, there is minimal morals. In essence, it asks to what extent even egoists have reasons for being moral. Though an egoist has no intrinsic preference for satisfying other people's preferences, doing so will sometimes be the most promising means to her ends. (For instance, when people whom she would frustrate could get back at her. Game theory is of great help in analysing motivational structures of this type.) Since minimal moralists marshal empirical facts to buffer the effects of egoism, their results will be limited to cases in which the relevant facts obtain.

These were examples of attempts to provide morality with a foundation. It is worth pointing out two standard mistakes in the discussion. One is to look at such attempts as competitors. The tacit assumption that, if one of them is right, the others have to be wrong, pervades large parts of the debate. It is, however, unwarranted, just like the claim that, of any two tools, one is better than the other. They needn't be. For some jobs, a hammer is better than a screwdriver, for others vice versa; this is why we have toolboxes. Ditto for the foundations of morals. Our task is to show for as many combinations as possible of agent, others, and desiderata that the agent wants what the others want. Maybe different arguments work for different combinations. Founding morals is like a collection: every contribution is welcome.

The second mistake is to blame an approach to the foundations of morals for the gaps it leaves: "Mary ought to do F, but system X gives her no reason to do so; hence, system X is inadequate." This is a frequent move, and there are two reasons why it is too rash. One is that, while system X might leave parts of the task unperformed, it might perform other parts of it, and that perhaps better than anybody else; it might be incomplete rather than mistaken. The other reason is that perhaps some gaps cannot be overcome by any argument. Maybe there are situations in which agents have no reasons to take other people's preferences seriously. This could be the sad truth.

Suppose it is. Suppose that some alleged obligation will be left dangling in the air; say, we cannot find a reason for Mary to do something she ought to do. What then?
We might still want to tie morality to preferences, or reasons, in a weaker sense. We might want to say, just like our second example of attempts to found morality did, that at least the person who acknowledges an obligation (subscribes to a moral judgement) must want, or have reasons, to meet it. Theories that say something along these lines are called internalist (as opposed to externalist) or prescriptivist (as opposed to descriptivist).

The question of internalism looms particularly large in the discussion of moral realism. Moral realists say that we perceive value, roughly in the same sense as we perceive colours; therefore, the question what ought to be done is cognitive just as much as the question whether tomatoes are red. Answers to it can be true or false, and somebody is handicapped (blind) if he doesn't experience the relevant data, and epistemically irrational if he does not process them 'properly'. However, if internalism is true — if, that is, accepting a moral judgement entails having reasons, or preferences, to act in accordance with it —, then it is very difficult to see how moral realism could be the whole story. Preferences, it would seem, can hardly be true or false.

But internalism is just one of many possibilities. More generally, there are, if we discover a dissonance between an alleged reason and an alleged obligation, three basic types of options. Either we take the lack of reasons to refute the alleged obligation: if somebody has no reason for doing something, then a morality saying that he ought to do it is inadequate. Or — vice versa, and less commonly — we take the alleged lack of reasons to refute the underlying idea of rationality: if somebody ought to do something, then a conception of rationality saying that he has no reason to do it is inadequate. Or we let the obligation dangle, and parts of our morality will be unfounded, parts of our rationality immoral: some people have obligations they have no reason to meet. (Even then, we can still require some connections, for example the internalist ones.)

The choice among the basic options and their combinations is controversial. In essence, it concerns the pecking order between rationality and morality. Where the two of them don't match, should the latter yield to the former, or vice versa, or should they stay out of each other's way?

Here our tour ends. We concentrated on the example of preferentialism. It should be obvious, however, that most of the questions we touched upon arise for most theories of rationality or morality. They certainly arise for every the-
ory that assigns any role at all to preference or welfare (even if – unlike preferentialism and welfarism – it assigns roles to other players as well). Though we mentioned dozens of problems and debates, there are at least as many that we did not mention. Still, the structure of the field is perhaps tolerably clear.

Let us recall some of the distinctions and some of the connections. Preferences can come up on different levels of practical reasoning. It is, for example, one thing to say that the format of morality is that of a normative preference relation; it is another thing to say that preference satisfaction is what the normative relation is all about (in other words, that preference satisfaction makes one world better than another); it is yet another thing to say that rationality is a matter of preference satisfaction; and so on. Bayesian decision theory, instrumentalism, consequentialism, (preference-)utilitarianism, welfarism, and the standard approach to deontic logic are all distinct from each other; it is quite possible that some of them are true, or acceptable, and others not.

Furthermore, the claim that, on any given level, preferences are not the whole story is distinct from the claim that they are not an important part of the whole story. And theories that employ the word “preference” may each define it as they please and can then require a critic to base his objections on their usage rather than on his.

From the differences to the connections. Take, for instance, the idea of intrapersonalization (mentioned on pp. xxxvii f.). That was the idea – intimately related to the well-known maxim to do to others as we wish others to do to us – that morality equals rational choice minus identity. In one version of it, the moral evaluation of outcomes is based on the question which of them a rational agent would choose if she knew how people live in them, but had no idea which of these lives she herself would have to live. If you endorse this type of criterion, then your theory of rational choice will determine the content of your morality, or vice versa, or both. Ditto if you think (following a claim we met in section 3.3) that people only have obligations that it would be rational for them to meet.

Other connections are harder to put into words. It seems that some combinations of answers are particularly concordant, in a rather aesthetical sense of the term. Consider the preferentialist standard package: Bayesian decision theory, consequentialism, preference-utilitarianism, deontic logic based on normative preference relations. The overall picture of practical reason that this package offers us is, this much has to be admitted, breathtakingly unified. It bristles with structural and substantial analogies between its components, and each part of it looks like the ‘natural’ or ‘obvious’ complement to all the others.

Is this an argument? Other packages, with radically different components, might look equally unified. And the sheer amount of unity could raise the
question whether the preferentialist system has simply bewitched us: maybe the vision has been too coherent to permit a clear judgement of its components. And maybe not.
This bibliography – not quite as short as we had once intended it to be – focuses on preferences (desires, etc.) in rationality and morals.

Nota bene: Important articles are usually not listed separately if they are contained in a collection that is listed. Contributions to this volume have not been listed. Comments that will enable us to correct or to update this bibliography are welcome.


Preferences – a Short Bibliography

Handbuch Philosophie: Ethik, Freiburg 1997; Peter Singer (ed.): A Companion to Ethics, Oxford 1991; Peter Singer (ed.): Ethik, Oxford 1994. Since dipping into the Beckers’ Encyclopaedia, into Schaber’s and Wolf’s Handbuch, into Singer’s Companion, and into The Encyclopaedia of Philosophy (ed. by Paul Edwards, London 1967) is always a good idea (e.g., s.v. “pleasure”, “desire”, “happiness”, “motive”, “practical reason”, “welfare”), contributions to these works have not been listed.

Other bibliographies: Some other bibliographies are listed at the appropriate places in this one. At least two more deserve special mention: volumes 9 and 10 in the series Study Aids, published by the Oxford Sub-faculty of Philosophy, are excellent and cover among other issues most of those that we do. Vol. 9 is James Baker: A Select Bibliography of Moral Philosophy (1977); vol. 10 is Susan L. Hurley, Jeff McMahan, and Madison Powers: A Select Bibliography of Moral and Political Philosophy (1987). The overlap between these two, as well as between them and this bibliography, is considerable in some areas, but so are the differences.

The bibliography is structured as follows:

1. The Concept of Preference
   1.1 The Logic of Preference
   1.2 Attitudes in General
   1.3 Preference in Particular: Representation and Charge
   1.4 The Taxonomy of Preferences

2. Preference and Rationality
   2.1 General Problems
   2.2 Rational Decision Theory

3. Preference and Morality
   3.1 Normative Preferences and the Format of Morality
   3.1.1 Consequentialism
   3.1.2 Deontic Logic
   3.2 Preference Satisfaction and the Content of Morality
   3.2.1 Preference Satisfaction and the Concept of Welfare
   3.2.2 Distributing Preference Satisfaction
   3.3 Shared Preferences and the Foundations of Morality
   3.3.1 General: Wanting What's Good (Internalism, Prescriptivity, “Why be Moral?”)
   3.3.2 Special: Wanting What Others Want
Preferences – a Short Bibliography

1. The Concept of Preference

1.1. The Logic of Preference

Since the boundary between decision theory and the logic of preference is blurry, section 2.2 is also of interest. And since many authors regard the logics of “want” and “ought” as identical, section 3.1.2 should be consulted as well.

Aristotle: *Prior Analytics*, 68a25–b7; this reference follows the standard system based on Bekker’s edition and indicated in most modern editions.

—: *Rhetoric*, book I.
—: *Topics*, book III.

Danielsson, Sven: *Preference and Obligation*, Uppsala 1968.

von Kutschera, Franz: *Einführung in die intensionale Semantik*, Berlin 1976, sect. 5.3.


1.2. Attitudes in General

Block, Ned (ed.): *Readings in Philosophy of Psychology*, 2 vols., Cambridge, Mass., 1980 (vol. 1) and 1981 (vol. 2). (On mental representation, see vol. 2, part I.)


Kemmerling, Andreas: “Mentale Reprasentationen”, *Kognitionswissenschaft* 1 (1990/91). (See also the exchange between the author and Ansgar Beckermann in vol. 2 (1991) of the same journal.)


1.3. Preference in Particular: Representation and Charge


first publ. in 1969). The volumes by Beckermann, Binkley et al., and Meggle all include splendid bibliographies.


von Brentano, Franz: Psychologie vom empirischen Standpunkts, 2 vols., Hamburg 1924 (vol. 1) and 1925 (vol. 2, second ed.); vol. 1 first publ. in 1874; second ed. of vol. 2 first publ. in 1911; first ed. of vol. 2 first publ. in 1874.
Egonsson, Dan: Interests, Utilitarianism and Moral Standing, Lund 1990. (Contains extensive and lucid discussions of the major options for explicating the concept of preference; mentions and considers many of the canonical writings.)
Green, Thomas Hill: Prolegomena to Ethics, third ed., Oxford 1890, book II; first ed. of Prolegomena first publ. in 1883.
Lewis, David: "Desire as Belief", *Mind* 97 (1988). (More papers on the same topic can be found in the same, the preceding and the subsequent volumes of *Mind*.)


Marks, Joel (ed.): *The Ways of Desire: New Essays in Philosophical Psychology on the Concept of Wanting*, Chicago 1986. (With a large bibliography.)


Nowell-Smith, P. H.: *Ethics*, Harmondsworth 1954, chs. 8 f.


Plato: *Republic*, book IV.


Prichard, H. A.: *Moral Obligation: Essays and Lectures*, Oxford 1949. (Contains several pieces on desires, the will, etc.)


Russell, Bertrand: *The Analysis of Mind*, London 1921, ch. 3.

Ryle, Gilbert: *The Concept of Mind*, London 1949, ch. III.


Seebass, Gottfried: *Wollen*, Frankfurt/Main 1993. (With, among other qualities, notes that contain a splendid number of references to, and discussions of, earlier sources, like Brentano, Gomperz, Lorze, Nietzsche, Reid, to name but a few.)


Preferences – a Short Bibliography


1.4. The Taxonomy of Preferences

Distinctions of types of preferences are usually motivated by specific problems, and are therefore scattered all over the literature. Most of the writings mentioned in section 3.2.1, for instance, contain some of them. The subject index of Zum moralischen Denken (ed. by Christoph Fehige and Georg Meggle, 2 vols., Frankfurt/Main 1995) lists types of preferences in abundance (s.v. "Präferenzen"); numerous debates, definitions, and writings can be spotted with its help. Many major distinctions are discussed at some point in the following works:

Bentham, Jeremy: A Table of the Springs of Action, Oxford 1983; first printed in 1815; first publ. in 1817.
Heil, John (ed.): Rationality, Morality, and Self-Interest: Essays Honouring Mark Carl Overvold, Lanham, Md., 1993. (See esp. the contributions by Brad Hooker and Alan E. Fuchs.)
2. Preference and Rationality

2.1. General Problems

Consequentialism: See section 3.1.1.

Psychology, free will, and the theory of action: See the beginning of section 1.3; ditto (paragraph on the theory of action) if you're looking for material on the role of preferences in the rational explanation, rather than justification, of actions.


Aristotle: The Nicomachean Ethics.
Harrison, Ross (ed.): Rational Action: Studies in Philosophy and Social Sciences, Cambridge 1980. (See esp. the papers by Williams, i.e. the much-quoted "Internal and External Reasons", and by Wollheim.)
Nozick, Robert: The Nature of Rationality, Princeton 1993, ch. V.


Raz, Joseph (ed.): *Practical Reasoning*, Oxford 1978. (With a good selected bibliography.)


### 2.2. Rational Decision Theory

Since the boundary between decision theory and the logic of preference is blurry, section 1.1 should be consulted as well. As to "utility" and its measurement, see also section 3.2.1. Game-theoretical writings with an immediate bearing on the rational foundation of morals are mostly listed in section 3.3.


Bell, David E., Ralph L. Keeney, and Howard Raiffa (eds.): *Conflicting Objectives in Decisions*, Chichester 1977.


the axiomatic derivation of numerical utility appears first in the second ed., of 1947).


3. Preference and Morality

3.1. Normative Preferences and the Format of Morality

3.1.1. Consequentialism


### 3.1.2. Deontic Logic

Since many authors regard the logics of "ought" and "want" as identical (for example, because they base their deontic logic on a normative preference relation), works on the logic of preference (cf. sect. 1.1) should be consulted as well. Much of the current work in deontic logic is presented at the biannual *Workshop on Deontic Logic in Computer Science*, and is documented in the corresponding proceedings; so far these are Meyer/Wieringa (1993), Jones/Sergot (1993), and Brown/Carmo (1996).


di Bernardo, Giuliano (ed.): *Logica deontica e semantica*, Bologna 1977. (With an enormous bibliography.)

3.2. Preference Satisfaction and the Content of Morality

3.2.1. Preference Satisfaction and the Concept of Welfare

Since the discussions of what welfare is and how it should be distributed are often intimately connected, section 3.2.2 is also of interest. So is section 2, since an agent's
rationality is frequently taken to be accountable to her welfare (utility, preference satisfaction, or the like).

Psychologists and social scientists: They, too, grapple with the concept and the measurement of welfare. To pick up the thread in psychology, see, for instance, Argyle (1987), Beebe-Center (1932), and Veenhoven (1984) and (1994). In sociology, the pertinent area is “social indicators research”; for problems, approaches, and literature, see Baldwin et al. (1990), Noll/Zapf (1994), and Zapf (1984).

Allais, Maurice, and Ole Hagen (eds.): Cardinalism: A Fundamental Approach, Dordrecht 1994. (With a long paper by Tore Ellingsen on the history of hedonimetry.)


—- (ed.): Value and Obligation, New York 1961, part I.


DeGrazia, David: Taking Animals Seriously: Mental Life and Moral Status, Cambridge 1996, chs. 6 and 8. (Of general interest, not just related to animals.)
Frey, R. G.: Interests and Rights: The Case Against Animals, Oxford 1980. (Includes numerous discussions that are not limited to animals.)
Green, Thomas Hill: Prolegomena to Ethics, third ed., Oxford 1890, books III ff.; first ed. of Prolegomena first publ. in 1883.
Knight, Frank Hyneman: The Ethics of Competition and Other Essays, New York 1935, essays I–III, esp. essay II.
Lewis, Clarence Irving: An Analysis of Knowledge and Valuation, La Salle, Ill., 1946, esp. chs. XIII and XVI.


Miller, Harlan B., and William H. Williams (eds.): *The Limits of Utilitarianism*, Minneapolis 1982, sect. III. (See esp. Richard B. Brandt’s paper.)


Narveson, Jan: *Morality and Utility*, Baltimore 1967, ch. III.


Parker, Dewitt H.: *An Interpretation of Ethics Based on a Study of Values*, New York 1931, esp. chs. II, V, and VII, and app. I.


Raitt, Peter: "Facts and Values", *Philosophical Topics* 14 (1986).

—: “Fairness to Goodness”, *Philosophical Review* 84 (1975).
—, and Bernard Williams (eds.): *Utilitarianism and Beyond*, Cambridge 1982.
Veenhoven, Ruut: *Conditions of Happiness*, Dordrecht 1984. (Mostly empirical.)


3.2.2. Distributing Preference Satisfaction

Problems of aggregation are closely connected, firstly, to problems of comparability and measurement (therefore, section 3.2.1 should be consulted as well); secondly, via the foundation of morals, to sharing, or acquiring, other people's preferences (therefore, section 3.3.2 should be consulted as well).


Ewing, A. C.: Ethics, London 1953, ch. III.


Fechner, Gustav Theodor: Über das höchste Gut, Leipzig 1846, ch. XI.

Fehige, Christoph, and Georg Meggle (eds.): Zum moralischen Denken, 2 vols., Frankfurt/Main 1995.


Höffe, Otfried (ed.): Einführung in die utilitaristische Ethik, second ed., Tübingen 1992. (Mostly translations, which are normally not listed here; we mention the book because of its useful bibliography.)


von Kutschera, Franz: Grundlagen der Ethik, Berlin 1982, sect. 4.3.


Mason, Andrew (ed.): Ideals of Equality, special issue of Ratio 10 (1997; new series).


Miller, Harlan B., and William H. Williams (eds.): The Limits of Utilitarianism, Minneapolis 1982.

Narveson, Jan: Morality and Utility, Baltimore 1967, ch. VII.
Preferences – a Short Bibliography


— —: *Inequality Reexamined*, Oxford 1992. (Also a good bibliographical guide, see esp. the enormous footnote on p. 93.)


del Vecchio, Giorgio: *La Giustizia*, second ed., Rome 1951; first ed. first publ. in 1946, proto-version publ. in 1923. (Helpful for the history of the idea of justice.)


3.3. Shared Preferences and the Foundations of Morality

3.3.1. General: Wanting What's Good (Internalism, Prescriptivity, "Why Be Moral?")

Does an agent have to want to do what (she thinks) she ought to do? (Or, similarly, want the things to be the case that (she thinks) ought to be the case?) This is the general type of question, often discussed under headings like "internalism", "prescriptivity", "Why be moral!", or "the rational foundation of morality". Since the title of section 3.3.2 can be seen as a special version of it (one that arises if we equate what ought to be the case with the satisfaction of everybody’s preferences), the items mentioned there should be consulted as well.


Ewing, A. C.: *Ethics*, London 1953, ch. II.


Foot, Philippa: *Virtues and Vices and Other Essays in Moral Philosophy*, Berkeley 1978.


———: Freedom and Reason, Oxford 1963, esp. part I.
Hooker, Brad (ed.): Truth in Ethics, special issue of Ratio 8 (1995; new series).
Little, Margaret: "Recent Work on Moral Realism", Philosophical Books 35 (1994), parts I and II.
Lumer, Christoph: Praktische Argumentationstheorie, Brunswick 1990.
——— (ed.): British Moralists 1650-1800, 2 vols., Oxford 1969. (See esp. the extracts from Hutcheson and Shaftesbury.)
Regis, Edward, Jr. (ed.): Gewirth's Ethical Rationalism: Critical Essays with a Reply by Alan Gewirth, Chicago 1984. (See esp. the papers by Richard M. Hare and Kai Nielsen.)


### 3.3.2. Specific: Wanting What Others Want

If it is good to satisfy everybody's preferences, then the general question from section 3.3.1 (viz., Does the agent want what is good?) gives rise to a more specific one: Does the agent want what others want (or, similarly, others to get what they want)?

Thus, the literature on the more general question from 3.3.1 should be consulted as well.


———: “Could Kant have been a Utilitarian?”, Utilitas 5 (1993).


Hobbes, Thomas: Leviathan, Harmondsworth 1968; first publ. in 1651.


———: An Enquiry Concerning the Principles of Morals, Oxford 1975; first publ. in 1751.

Kant, Immanuel: Grundlegung zur Metaphysik der Sitten, vol. 4 of the Royal Prussian Academy’s edition Kant’s gesammelte Schriften, Berlin 1911; Grundlegung first publ. in 1785.


Plato: Republic.


Scheler, Max: Wesen und Formen der Sympathie, third ed., Bonn 1985; first ed. first publ. in 1913; third ed. first publ. in 1926.


