Theory and Decision Library A: Rational Choice in Practical Philosophy and Philosophy of Science

John R. Welch

Moral Strata

Another Approach to Reflective Equilibrium

Springer

Theory and Decision Library A:

Rational Choice in Practical Philosophy and Philosophy of Science

Volume 49

Series editor
Julian Nida-Rümelin
Universität München, Munich, Berlin, Germany

This series deals with practical and social philosophy and also foundational issues in philosophy of science in general that rely on methods broadly based on rational choice. The emphasis in the Series A is on well-argued, thoroughly analytical and philosophical rather than advanced mathematical treatments that use methods from decision theory, game theory and social choice theory. Particular attention is paid to work in practical philosophy broadly conceived, the theory of rationality, issues in collective intentionality, and philosophy of science, especially interdisciplinary approaches to social sciences and economics.

More information about this series at http://www.springer.com/series/6616

John R. Welch

Moral Strata

Another Approach to Reflective Equilibrium

Springer

John R. Welch
Saint Louis University – Madrid Campus
Madrid
Spain

ISSN 0921-3384 ISSN 2352-2119 (electronic)
ISBN 978-3-319-08012-3 ISBN 978-3-319-08013-0 (eBook)
DOI 10.1007/978-3-319-08013-0
Springer Cham Heidelberg New York Dordrecht London

Library of Congress Control Number: 2014945757

Printed on acid-free paper

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

These heavy sands are language
—James Joyce, Ulysses

Preface

This book began an embarrassingly long time ago. In looking for discussion material for an ethics class, I chanced across my copy of William Styron's *Sophie's Choice*. The work's central dilemma intrigued my students, who concluded that an act-utilitarian and a Kantian using the second formulation of the categorical imperative would make very different decisions about the case. The question was then unavoidable: Which theory offers the better ethical advice? I was dissatisfied with my own attempt to answer.

Meanwhile, I had become accustomed to responding to students' contrasts of science and ethics by citing Alan Gewirth's "Positive 'Ethics' and Normative 'Science,'" which warns of the fallacy of disparateness: "the fallacy of discussing one field on one level or in one respect and the other field on a quite different level or in a quite different respect." To contrast a scientific discovery such as the molecular structure of DNA with an ethical discussion of the highest good is an instance of this fallacy. Yet I was unable to be clearer about Gewirth's levels or respects until I discovered Larry Laudan's analysis of scientific discourse in *Science and Values*. In reading the work, I formed the hypothesis that moral discourse, like scientific discourse, could be analyzed into factual, methodological, and axiological levels. This hypothesis gained momentum when, in the book's Epilogue, Laudan himself mentioned the possibility of extending his approach to moral theory.

Rationality in morality, I thought, appears to be governed by the cognitive ideal of reflective equilibrium among levels of moral discourse analogous to Laudan's levels of scientific discourse. I proposed this ideal in "Science and Ethics: Toward a Theory of Ethical Value," which can be seen as a kind of mission statement for this book. But the article offered only the sketchiest indications of how reflective equilibrium might be attained at each of these levels, and very difficult technical problems lay half-submerged in each case. My attempts to resolve these problems led to explorations of quantitative inductive logics and comparative decision theory.

In the course of these explorations, I began to see what I take to be rational grounds for choice among theories. Rival ethical theories can offer conflicting advice about dilemmas, and quantitative inductive logics can be used to resolve a common sort of dilemma. Whenever this occurs, any theory that recommends the inductively preferred option secures an advantage over theories that recommend

other options. In addition, decision theory can be employed to guide the choice between one theory and another, particularly when formulated in terms of comparative plausibilities and utilities. This book presents such a version of decision theory, offered in the hope that it will aid in the quest for reflective equilibrium.

Quotation marks in this work are handled as follows. Double quotes are employed for short quotations, whether attributed or not, and quotations within long quotations. No quotes are used for long quotations, which are set off from their context by indents and smaller type. Single quotes are used for quotations within short quotations and words cited as words, such as the predicate ‘just’.

The development of the outlook presented in this volume was facilitated by interactions with many people: students, colleagues, conference participants, anonymous reviewers, and editors, among others. These interlocutors are too numerous to be listed individually, but I cannot fail to mention my parents, the late Mary V. and Robert J. Welch, who showed by example the centrality of morality to human life. Nor could I omit my colleagues Renzo Llorente, Olga Ramírez Calle, and Jawara Sanford, who commented insightfully on sundry parts of the manuscript. Talented people at Springer who played vital roles in this project include Associate Editor Lucy Fleet, Assistant Editor Martin Rechenauer, and Senior Editorial Assistants Diana Nijenhuijzen and Mireille van Kan. Finally, my wife Cristina and son Guillermo formed the uniquely supportive environment that enabled this work to be completed. Each, in different ways, has helped me through this project. It is an unmixed pleasure to thank you all.

Contents

Chapter 1
Discursive Strata

Abstract Chapter 1 introduces the strata that structure this work: phenomenal, instrumental, and teleological moral discourse. After an overview of the approach to moral strata presented in this volume, the chapter offers a quick reprise of the method of reflective equilibrium as elaborated by Goodman, Rawls, and Daniels. It then considers five objections to this method in its canonical formulations. Objections couched in terms of moral conservatism, moral diversity, and the moral weight of considered judgments are judged unsuccessful, while objections based on the nature of considered judgments and the relation to intuitionism are found to be more problematic. In order to meet these last two objections, the chapter reworks the received view of reflective equilibrium by defining an alternative notion of wide reflective equilibrium. This alternative is presented as a cognitive ideal: coherence among phenomenal, instrumental, and teleological discursive strata in addition to background theories. How moral discourse might achieve coherence of this sort is the subject of successive chapters.

1.1 Moral Strata

Where goals are concerned, the clearer the better. This book is an attempt to clarify the goal of moral inquiry. It undertakes this project in two stages. Its opening chapter proposes that the goal of moral inquiry is a specific kind of reflective equilibrium. Subsequent chapters then suggest ways of reaching (or at least approximating) this goal.

To specify the requisite form of reflective equilibrium, the book invokes the notion of discursive strata. In any field, discursive strata are formed by sentences and differentiated by function. Sentences that perform the same kind of linguistic function clump together, so to speak, to form a stratum. Where the field is moral inquiry, three linguistic functions appear to be central. In the morally phenomenal stratum, predicates like ‘dishonest’, ‘loyal’, ‘cowardly’, ‘generous’, and ‘cruel’ are applied to actions, policies, and persons; ‘That took courage’ is an instance. The morally instrumental stratum concentrates on means for attaining moral ends, as in ‘Justice requires an independent judiciary’. The correlative stratum is morally teleological, formed by sentences that focus on moral ends like ‘Seek the greatest happiness of the greatest number’.

J. R. Welch, *Moral Strata*, Theory and Decision Library A:, 49,
DOI 10.1007/978-3-319-08013-0_1,

The book's opening chapter recruits these strata to delineate the appropriate form of reflective equilibrium. The result is a version of wide reflective equilibrium here proposed as a goal for moral inquiry. How this goal might be achieved, stratum by stratum, is the subject of the rest of the work.

Chapter 2 addresses the phenomenal stratum. Its initial step is to introduce a fundamental form of classification called 'core classification'. To develop this concept, the chapter formulates the analogy thesis: core classification in general, and moral core classification with terms like 'honest' in particular, is carried out by analogy. In addition, the chapter holds that good analogies can be distinguished from bad ones by appeal to a standard of inductive cogency. This standard is specified with the help of quantitative inductive logics in the tradition of Carnap, Hintikka, Kuipers, and Niiniluoto. The chapter shows how the use of these logics can reduce vagueness and provide an in-principle solution to morally phenomenal disagreements. It illustrates these claims through an extended discussion of the moral dilemma faced by Cicero's grain merchant.

Chapter 3 postpones the treatment of instrumental and teleological moral strata in order to marshal technical resources needed in subsequent chapters. The chapter opens with a survey of four approaches to theory choice. The result of this survey is an endorsement of decision theory as a guide to theory choice. However, decision theory has a serious problem of numeric poverty: standard applications of decision theory require point-valued utilities (or point-valued probabilities and utilities), but we rarely have precise and reliable values for these inputs. Consequently, this chapter pleads for a more widely applicable form of decision theory. It proceeds to argue that a comparative version of decision theory fills the bill. It shows that many choices among theories can be based on merely comparative plausibilities and utilities.

Chapter 4 analyzes the instrumental stratum of moral discourse. Since this stratum is composed of substrata that can be demarcated in different ways, the chapter considers three different groupings: individual sentences, inferences, and theories. It contends that individual sentences can be confirmed or disconfirmed through scrupulous observation and inductive logic. The chapter also treats the justifiability of practical inferences, proposing that they be evaluated by the standard of inductive cogency introduced in Chap. 2. Finally, the chapter grapples with the issue of moral theory choice. It holds that a moral theory can be chosen on instrumental grounds by applying the comparative decision theory of Chap. 3. To illustrate the procedure, it undertakes a comparative evaluation of Kantian, Benthamite, and Frankenian theories as applied to Sophie's choice.

Chapter 5 investigates the teleological stratum of moral discourse. It distinguishes teleological descriptions such as 'The highest good is the greatest happiness of the greatest number' from teleological directives like 'Act from the good will'. The chapter maintains that teleological descriptions can be viewed as hypotheses. As such, they can be confirmed or disconfirmed through hypothetico-deductive reasoning analogous to that employed in the sciences. In addition, the chapter shows that choice among teleological directives can be reasonably guided by the comparative decision theory of Chap. 3. Finally, the chapter ponders the possibility of justifying the higher-order end of being moral. It urges that coherence requires accepting

'I ought to be moral', where coherence is understood in terms of wide reflective equilibrium.

The concluding Chap. 6 proposes remedies for a common affliction: reflective disequilibrium. This affliction can result from inconsistencies within moral strata or between moral and nonmoral discourse. The chapter claims that reflective disequilibrium within the phenomenal stratum can be relieved in some cases through recourse to the standard of inductive cogency defined in Chap. 2. Reflective disequilibrium within the instrumental stratum may be resolved intra-theoretically, as illustrated by the application of Frankenian theory to the case of *United States v. Holmes*, or finessed inter-theoretically through ascent to the teleological stratum of moral discourse. Reflective disequilibrium within the teleological stratum can often be reduced inter-theoretically with the help of comparative decision theory or intra-theoretically on consequentialist grounds. Finally, disequilibrium between moral and nonmoral discourse can be reduced by hewing to a modest version of the overridingness thesis.

1.2 Origins of Reflective Equilibrium

C. S. Peirce famously distinguished two forms of thought: "Thought in action has for its only possible motive the attainment of thought at rest" ([1878] 1986, p. 263). Thought in action is the activity of inquiry; thought at rest is inquiry's goal. Thought at rest, according to Peirce, is settled belief. So inquiry aims at settled belief.

The view that the goal of inquiry is settled belief has been held in one form or another by a number of recent philosophers, notably those influenced by Nelson Goodman.[1] Its best-known restatement is John Rawls' concept of reflective equilibrium, routinely described as the most widely used method in contemporary ethics (e.g., Varner 2012, p. 11). Attaining reflective equilibrium, according to Rawls, requires the "mutual adjustment of principles and considered judgments" (1971, p. 20 n. 7). Even though this description is rooted in Rawls' discussion of justice, he remarks at least twice that the process of mutual adjustment of principles and considered judgments is "not peculiar to moral philosophy" (1971, p. 20 n. 7, 49), and he cites Goodman's observations about the justification of general rules and particular inferences in both deductive and inductive logic: "The process of justification is the delicate one of making mutual adjustments between rules and accepted inferences; and in the agreement achieved lies the only justification needed for either" (Goodman 1979, p. 64).

The process of justification Goodman refers to is not peculiar to logic. Russell and Whitehead, for example, make cognate remarks about justification in mathematics:

[1] For example, Rawls (1971, p. 20, 46–51), Putnam (1983, pp. 201–202), and Elgin (1989, p. 91, 94; 1996).

> [T]he chief reason in favour of any theory on the principles of mathematics must always be inductive, *i.e.* it must lie in the fact that the theory in question [which enunciates principles] enables us to deduce ordinary mathematics [considered mathematical judgments]. In mathematics, the greatest degree of self-evidence is usually not to be found quite at the beginning, but at some later point; hence the early deductions, until they reach this point, give reasons rather for believing the premisses because true consequences follow from them, than for believing the consequences because they follow from the premisses. (Russell and Whitehead 1927, vol. I, p. v)

Similarly, Vann McGee draws the methodological parallel between justification in logic and justification in the empirical sciences:

> The methodological moral to be drawn from this [putative counter-examples to modus ponens] is that, when we formulate general laws of logic, we ought to exercise the same sort of caution we exercise when we make inductive generalizations [empirical principles] in the empirical sciences. We must take care that the instances [considered empirical judgments] we look at in evaluating a proposed generalization are diverse as well as numerous. (1985, p. 468)

Hence this process of mutual adjustment of principles and considered judgments goes on in moral philosophy, logic, the foundations of mathematics, and the empirical sciences. It is not limited to these contexts, however. It appears to occur wherever rational inquirers employ principles and considered judgments to inquire.

The goal of reflective equilibrium can take more than one form. Rawls' distinction between narrow and wide reflective equilibrium is present in all but name in *A Theory of Justice* (1971, p. 48, 49), and Rawls began to counterpose the two forms explicitly not long after the publication of his best-known work (e.g., 1975, p. 8, 21). Suppose that we undertake a process of mutual adjustment of our principles and considered judgments; if the end result is coherent, we have reached a state of narrow equilibrium. The critical edge of this process is likely to be dull, however. Though we may reject a considered judgment here and revise a principle there, the process is meant to salvage as much of our belief structure as possible. Suppose, on the other hand, that we attempt to subject the entire structure of relevant beliefs to philosophical scrutiny, contrasting our native point of view with alternative conceptions. In such cases,

> we are interested in what conceptions people would affirm when they have achieved wide and not just narrow reflective equilibrium, an equilibrium that satisfies certain conditions of rationality. That is, adopting the role of observing moral theorists, we investigate what principles people would acknowledge and accept the consequences of when they have had an opportunity to consider other plausible conceptions and to assess their supporting grounds. Taking this process to the limit, one seeks the conception, or plurality of conceptions, that would survive the rational consideration of all feasible conceptions and all reasonable arguments for them. We cannot, of course, actually do this, but we can do what seems like the next best thing, namely, to characterize the structures of the predominant conceptions familiar to us from the philosophical tradition, and to work out the further refinements of these that strike us as most promising. (Rawls 1975, p. 8)

Hence wide reflective equilibrium, as Rawls conceives it, emerges only if we escape the narrow circle of our own favored principles and considered judgments and confront them with the widest possible range of alternative conceptions. As Norman

Daniels observes, "We must show why it is reasonable to hold these principles and beliefs, not just that we happen to do so" (1996, p. 1).

Daniels stresses the importance of wide reflective equilibrium for moral theory, but he rejects a "two-tiered view of moral theories" that relies on a set of considered judgments and a set of principles (1979, p. 256). Instead, he proposed a more complex view of moral theories that registers the importance of background theories. For example, Rawls' conclusions about justice as fairness are derived with the help of a set of background theories that includes "a theory of the person, a theory of procedural justice, general social theory, and a theory of the role of morality in society (including the ideal of a well-ordered society)" (Daniels 1979, p. 260). Hence we want coherence not only among our considered judgments and principles; we also want both to cohere with our background theories. Consequently, Daniels expanded the definition of wide reflective equilibrium as follows: "The method of wide reflective equilibrium is an attempt to produce coherence in an ordered triple of sets of beliefs held by a particular person, namely, (a) a set of considered moral judgments, (b) a set of moral principles, and (c) a set of relevant background theories" (1979, p. 258).

1.3 Problems with Reflective Equilibrium

Like other cognitive ideals, the ideal of reflective equilibrium has been challenged. In this section I want to air five of these challenges. The discussion aims to strike a sort of Aristotelian mean between too much and too little. To say too much would be to shoehorn the rest of the book into this one preliminary section, for one way to look at the book as a whole is as a defense of wide reflective equilibrium. But to say too little would leave the reader with the sense that substantive objections to this cognitive ideal have simply gone unheeded.

1.3.1 Moral Conservatism

One complaint about the method of reflective equilibrium is that it is uncritically conservative—a volley returned in the opposite direction by the method's proponents.[2] An early advocate of this point of view is Richard Brandt, who worries that the coherence proper to the method "may be no more than a reshuffling of moral

[2] Rawls targets utilitarianism in particular: "the [utilitarian] choice [of "ideals of the person"] does depend upon existing desires and present social circumstances and their natural continuations into the future." By contrast, justice as fairness and perfectionism "establish independently an ideal conception of the person and of the basic structure so that not only are some desires and inclinations necessarily discouraged but the effect of the initial circumstances will eventually disappear" (1971, p. 262). Daniels also claims that "utilitarianism is biased toward the status quo" (1996, p. 94).

prejudices" (1979, p. 22). More recently, Gilbert Harman observes "The method is conservative in that we start with our present views and try to make the least change that will best promote the coherence of our whole view" (2003, p. 416). Indeed, Allen Wood rejects the method because some of our beliefs are corrupted by "the radical evil of our social condition," which would contaminate any equilibrium that retains such beliefs (2008, p. 5).

These concerns appear to be valid for the narrow variety of reflective equilibrium but not for the wide. Wide reflective equilibrium requires the widest possible reflection, which includes all the critical resources we can bring to bear. Ron Amit has compared Rawlsian method to the Frankfurt School's method of immanent criticism, arguing that it holds out "an idealized and improved mirror against which we confront our own institutions" (2006, p. 182). In addition, Katarzyna de Lazari-Radek and Peter Singer point out that commitment to wide reflective equilibrium includes commitment to scientific beliefs, which means that evolutionary theory could be used "to reject many widely shared moral intuitions" (Lazari-Radek and Singer 2012, p. 30). Consequently, the method does not appear to be inherently conservative.

1.3.2 Moral Diversity

Some of these same critics charge that moral diversity derails the quest for reflective equilibrium. Singer puts the point as follows:

> If I am right in attributing this version of the reflective equilibrium idea to Rawls, then Rawls is a subjectivist about morality in the most important sense of this often-misused term. That is, it follows from his views that the validity of a moral theory will vary according to whose considered judgments the theory is tested against. There is no sense in which we can speak of a theory being objectively valid, no matter what considered moral judgments people happen to hold. If I live in one society, and accept one set of considered moral judgments, while you live in another society and hold a quite different set, very different moral theories may be "valid" for each of us. There will then be no sense in which one of us is wrong and the other right. (1974, p. 494)

Brandt objects along similar lines:

> Moreover, moral intuitions differ from one individual or culture to another. Where one person thinks promise-keeping or sexual taboos are highly important—these beliefs have high initial credence level—and another does not, the search for reflective equilibrium will only produce different moral systems, and offers no way to relieve the conflict. Nor is this matter trivial. Moral disagreement does not exist only between our own reflective equilibria and those of some primitive tribes, or on relatively superficial matters. It exists among sophisticated civilized persons and in core areas…. (1979, p. 22)

Singer and Brandt are actually making four interrelated points: there are different considered judgments; because there are different considered judgments, there are different moral systems; different moral systems can be in reflective equilibrium; and reflective equilibrium will not relieve conflicts among moral systems. I will comment briefly on each in turn.

The claim about differing considered judgments is an empirical point, of course. Brandt (1954) contributed to research on this subject, and I think it likely that, in some carefully qualified sense, the claim is true. Some differences among considered judgments may be attributable to error by one or more parties; others may not. Errors that affect considered judgments would not create methodological problems, for reflective equilibrium requires correcting considered judgments when necessary. But if differing considered judgments cannot be explained away as errors, the proponent of reflective equilibrium will simply accept the differences. Accordingly, the analysis of morally phenomenal discourse in Sect. 2.3.2 explicitly recognizes the possibility of moral prototypes that are culturally variant. I will not assume it impossible for one culture to stress moral properties that are minimized or ignored by another.

However, if two cultures do in fact have different considered judgments, different moral systems would result, just as Singer and Brandt say. Indeed, different considered judgments that did *not* lead to different moral systems would require strenuous explanation. Note, however, that 'different' here means 'non-identical', not 'non-overlapping'. Careful analysis may show that different sets of equilibrium beliefs nevertheless overlap. In fact, this is what the historical record would lead us to expect. It shows substantial areas of moral agreement and narrowly delimited areas of moral disagreement, typically over exceptions to agreed-upon moral principles (Rachels 1986, pp. 19–22).

Different sets of moral beliefs, Singer and Brandt imply, can be in reflective equilibrium.[3] I suggest, once again, that they are right; "the prospects of divergence in wide reflective equilibrium remain significant" (Daniels 1996, p. 8). But this does not show that there is something wrong with reflective equilibrium. On the contrary, it shows that there is something right with it. Reflective equilibrium is a criterion of justification, not a criterion of truth (Daniels 1979, p. 277; Knight 2006, p. 220). As such, it is fully consistent with the fact that beliefs can be justified yet nonetheless false. Just as we can say that different historical beliefs about the shape of the earth, say, can be justified without committing ourselves to more than one geological truth, we can recognize that different moral beliefs can be justified without committing ourselves to more than one moral truth. That different sets of moral beliefs can be in reflective equilibrium is just what we would expect from a reliable criterion of justification.

Finally, we have Singer's and Brandt's point that reflective equilibrium would provide no way to relieve conflicts among different moral systems. Take people in a culture with a collective belief that human sacrifice is necessary to prevent the world from being destroyed, for instance. Thinking people in such a culture might attain reflective equilibrium provided they lack the empirical resources that would show the underlying cosmological belief to be false. Imagine such people being confronted by contemporaries from a second culture who object to human sacrifice on the grounds that it is immoral and causally independent of the world's survival.

[3] Rawls himself mentions this possibility but dismisses it for being "far beyond our reach" (1971, p. 50).

Would there be a way to relieve the conflict? In principle, yes: halt the sacrifices and watch what happens. The watching—short or long—should settle the conflict. Here, then, is a counter-example to the claim that there is no way to relieve conflicts among different moral systems. Obviously, though, this in-principle solution might not work in practice; people in the first culture might refuse to pursue this solution out of fear. I suggest, then, that blanket statements that there is or is not a way to relieve such conflicts are just too coarse. I hope to show, however, that when conflicts are identified as morally phenomenal, instrumental, and teleological, resolution can be sought in rational ways.

The upshot, I think, is that even though much of what Singer and Brandt say is true, none of it spoils the party for reflective equilibrium. Suitably understood, reflective equilibrium is not committed to unanimity of considered judgments, uniqueness of reflective equilibria, or universal resolution of conflicts among moral systems. But the pursuit of reflective equilibrium can, I claim, reduce or eliminate moral conflict in a rational way. The defense of this claim is the work of later chapters.

1.3.3 The Moral Weight of Considered Judgments

The moral weight that reflective equilibrium grants to considered judgments has been criticized. Singer, for example, suggests that it would be better to ignore them and work with moral axioms instead:

> Why should we not rather make the opposite assumption, that all the particular moral judgments we intuitively make are likely to derive from discarded religious systems, from warped views of sex and bodily functions, or from customs necessary for the survival of the group in social and economic circumstances that now lie in the distant past? In which case, it would be best to forget all about our particular moral judgments, and start again from as near as we can get to self-evident moral axioms. (1974, p. 516)

Similarly, Brandt remarks that considered judgments may have credence but not credibility:

> There is a problem here quite similar to that which faces the traditional coherence theory of justification of belief: that the theory claims that a more coherent system of beliefs is better justified than a less coherent one, but there is no reason to think this claim is true unless some of the beliefs are initially credible—and not merely initially believed—for some reason other than their coherence, say, because they state facts of observation. In the case of normative beliefs, no reason has been offered why we should think that initial credence levels, for a person, correspond to credibilities. The fact that a person has a firm normative conviction [considered judgment] gives that belief a status no better than fiction. Is one coherent set of fictions supposed to be better than another? (1979, p. 20)

Two points, I think, can be offered in response. The first is that Brandt (though not Singer) eventually softened his position on considered judgments. In fact, the science-based theory of ethics Brandt came to favor can heed considered judgments. This theory, he says,

> will respect "considered opinions" about what is right, at least as warning signs where there is disparity, for in the form these take in ordinary life they seem mostly to occur either as expressions of compassion (being shocked by harm to others) or repugnance at some forms of behavior which members of the society normally have come to find repugnant because of their usual connection with tendencies to harm. The "considered judgments" doubtless have been strongly influenced by the norms of society, but since there seems to be a process of pruning which modifies these norms so as to drop the useless or harmful norms and also some process of thoughtfully addressing new problems, one's society-produced judgments deserve some respect. Norms have been much influenced by what thoughtful people have wanted in the moral code of the society. (Not always: see the… morality of homosexual behavior.) So it is reasonable to accept a principle of conservatism here. (1990, p. 277)

Secondly, I hope to show in successive chapters that if reflective equilibrium is reconceptualized with respect to phenomenal, instrumental, and teleological strata, discourse within each stratum can not only express belief but acquire credibility. I argue for this conclusion within the phenomenal stratum in Chap. 2; within the instrumental stratum in Chap. 4; and within the teleological stratum in Chap. 5. These arguments are fully consistent with the conviction that justification of moral discourse within each of these strata is an intricate and arduous affair and that the mutual adjustment of these strata is more demanding still. Moral discourse is fallible through and through. But in no stratum, I maintain, can a no-credibility argument be sustained.

1.3.4 The Nature of Considered Judgments

The other difficulty with considered judgments is just what they are supposed to be. Rawls defined them as "those judgments in which our moral capacities are most likely to be displayed without distortion" (1971, p. 47). However, Goodman's seminal contrast was between general rules and particular inferences:

> Justification of general rules thus derives from judgments rejecting or accepting particular deductive inferences. . . . The point is that rules and particular inferences alike are justified by being brought into agreement with each other. . . . Thus the interplay we observed between rules of induction and particular inductive inferences is simply an instance of this characteristic dual adjustment between definition and usage, whereby the usage informs the definition which in turn guides extension of the usage. (1979, p. 64, 66)

In the view of many readers, Rawls likewise meant to distinguish considered judgments, which are particular, from principles, which are general (e.g., Daniels 1996, p. 1). But Rawls uses the injustice of religious intolerance and racial discrimination as examples of considered judgments in *A Theory of Justice* (1971, p. 19), and the post-*Theory* Rawls says explicitly that considered judgments can have any level of generality: "People have considered judgments at all levels of generality, from those about particular situations and institutions up through broad standards and first principles to formal and abstract conditions on moral conceptions" (1975, p. 8; cf. Richardson 1994, p. 178). However, this complicates the contrast between considered judgments and principles considerably. Just how are they supposed to differ? We return briefly to this problem in Sect. 1.5.

1.3.5 Intuitionism

A final difficulty with the method of reflective equilibrium is its relation to intuitionism.[4] Some commentators have accused Rawls of espousing a form of subjective intuitionism. R. M. Hare claims "Intuitionism is nearly always a form of disguised subjectivism. Rawls does not call himself an intuitionist; but he certainly is one in the usual sense" (1973, p. 146). Similarly, Brandt includes Rawls among those who espouse "the method of intuitions" (1979, pp. 19–22). In a related line of argument, Stephen Stich draws on the heuristics and biases literature to urge that people under the sway of a fallacious heuristic—an intuition—might adopt some "daffy inferential rule" and yet be in a state of reflective equilibrium (1990, p. 86).

The tendency to link reflective equilibrium with intuitionism is reinforced by a penchant for describing considered judgments as intuitions. Jared Bates, for example, understands the method of reflective equilibrium in epistemology to require testing theories against "our intuitions about cases of justified and unjustified belief" (2004, p. 45). Analogously, David Gauthier takes the method of reflective equilibrium in morality to entail appeal to moral intuitions (that is, considered moral judgments):

> If the reader is tempted to object… on the ground that his *moral intuitions* are violated, then he should ask what weight such an objection can have, if morality is to fit within the domain of rational choice. We emphasize the radical difference between our approach, in which we ask what view of social relationships would rationally be accepted *ex ante* by individuals concerned to maximize their utilities, from that of moral coherentists and defenders of "reflective equilibrium," who allow initial weight to our *considered moral judgements*. (1986, p. 269; emphasis added)

To clarify the relation between reflective equilibrium and intuitionism, let us begin with Rawls' own definition of intuitionism: "Intuitionist theories, then, have two features: first, they consist of a plurality of first principles which may conflict to give contrary directives in particular types of cases; and second, they include no explicit method, no priority rules, for weighing these principles against one another: we are simply to strike a balance by intuition, by what seems to us most nearly right" (1971, p. 34). Evaluated in the light of this definition, Rawls' principles of justice do not form an intuitionist system. Even though he admits a plurality of first principles, his first principle of justice has clear priority over the second.

In addition, the method of reflective equilibrium is not intuitionist in a straightforward sense. Intuitionism in the usual sense is a form of foundationalism, and some commentators accordingly take reflective equilibrium to be foundationalist (e.g., Bates 2004, pp. 48–50). However, as Daniels points out, reflective equilibrium is not foundationalist (1979, pp. 264–265, 1996, p. 4). In the quest for reflective equilibrium, none of our beliefs is held immune from revision. Because there

[4] There are at least four different species of intuitionism: reliabilism, experientialism, reflectionism, and contextualism (Sinnott-Armstrong 2006, p. 186). The reliability of moral intuitions has been attacked (Singer 2005) and defended (Tersman 2008) in the light of recent neuroscience.

is no priority among considered judgments, principles, and background theories, a considered judgment might be rejected because it conflicts with principles or background theories, or a principle might be abandoned because it is inconsistent with considered judgments or background theories. Consequently, since reflective equilibrium is not foundationalist, it is not intuitionism in the usual sense. Reflective equilibrium is arguably not foundationalist in another, more radical sense. Michael DePaul claims that reflective equilibrium and foundationalism "are not really positions on the same topic" because reflective equilibrium is a method, while foundationalism is an epistemic account of beliefs (1986, p. 68).

Nevertheless, some might grant that reflective equilibrium is not intuitionist in the two previous senses but counter that it is in a third. That is, even though Rawls' principles of justice are prioritized and reflective equilibrium permits neither considered judgments nor principles to be foundational, the method does rely on moral intuitions in cases of reflective disequilibrium. Given the real possibility that common morality is inconsistent (Brand-Ballard 2003, pp. 231–232), reflective disequilibrium is not a peripheral matter. Say, then, that we notice an inconsistency between considered judgments and principles but resolve the inconsistency through an intuition that the principles are more reliable. In another instance, of course, we might appeal to an intuition in favor of the considered judgments. But no matter which way we lean, the argument goes, we are relying on moral intuitions, and these intuitions serve as a cognitive foundation.

One response to this argument is to suggest that those who read Rawls as an intuitionist fail to distinguish between narrow and wide reflective equilibrium (Daniels 1979, p. 267 n. 17). For if the goal is not just coherence of our native belief system but coherence acquired through the widest possible reflection, through radical contrasts of our beliefs with alternative conceptions, the choice of considered judgments over principles (or vice versa) is not due to an intuition; it is due to disciplined philosophical reflection. This response is on the right track, I think, but there is more to be said. If wide reflective equilibrium is suitably reconceptualized, any residual tendency to conflate it with intuitionism vanishes. A proposal along these lines occupies the following section.

1.4 A Proposal for Wide Reflective Equilibrium

Consistent with the view that the quest for reflective equilibrium is not restricted to moral philosophy, a reconceptualization of reflective equilibrium can be extrapolated from the philosophy of science.[5] Larry Laudan has proposed what he calls the reticulated model of scientific rationality (1984, pp. 50–66). Perhaps the simplest way to introduce its central ideas is by contrast with the older, hierarchical model of justification associated with Popper, Hempel, and Reichenbach, among others. The

[5] The following proposal builds on Welch (1994, pp. 281–282).

proponents of this model identified three levels of scientific discourse. The first is factual, wherein scientists might agree or disagree about the structure of DNA or the existence of phlogiston. The second is methodological; agreement or disagreement here could be over rules of scientific procedure such as the Royal Society's "Nullius in verba" or how to use a bubble chamber. The third, axiological level concerns the goals of science: to avoid action at a distance, say, or to strive for falsifiable theories.

How the model explains the resolution of disagreement clearly reveals its hierarchical structure. Disagreements at the factual level are resolved by ascending to the methodological level; disagreements over methodology are settled by moving up to axiology. But what about axiological differences? That was the flaw, of course. The model provided no resources whatsoever for resolving them in a rational way. And since the model makes all justification depend ultimately on axiology, to claim that axiology is impervious to reason is to admit that, ultimately, so too is science. The hierarchical model builds science on sand.

To redress that problem, Laudan does away with the hierarchy. The reticulated model maintains the distinction between factual, methodological, and axiological discourse. But it admits "a complex process of mutual adjustment and mutual justification" among all three levels, and it subjects these levels to "a kind of leveling principle that emphasizes the patterns of mutual dependence between these various levels" (1984, pp. 62–63). Whereas the hierarchical model permitted justification to flow only downward, from axiology to methodology to fact, the reticulated model allows justification to flow upward as well. Not only, then, are facts justified by methods and goals, but methods and goals are justified by facts.

Whereas Laudan treats the mutual adjustment of factual, methodological, and axiological discourse in science, Rawls discusses the mutual adjustment of considered judgments and principles in ethics. But there is clearly a structural similarity between the two approaches. Rawls himself refers to considered judgments in reflective equilibrium as "facts" (1971, p. 51), which makes these considered judgments comparable to Laudan's factual discourse. In addition, Rawls' principles are kin to Laudan's methodological and axiological discourse. Nevertheless, Laudan's model has one important advantage over Rawls': his "principles" are differentiated into methodological and axiological levels. Because of this greater explicitness, I will work with it in the following pages.

Though the reticulated model was developed for scientific rationality, it is not limited to scientific rationality. Laudan remarks that axiological change is driven by "a theory of rationality," which is "acting to overcome a state of disequilibrium," and that the reticulated model is meant for inquiry (1984, p. 55, 63–64). In addition, he explicitly mentions the possibility of extending the reticulated model to moral theory (1984, pp. 138–139), though he naturally does not do so in his discussion of science. I would like to make the attempt. To facilitate matters, I want to modify Laudan's terminology in the interest of greater generality. These modifications make Laudan's insights more widely applicable and, of special interest to us here, more congenial to moral discourse.

The first adjustment is that I will speak of moral phenomena instead of moral facts. Though there is a literature on moral facts,[6] the term's connotations are jarring. By contrast, the term 'phenomena' comfortably subsumes Laudan's facts, which might concern observables like an earthquake or unobservables such as tectonic plates, as well as moral matters like the generosity of a gift or the fairness of a policy. 'That was cruel' and 'She is a loyal friend' are examples of morally phenomenal discourse.

Secondly, I will refer to morally instrumental rather than morally methodological discourse. The reason is straightforward: all methods are instruments, but not all instruments are methods. A scientific hypothesis, for instance, is instrumental even if no method has been used to elaborate it. Although some ethical theories include what might be taken to approximate a method—the hedonistic calculus comes to mind—much ethical discourse is bereft of method. Much is instrumental, however, and I will therefore use the more general term. 'Be good in order to be happy' is an example of morally instrumental discourse.

Finally, I will discuss moral teleology rather than moral axiology. Granted, both the ends of teleology and the values of axiology are germane if, as Aristotle claimed, all action is aimed at some good (*Nicomachean Ethics* 1094a1–3; *Politics* 1252a1–4). For all ends are values; that is, all ends of action reflect the values of the agents performing the actions. But not all values are ends. I may value a pocket calculator, for instance, though I value it as a means rather than an end. Since axiology's scope includes what we value as means and what we value as ends, it tends to blur any investigation that relies on the distinction between them. For my purposes, then, teleology rather than axiology is the right stuff. 'Do good and avoid evil' is an example of morally teleological discourse.

To help motivate the distinction between teleology and instrumentality, I note that it illuminates the Western moral tradition from end to end. Sidgwick's systematization of ethics, for example, differentiated ultimate ends from methods. The ultimate ends of moral action, he claimed, are self-perfection and happiness, where happiness might be conceived individually or collectively (1907, I.i.4, pp. 9–11). These differences, which are teleological, demarcate perfectionism of an Aristotelian sort, say, from ethical egoism and utilitarianism. He also distinguished intuitional, egoistic, and utilitarian methods, remarking "almost any method may be connected with almost any ultimate reason [or end] by means of some—often plausible—assumption" (1907, I.vi.3, p. 83). Sidgwick's methods belong to the instrumental level of morality and, for reasons stated two paragraphs previous, are better conceived as such.[7]

[6] Some examples are Dewey ([1922] 1983, pp. 166–167); Rawls (1971, p. 51); Harman (1977, pp. 131–132); Mackie (1977, pp. 16–17, 25–27, 41); Ross (1991, pp. 243–269); Timmons (1991, pp. 382–383); Passell (1995, pp. 463–480); Smith (2001, pp. 70–77); Dreier (2006, pp. 197–282); Sinnott-Armstrong (2006, p. 27, 32–59, and passim); and Prinz (2007, p. 89).

[7] In addition, Sidgwick's discussion of intuitional methods relies on distinctions that are structurally similar to Laudan's levels in some respects: "perceptional" intuitionism yields immediate intuitions about particular actions; "dogmatic" intuitionism relies on the rules of common sense morality; and "philosophical" intuitionism seeks deeper explanations of these common sense rules (1907, I.viii.2–4, pp. 98–104).

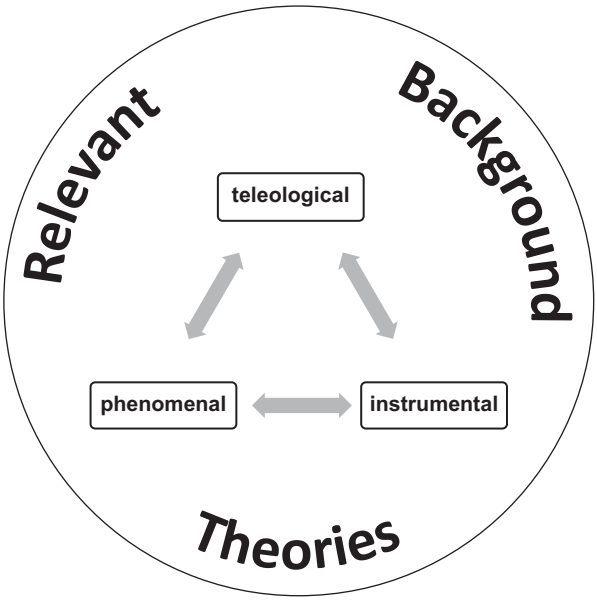


Fig. 1.1 Wide reflective equilibrium

What we have, then, is a refocusing of wide reflective equilibrium as coherence among phenomenal, instrumental, and teleological levels of moral discourse together with background theories. Their interrelations are represented in Fig. 1.1.

1.5 Conclusion

This chapter has offered a reinterpretation of wide reflective equilibrium as coherence among phenomenal, instrumental, and teleological discursive strata in addition to background theories. Some strengths of this interpretation can be suggested by reverting to two problems that afflict the original concept of reflective equilibrium: the nature of considered judgments (Sect. 1.3.4) and the relation to intuitionism (Sect. 1.3.5).

That the nature of considered judgments is no longer a problem is evident: in our proposed view of reflective equilibrium, there are no considered judgments to consider. We have morally phenomenal statements such as ‘This policy is discriminatory’; morally instrumental statements such as ‘That action maximizes utility’; and morally teleological statements such as ‘The good will is good without qualification’. Granted, there are still problems of justification for each type of discourse, but these problems are often solvable problems. How to solve them when they emerge within the morally phenomenal stratum is the topic of Chap. 2; within the morally instrumental stratum, of Chap. 4; and within the morally teleological stratum, of Chap. 5. In addition, the prospects for reducing or eliminating the reflective disequilibrium that can plague moral discourse are explored in Chap. 6.

The temptation to confuse reflective equilibrium with intuitionism is undercut as well. The reconceptualization of reflective equilibrium imposes a division of intellectual labor that facilitates direct links to resources for rational choice. These resources include inductive logic and decision theory. How quantitative inductive logics can help with certain issues in the morally phenomenal stratum is the theme of Chap. 2. The value of comparative decision theory in guiding morally instrumental and teleological decisions is shown in Chaps. 4 and 5. These linkages make it quite clear that the method being employed is not intuitionist.

References

Amit, Ron. 2006. Rawls as a critical theorist: Reflective equilibrium after the 'deliberative turn'. *Philosophy and Social Criticism* 32:173–191.

Aristotle. 1984. *Nicomachean Ethics*. In *The complete works of Aristotle,* ed. Jonathan Barnes, vol. 2, 1729–1867. Princeton: Princeton University Press.

Aristotle. 1984. *Politics*. In *The complete works of Aristotle,* ed. Jonathan Barnes, vol. 2, 1986–2129. Princeton: Princeton University Press.

Bates, Jared. 2004. Reflective equilibrium and underdetermination in epistemology. *Acta Analytica* 19:45–64.

Brand-Ballard, Jeffrey. 2003. Consistency, common morality, and reflective equilibrium. *Kennedy Institute of Ethics Journal* 13:231–258.

Brandt, Richard B. 1954. *Hopi ethics: A theoretical analysis*. Chicago: University of Chicago Press.

Brandt, Richard B. 1979. *A theory of the good and the right*. Oxford: Clarendon Press.

Brandt, Richard B. 1990. The science of man and wide reflective equilibrium. *Ethics* 100:259–278.

Daniels, Norman. 1979. Wide reflective equilibrium and theory acceptance in ethics. *The Journal of Philosophy* 76:256–282.

Daniels, Norman. 1996. *Justice and justification: Reflective equilibrium in theory and practice*. Cambridge: Cambridge University Press.

DePaul, Michael R. 1986. Reflective equilibrium and foundationalism. *American Philosophical Quarterly* 23:59–69.

Dewey, John. 1922. *Human nature and conduct: An introduction to social psychology*. In The middle works, 1899–1924, ed. Jo Ann Boydston and Patricia Baysinger, vol. 14. Carbondale: Southern Illinois University Press, 1983.

Dreier, James. 2006. *Contemporary debates in moral theory*. Oxford: Blackwell.

Elgin, Catherine Z. 1989. The relativity of fact and the objectivity of value. In *Relativism: Interpretation and confrontation,* ed. Michael Krausz, 86–98. Notre Dame: University of Notre Dame Press.

Elgin, Catherine Z. 1996. *Considered judgment*. Princeton: Princeton University Press.

Gauthier, David. 1986. *Morals by agreement*. Oxford: Clarendon Press.

Goodman, Nelson. 1979. *Fact, fiction, and forecast*. 3rd ed. Indianapolis: Hackett.

Hare, R. M. 1973. Rawls' theory of justice. *The Philosophical Quarterly* 23:144–155, 241–252.

Harman, Gilbert. 1977. *The nature of morality*. New York: Oxford University Press.

Harman, Gilbert. 2003. Three trends in moral and political philosophy. *The Journal of Value Inquiry* 37:415–425.

Knight, Carl. 2006. The method of reflective equilibrium: Wide, radical, fallible, plausible. *Philosophical Papers* 35:205–229.

Laudan, Larry. 1984. *Science and values: The aims of science and their role in scientific debate*. Berkeley: University of California Press.

Lazari-Radek, Katarzyna de and Peter Singer. 2012. The objectivity of ethics and the unity of practical reason. *Ethics* 123:9–31.
Mackie, J. L. 1977. *Ethics: Inventing right and wrong*. Harmondsworth: Penguin.
McGee, Vann. 1985. A counterexample to modus ponens. *The Journal of Philosophy* 82:462–471.
Passell, Dan. 1995. Natural fact, moral reason. *Journal of Philosophical Research* 20:463–480.
Peirce, Charles S. 1878. How to make our ideas clear. In *The writings of C. S. Peirce*, vol. 3, 257–276. Bloomington: Indiana University Press, 1986.
Prinz, Jesse. 2007. *The emotional construction of morals*. Oxford: Oxford University Press.
Putnam, Hilary. 1983. *Realism and reason*. *Philosophical papers,* vol. 3. Cambridge: Cambridge University Press.
Rachels, James. 1986. *The elements of moral philosophy*. New York: Random House.
Rawls, John. 1971. *A theory of justice*. Cambridge: The Belknap Press of Harvard University Press.
Rawls, John. 1975. The independence of moral theory. *Proceedings and Addresses of the American Philosophical Association* 48:5–22.
Richardson, Henry S. 1994. *Practical reasoning about final ends*. Cambridge: Cambridge University Press.
Ross, Stephen. 1991. The nature of moral facts. *The Philosophical Forum* 91:243–269.
Russell, Bertrand, and Alfred North Whitehead. 1927. *Principia mathematica*. 2nd ed. Cambridge: Cambridge University Press.
Sidgwick, Henry. 1907. *The methods of ethics*. 7th ed. London: Macmillan.
Singer, Peter. 1974. Sidgwick and reflective equilibrium. *The Monist* 58:490–517.
Singer, Peter. 2005. Ethics and intuitions. *The Journal of Ethics* 9:331–352.
Sinnott-Armstrong, Walter. 2006. *Moral skepticisms*. Oxford: Oxford University Press.
Smith, Barry. 2001. The Chinese rune argument. *Philosophical Explorations* 4:70–77.
Stich, Stephen. 1990. *The fragmentation of reason: Preface to a pragmatic theory of cognitive evaluation*. Cambridge: The MIT Press.
Tersman, Folke. 2008. The reliability of moral intuitions: A challenge from neuroscience. *Australasian Journal of Philosophy* 86:389–405.
Timmons, Mark. 1991. Putnam's moral objectivism. *Erkenntnis* 34:371–399.
Varner, Gary E. 2012. *Personhood, ethics, and animal cognition: Situating animals in Hare's two-level utilitarianism*. Oxford: Oxford University Press.
Welch, John R. 1994. Science and ethics: Toward a theory of ethical value. *Journal for General Philosophy of Science* 25:279–292.
Wood, Allen W. 2008. *Kantian ethics*. Cambridge: Cambridge University Press.

Chapter 2
Saving the Moral Phenomena

Abstract Chapter 2 focuses on morally phenomenal statements such as 'That was generous' and 'He is unfair'. Disagreements over such statements are rooted in the vagueness of terms like 'generous' and 'unfair', which makes the use of these terms to classify actions and people problematic. This chapter introduces core classification as the fundamental form of classification, linguistic or not. To develop the concept of core classification, the chapter proposes the analogy thesis: positive core classification is by analogy; negative core classification is by disanalogy. This is a descriptive claim, but there is an attendant normative thesis: good core classifications result from good analogies. When is an analogy good? The chapter proposes that arguments by analogy can be evaluated by appeal to a standard of inductive cogency. Like the standard of deductive soundness, inductive cogency imposes a condition on the argument's content and a condition on its form. The formal condition is that the form be inductively strong, where inductive strength can be defined by quantitative inductive logics in the tradition of Carnap, Hintikka, Kuipers, and Niiniluoto. The chapter claims that recourse to inductive cogency affords an in-principle solution to morally phenomenal disagreements. To illustrate this claim, the chapter addresses the moral dilemma faced by Cicero's grain merchant.

2.1 Inductive Molding

Plato tasked astronomers with saving the phenomena: to account for the observed motions of the planets with hypotheses of uniform circular motion. This chapter undertakes to save the phenomena as well, though the phenomena it treats are linguistic, not astronomical. More specifically, the chapter ponders our linguistic practices of classification. The discussion treats classification in general, but it does so in order to focus on moral classification with predicates proper to the phenomenal stratum of moral discourse. Unproblematic moral classifications (that the torture of a child is cruel, say) and unproblematic nonmoral classifications (such as identifying Secretariat as a horse) are the linguistic counterparts of the observed motions of the planets. They are the phenomena to be saved.

J. R. Welch, *Moral Strata,* Theory and Decision Library A:, 49,
DOI 10.1007/978-3-319-08013-0_2,

These phenomena do not exhaust our classificatory experience. The predicate that we apply with aplomb to Secretariat may stump us in the presence of a mule. Predicates that are clearly applicable in some situations are doubtfully applicable in others. For predicates, as we know, are vague. Vagueness is so ubiquitous that Peirce claimed "*No* concept, not even those of mathematics, is absolutely precise; and some of the most important for everyday use are extremely vague" ([c. 1906] 1931–1958, 6.496; emphasis in original).

Just what vagueness is has been strenuously debated.[1] Vagueness is epistemic, according to some (Sorensen 1988; Williamson 1994). Objects have clear boundaries, and vague language reveals ignorance of these boundaries. Vagueness is ontological, according to others (Tye 1990). Objects like Mount Everest are inherently fuzzy, and vague language reflects the underlying fuzz.

This chapter deploys what I take to be a compromise position. It discusses the central case of vague predicates, though adjectives, adverbs, quantifiers, definite descriptions, and proper names may also exhibit vagueness. Predicates are coined in specific contexts for specific purposes, but these limited practices do not automatically fix the extensions of predicates over the domain of all objects. The linguistic community using the predicate has rarely considered, much less decided, all questions that might arise about the predicate's extension. To this extent, then, I take the ontological view to be correct: there may be no fact of the matter of whether a man with 229 hairs on his head is bald. But this is not the end of the matter. A predicate that clearly applies in some contexts can be reasonably extended to others where it is initially vague. This process of development approximates the cognitive remedy for vagueness that the epistemic view prescribes. Provided the line of development runs from clear to problematic cases, it is comparable to a hypothesis that saves the phenomena.

This developmental process appears to be what G. H. von Wright was groping for in his reflections on molding concepts (1963, p. vii, 5, 138, 171). The urge to undertake conceptual investigation is one of the main reasons for doing philosophy, he claimed. This urge arises from bewilderment about the meaning of words. But this is not the type of bewilderment produced by unfamiliar terms. It arises in connection with familiar terms when the grounds for their appropriate use are incompletely known. The aim of this type of conceptual investigation

> is not to "uncover" the existing meaning (or aspect of meaning) of some word or expression, veiled as it were behind the bewildering complexities of common usage. The idea of the philosopher as a searcher of meanings should not be coupled with an idea or postulate that the searched entities actually *are there*—awaiting the vision of the philosopher. If this picture of the philosopher's pursuit were accurate, then a conceptual investigation would, for all I can see, be an *empirical* inquiry into the actual use of language or the meaning of expressions.
>
> Philosophical reflexion on the grounds for calling a thing '*x*' is challenged in situations, when the grounds have not been fixed, when there is no settled opinion as to what the grounds are. The concept still remains to be *moulded* and therewith its logical connexions with other concepts to be *established*. The words and expressions, the use of which bewilder the philosopher, are so to speak *in search of a meaning*. (1963, p. 5)

[1] This chapter is a revised version of Welch (2007).

In the spirit of these remarks, what I propose in this chapter is a strategy for conceptual investigation. The basic idea is to mold concepts and thereby reduce vagueness through a process of inductive inference. As stated, the emphasis will be on molding moral concepts proper to the phenomenal stratum of moral discourse. That is, "thick" terms such as 'honest' and 'courageous' are notoriously vague, and the chapter proposes a strategy for reducing their vagueness. The argument is modest in that I will not claim that inductive molding can eliminate vagueness. Unfortunately, the law of excluded middle does not always hold. But I will argue that the truth-value gaps associated with these failures need not be permanent, that they can be reduced on a piecemeal basis. The engine of reduction, I claim, is inductive logic.

Vagueness cannot be understood apart from the backdrop of classification, for vagueness is classification gone awry. Hence these pages explore the classification of particulars, both its clear successes and vague failures. How we classify particulars is the theme of the next two sections, which are primarily descriptive. Section 2.2 identifies a way of classifying particulars that pervades discourse of all sorts, and Sect. 2.3 illustrates its use in moral discourse. Why a certain particular should (or should not) be classified in a certain way is a normative question, however, and it occupies the two following sections. Section 2.4 proposes a standard for cogent arguments by analogy, and Sect. 2.5 illustrates how the standard might resolve vagueness in one kind of moral dispute. This standard, which has a strong probabilistic component, is one way of affirming that probability is a guide to life.

2.2 Core Classification

About 150 Yanomama Indians eke out a Stone Age existence deep in the Amazon rain forest, in an isolated village on the border between present-day Venezuela and Brazil. A team of anthropologists and journalists visited the area some years ago, and one of the journalists reported the following encounter.

> The little bull of a man with brushcut hair and only a bark string around his waist was studying our Venezuelan Air Force Super Puma helicopter like a scientist. Once before he had seen something similar, and he drew his arm high across the sky in an arc. "I kept calling him to come down, but no luck." What about the chopper, we asked, what was it? He paused for a moment, then tentatively offered: "It's an animal, a *hashimo*"—a smooth-feathered green grouse—and by way of explanation waved his arms and made the thrashing sound of a big bird exploding from the underbrush. What kept it here? "It's a pet." Did he want to go for a ride? "Maybe later. A long time from now. A really long time." (Reiss 1990, p. 46)

Of the several remarkable features of this encounter, I would like to focus on just one: Yokokoma's classification of the helicopter as a *hashimo*. How did he manage to do that? And how do we manage to classify it otherwise?

Suppose we refer to statements like 'That is a *hashimo*' and 'That is a helicopter' as *core classifications*. These rudimentary orderings of the world have the form 'δ is *T*', where 'δ' stands for a demonstrative pronoun or a proper name and '*T*' for a class

term. The notion of core classification is ample enough to include a perception that something is wet, for instance, since the perception can be expressed in the form 'δ is *T*'.[2] Proficiency in core classification is essential for getting around in the world, for survival requires reliable identifications of food, danger, and potential mates. How we core classify is therefore a question of cardinal importance.

Though the question can be addressed from many points of view, the biggest part of the answer can be put in the fewest words. Positive core classification is by analogy. The astronomer identifies a quasar, the camper a lichen, the musician a half-tone by perceiving the similarities between a unique object and previously cognized quasars, lichens, and half-tones. In all these cases, there are the known old, the unknown new, and the assimilation of the latter to the former through the relation of similarity. This assimilation is all but transparent in ordinary English expressions such as 'This looks like a mantis' and 'That smells like fire'. Negative core classification, on the other hand, is by disanalogy. The numismatist's judgment that this is not gold arises from the known old, the unknown new, and the perceived dissimilarity between the two. Hence I propose the following thesis: positive core classification is by analogy; negative core classification is by disanalogy. Call this the *analogy thesis* for short.

The analogy thesis as just formulated needs at least two qualifications, however. The first is to recognize that core classification may occur with abstract as well as concrete terms. Take the core classification 'This is an animal', for instance. Its epistemic base might well be a concrete core classification that this is a paramecium and a linguistic truth that all paramecia are animals. If this is so, however, the concrete classification would be carried out by analogy. Hence 'This is an animal' would be grounded mediately by analogy, not immediately as in concrete classifications. To cover the abstract case, I will claim that positive core classification is ultimately by analogy; negative core classification is ultimately by disanalogy.

A second qualification is needed for a relatively infrequent but key occurrence of core classification that is not analogical. When someone inaugurates a class term, there is no identification of the new by analogy with the old for the simple reason that there is no old. In the case of a newly-identified species, for example, there may be analogies between the first known individual of that species and members of neighboring species, but there can be no analogy within the species while only one exemplar is known. Call these seminal classifications *coinages*. The analogy thesis can then be restated: Except for coinages, positive core classification is ultimately by analogy; negative core classification is ultimately by disanalogy.

For all its brevity, the analogy thesis covers a lot of territory. A full appreciation of this point requires three considerations. The first is a reprise of the prior observation that core classification can be linguistic, as in 'That is a hake', or nonlinguistic,

[2] Perceptions can be interpreted as nonlinguistic beliefs. This opens up lines of inquiry into the controversial arena of animal belief. Plato and Aristotle split on the matter, as Richard Sorabji (1993) interprets them, with Plato attributing beliefs to animals and Aristotle not. In contemporary philosophy, Jeffrey (1985) is Platonic on this point while Davidson (1982) is Aristotelian. Here I side with Jeffrey.

as in a perception that Dana is tall. The analogy thesis embraces both types of core classification. The second consideration is that, despite the nature of the foregoing examples, the thesis is not limited to physical objects. Its scope is actually the class of events. This has immediate implications for physical objects, however, for each can be understood as an event that unfolds as long as the object exists. Finally, the analogy thesis applies to actions as well, for actions are purposed events.

2.3 Core Classification in Ethics

The preceding is all quite general; it makes no reference to specific domains of discourse. From here we could branch off into any domain at all. However, what I propose to explore is moral discourse of a specific kind. Morally phenomenal discourse is typified by concepts such as 'just', 'cruel', 'temperate', 'cowardly', 'honest', 'untrustworthy', 'loyal', 'unfair', 'compassionate', and their complements. Loci for the sort of thing I have in mind are the early Platonic dialogues, which can be mined for insights on core classification in ethics: courage in the *Laches*, justice in *Republic*, Book I, temperance in the *Charmides*, and so forth.

Had the search for definitions in the early dialogues been successful, or had the definitions in *Republic*, Book IV, been more than rough-cut, stopgap measures, or had there been breakthroughs in the definition of ethical terms between Plato's time and our own, we could understand ethical core classification as follows. Imagine that we have an accurate definition of justice. It tells us that an action is just if, and only if, it is *F* and *G* and *H*. We could then classify individual actions by using our definition as a criterion: since this action is *F* and *G* and *H*, it is just; and that action, because it is *F* and *G* but not *H*, is not just.

2.3.1 Prototype Theory

There is ample reason by now to think that this definitional approach is barking up the wrong tree. An extensive, multidisciplinary literature points toward a very different understanding of human classification (Lakoff 1987, Chap. 2). Though I will not survey this literature here, I will note that Wittgenstein's remarks on family resemblance (1953, § 65–78) are a point of departure for much later work in the field. And I will acknowledge the special importance of empirical work by Eleanor Rosch and her associates in cognitive psychology.[3]

Rosch is responsible for drawing together a number of separate empirical studies under the rubric of prototype theory. Part of the interest of prototype theory lies in its direct opposition to the classical conception of classification presupposed by

[3] Rosch (1973, 1978, 1983), Rosch and Mervis (1975), Rosch et al. (1976). Rosch (1978) is conveniently reprinted in Margolis and Laurence (1999, pp. 189–206) along with several papers that aim to supplement or replace prototype theory.

Plato and passed on to scores of generations of Western scholars. According to this classical view, class membership is determined by necessary and sufficient conditions, and accurate definitions state these conditions. If this were the case, however, there would be no best examples of a kind; any member of the class would serve equally well, for all would satisfy the same set of conditions. That there are best examples of a kind—prototypes—is a crucial result of Rosch's work. Speakers of American English, for example, consistently rate robins as better examples of bird than ostriches or penguins, and desk chairs as better examples of chair than beanbag chairs or electric chairs (Rosch et al. 1976). Class membership, so understood, is not an either-or affair; it is a matter of degree.

Even the classes recognized by the physical sciences seem to have prototypes. In discussing the theoretical identities 'Water is H_2O' and 'Temperature is mean kinetic energy', Hilary Putnam remarks that "the 'essence' that physics discovers is better thought of as a sort of *paradigm* that other applications of the concept ('water', or 'temperature') must *resemble* than as a necessary and sufficient condition good in all possible worlds" (1983, p. 64).

Moral classes also show prototype effects. We have no problem identifying Socrates' saving of Alcibiades' life as courageous, but how do we classify the suicide of Seneca's barbarian, who asphyxiated himself with the sponge he was given for wiping himself before his scheduled appearance in the circus to fight wild beasts? Even if we concur with Seneca's classification of it as courageous ([63–65] 1920, LXX.20–21, pp. 66–69), it is not obviously so. This graded sort of membership in moral classes is due in part to moral education, which in its early stages proceeds through introducing prototypically moral and immoral actions in fairy tales and other narratives. But it is also an effect of what happens next. Once we have learned to manage a handful of moral predicates in prototypical situations, we begin to extend these terms to new situations by analogy.

2.3.2 Washington's Cherry Tree

I want to consider one example of this process in some detail. The example is trivial, in a sense, but that is precisely its point. It is a prototype, the kind of action that serves as a moral reference point for a community—primarily, in this case, that of the United States. Despite the limitations of the example, comparable prototypes for other cultures would not be hard to find. Prototypes are culturally embedded, and the profusion of human life forms practically guarantees variety of prototypes.

The source for this sample prototype is Mason Weems' classic biography of the American President George Washington ([1809] 1962).[4] Weems was an Episcopal clergyman and bookseller who, a month after Washington's death in 1799, wrote to a business associate to propose a biography of Washington. His plan was to

[4] The incident appears for the first time in the fifth edition of 1806. The citations from Weems ([1809] 1962) are from the ninth edition of 1809.

demonstrate that Washington's "unparrelled [*sic*] rise & elevation were due to his Great Virtues" ([1809] 1962, p. xv). One of the virtues Weems attributed to Washington was honesty. To illustrate the point, he recounted the following incident:

> When George… was about six years old, he was made the wealthy master of a *hatchet*! of which, like most little boys, he was immoderately fond, and was constantly going about chopping every thing that came in his way. One day, in the garden, where he often amused himself hacking his mother's pea-sticks, he unluckily tried the edge of his hatchet on the body of a beautiful young English cherry-tree, which he barked so terribly, that I don't believe the tree ever got the better of it. The next morning the old gentleman finding out what had befallen his tree, which, by the by, was a great favourite, came into the house, and with much warmth asked for the mischievous author, declaring at the same time, that he would not have taken five guineas for his tree. Nobody could tell him any thing about it. Presently George and his hatchet made their appearance. *George*, said his father, *do you know who killed that beautiful little cherry-tree yonder in the garden?* This was a *tough question*; and George staggered under it for a moment; but quickly recovered himself: and looking at his father, with the sweet face of youth brightened with the inexpressible charm of all-conquering truth, he bravely cried out, "*I can't tell a lie, Pa; you know I can't tell a lie. I did cut it with my hatchet*." ([1809] 1962, p. 12; Weems' emphasis)

The first point to be made about this story is that even though Weems attributes it to "an aged lady, who was a distant relative [of Washington], and when a girl spent much of her time in the family," historians almost universally reject it as apocryphal ([1809] 1962, p. 9, xxiv–xxxiv). The second point is the mordant one that Weems succeeded, apparently through falsehood, in placing what he took to be Washington's honesty before the "admiring eyes" of many children. By the time of Weems' death in 1825, twenty-nine editions of his *Life* had appeared; by 1925, the number had grown to eighty ([1809] 1962, p. xx). Abraham Lincoln read it in "the earliest days of my being able to read," but so did many others ([1809] 1962, p. xxii). The story of the cherry tree in particular reached millions through being excerpted in a vast body of Sunday-school books and textbooks, notably McGuffey's *Readers*, 120 million of which were published in the United States between 1836 and 1920 ([1809] 1962, pp. xx–xxiii, xlvi–xlviii; Ong 1982, pp. 115–116). Given such a central place in moral education, there is little doubt that the incident has served as a base for Americans' understanding of honesty.[5] Almost in spite of itself, then, the incident became a moral prototype.

To see how such a prototype might be used in moral reasoning, let us return to Weems' account. From it we can extract an abstract description of the situation that includes the following features:

1. *g* is the child of *f*;
2. *g* believes that *p*;
3. *g* has a selfish desire that *f* not come to believe that *p*;
4. *f* asks *g* whether *p* is true;
5. *g* conveys to *f* that *p* is true.

[5] To avoid confusion over the kind of honesty that has to do with property, we might prefer the term 'truthfulness' here. I use 'honesty' to connect up with the case of the grain merchant below.

The result is a kind of template that is plainly applicable to other cases.

Let us consider a few. Incorrigible George next cuts down his father's pear tree, responding as before to his father's question. Because this recidivist action is identical to the prototype in the ways picked out by the template, the present action is honest as well. Now take an action that is like the preceding except that feature 1 of the template is absent: *f* and *g* are not related as parent to child. Despite the difference, the strong similarity between the prototype and this case would naturally lead us to classify the action as honest. A third case is like the second except that feature 3 of the template is missing as well (*g* is indifferent whether *f* comes to believe *p* or not, say). While the similarities between this case and the prototype make it easy enough to call the former honest, the exemplary honesty exploited by Weems has been lost.

Our reactions to these increasingly divergent cases suggest that, as a matter of fact, moral predicates are applied to novel actions or not on the basis of perceived similarities and dissimilarities between the actions and prototypes. This is most easily confirmed for predicates linked with "thick" concepts like honesty, brutality, and courage (Williams 1985, p. 129 f.). However, I submit that "thin" concepts such as the right and the good are tied to concepts like honesty via meaning postulates such as 'Honesty is *pro tanto* right'. That is, thin core classifications like 'This action is right' are epistemically grounded in thick core classifications such as 'This action is honest', and these thick classifications are carried out via analogy. If this is so, then ethical core classification proceeds ultimately by analogy and disanalogy, and the analogy thesis holds for moral discourse in particular.[6]

Observe that the analogies contemplated here are entirely factual. Moral properties picked out by thick core classifications supervene on factual properties such as instantiations of features 1–5. This standpoint should be compatible with an ample range of meta-ethical positions. It intersects with moral cognitivism of the sort defended by Alan H. Goldman, who thinks "Moral reasoning must begin with nonnormative descriptions of actions or situations and terminate in moral prescriptions" (2002, p. 13). It is also consistent with moral hybridism: the view that moral theories express both desires and beliefs (e.g., Schroeder 2009), for the beliefs can reflect factual properties that instantiate features 1–5. It should even be acceptable to some forms of moral noncognitivism. In developing a well-known form of moral subjectivism, J. L. Mackie concedes that whether an action is cruel or kind or just is an objective matter (1977, pp. 16–17, 25–27, 41). What is not objective, he thinks, is that we should be kind and just but not cruel.

A possible response to this sort of moral subjectivism is to shift our ground. We might try moving from factual analogy to mixed analogy with both descriptive and evaluative properties. Say, for example, that the Weemsian prototype has descriptive properties instantiating features 1–5 plus the evaluative property of being morally obligatory. Say that another action has the same descriptive properties. We

[6] Albert R. Jonsen and Stephen Toulmin's defense of particularism deploys a similar thesis on moral paradigms and analogy (1988, pp. 85–86, 251–252, 330). Although Alan H. Goldman rejects particularism, he makes a parallel case for moral and legal reasoning (2002, p. 2, 15, 161, 166, 168–169).

might then draw the analogical inference that the second action is morally obligatory. I submit that this inference is a good model for the psychology of many ethical analogies; we do often reason like that. But I do not believe it is the best way to justify these inferences. The reason is that mixed analogies simply assume the moral obligatoriness of the prototypical action. That the prototypical action is morally obligatory can often be shown, I believe, but not within the phenomenal stratum. Doing so requires ascent to instrumental and teleological strata. Sections 4.4–4.5 and 5.4–5.5 treat the matter directly.

2.4 A Standard for Analogy

The analogy thesis brings epistemological problems in its train. Suppose that perceptual analogies clash; one person perceives a color as mauve, say, while another perceives that it is not. Stating the conflicting analogies offers a way out by opening the analogies up to intersubjective criticism. But that is to lean on a slender and suspect base: the much-maligned argument by analogy.

Imagine some ancient sailor with a prototype of a fish in mind.[7] The sailor knows that that aquatic animal—a shark, perhaps—is a fish. Then it would be perfectly natural to reason by analogy that since this whale is an aquatic animal, this whale is a fish. Aristotle, on the other hand, thinking of some human being as a prototype of a mammal, would make the analogical inference that since that animal that nourishes its fetus with a placenta is a mammal, and since this whale is an animal that nourishes its fetus with a placenta, this whale is a mammal. Put the sailor and Aristotle together, and the result is what I will call convergent analogy: two chains of analogical inference converging at the same point—in this case, the whale. One chain reasonably identifies it as a fish; the other, just as reasonably, as a mammal.

We also find convergent analogies in ethics. Consider the African slave trade, for instance. One point of view was to justify the slave trade in the language of Aristotle's *Politics*, where the soul is said to govern the body, a human being an animal, a parent a child, all according to the natural dominance of inferior by superior (*Politics* 1254a18–1255a2, 1259b18–1260b26). Referring to one of these prototypes, a defender of the slave trade could reason as Aristotle himself would have: since that case of superior governing inferior is just, and since the treatment of Africans in the New World is a case of superior governing inferior, the treatment of Africans is just. On the other hand, the ancient ideology of slavery assimilated liberty to property: just as selling one's property need not violate justice, neither does selling one's liberty (Blackburn 1997, pp. 177–180). Accordingly, Bartolomé de Las Casas proposed in 1516 that justly enslaved Africans replace unjustly enslaved native Americans in the New World. But he was soon forced to change his mind. He saw that since native American enslavement was unjust and the conditions of African enslavement were comparable to those of native American enslavement, African

[7] The remainder of this section draws on Welch (1994, pp. 284–285).

enslavement was unjust as well (Las Casas [1561] 1994, iii.102, p. 2191; iii.129, p. 2324). Once again, then, we have analogical chains of inference converging at the same point. One says that the African slave trade is just, the other that it is unjust.

The clash of reasoned opinion that comes from the convergence of analogical chains can be transmitted almost spontaneously up the epistemological ladder. Having concluded through repeated analogies that a number of *F* are *G*, we could leap to the inductive conclusion that all *F* are *G*, which could serve in turn as a premise in a deductive inference that some as yet unexamined *F* is *G*. At the same time, rival analogical conclusions that various *F* are not *G* could ground the inference that all *F* are not *G* and, as a result, that some unexamined *F* is not *G*. The respective chains of inference just get further and further apart, it seems, the differences more and more unbridgeable. Is there any rational way to bring them together?

There are many ways, actually. Some of them work piecemeal, as the histories of the whale's taxonomy and the antislavery movement show. How long it took the rest of the world to catch up with Las Casas on slavery is as painful to contemplate as the lag in catching up with Aristotle on the whale. Yet catch up it did. But there is another approach that is at once more sweeping, more promising, and more problematic than any of these specific approaches. It is to seek general principles that would permit us to differentiate good and bad arguments by analogy. The remainder of this section attempts to specify what these principles might be.

2.4.1 Inductive Cogency

J. S. Mill once remarked that "There is no word… which is used more loosely, or in a greater variety of senses, than Analogy" ([1872] 1973–1974, III.20.1, p. 554). So let us rehearse some distinctions among kinds of analogy. Although the following typology refers exclusively to linguistic analogy, it should be kept in mind that nonlinguistic, perceptual analogies can be described *as if* they were linguistic. That is, a perceived analogy between two things can be described as if an analogical claim about them was being made.

One place to begin a typology of analogy is with the conspicuous divide between *general analogies*, analogies with at least one quantified sentence, and *singular analogies*, composed entirely of unquantified sentences. The latter, which are indispensable to core classification, can be subdivided into perfect and imperfect forms. In *perfect analogies* the relata are thought to share all relevant properties. A standard example is the traditional argument by analogy:

A_1:

$Fa \wedge Ga$.
Fb.
Hence Gb.

In *imperfect analogies*, on the other hand, the relata are thought to share some but not all relevant properties. Here is a simple example (Pietarinen 1972, pp. 68–69):

A_2:

$Fa \wedge Ga$.
$-Fb$.
Hence Gb.

A_2 makes the imperfectly analogical claim that *a* and *b* share the property *G* but not the property *F*. Though the perfect-imperfect distinction is just the beginning of a typology of singular analogy, we need pursue the matter no further here.[8]

Instead, let us turn to the crucial normative question: What is the difference between good and bad analogies? To sketch an answer, suppose we pull out the old critical saw about sound argumentation. A deductively sound argument must meet at least two necessary conditions. A condition on the argument's content requires all of its premises to be true. And a condition on the argument's form requires it to be valid in the sense that it is impossible that its conclusion is false when its premises are true. Arguments that are deductively sound are logically demonstrative.

Arguments that are inductively cogent are logically nondemonstrative. We can specify necessary conditions for inductive cogency by adapting the pattern of deductive soundness. The content condition remains the same: all the argument's premises must be true. But the formal condition is different; it must be weaker than deductive validity yet still demanding. Here the usual requirement is inductive strength, which stipulates that it be improbable that the argument's conclusion is false when its premises are true (Skyrms 1986, p. 7). Now it is improbable that the conclusion is false when the premises are true if, and only if, it is probable that the conclusion is true when the premises are true. For an argument to count as inductively strong, then, the conditional probability of its conclusion given the premises must be greater than or equal to that of any rival conclusion based on the same premises.

Hence the proposed standard of inductive cogency amounts to this: An argument is inductively cogent only if

a. all the argument's premises are true; and
b. the conditional probability of the argument's conclusion given its premises is greater than or equal to that of any rival conclusion based on the same premises.

Before relating these inductive conditions to arguments by analogy, let us note how neatly they dovetail with our deductive practice. The condition on deductive content is exactly the same, as we have noted. The condition on deductive form, the requirement that the argument be structured such that if the premises are true, then the conclusion must be true, actually implies the inductive condition on form. That is, if an argument is deductively valid, then it is also inductively strong, for its conclusion has a greater probability on the premises (probability 1) than any rival conclusion based on the same premises (probability 0). The condition on deductive form is thus a special case of the inductive condition on form.[9] This was Wittgenstein's point in

[8] A detailed typology of singular analogies can be found in Welch (1999, pp. 209–213).

[9] The inductive condition on form is in turn a special case of the plausibilistic condition on form: the plausibility of the argument's conclusion given its premises must be greater than or equal to

the *Tractatus*: "The certainty of logical inference is a limiting case of probability" (1922, 5.152; cf. Haack 1978, p. 17).

2.4.2 Analogy as Induction

Now let us link the foregoing to analogy. No argument by analogy is deductively sound, but they are not all equally unsound. To distinguish the better from the worse, I propose that we treat argument by analogy as one form of inductive argument. This is a time-honored view. Mill, for instance, remarked that arguments by analogy are "supposed to be of an inductive nature" ([1872] 1973–1974, III.20.1, p. 554), and Carnap handled analogy as induction from at least 1945 on (1945, pp. 87–88). There are a few dissenting voices, however, and they rely on two objections.

One objection has been urged by Stephen Barker, who maintains that while some analogies are inductive, others are not (2003, pp. 225–228).[10] Barker adduces an example of a non-inductive analogical argument involving a student who has passed a bad check and violated his university's honor code:

> Let us consider an example of an argument by analogy which is not inductive. At a certain college the student body has established a rigorous honor code to govern student behavior. The code specifically lists lying and cheating as punishable offenses. The students administer this code and take it seriously. Now suppose it is discovered that a student has written a bad check and used it to purchase merchandise in the town. The question arises whether this student has violated the honor code. Is writing a bad check a violation of the rule against lying and cheating? Let us suppose that those who wrote the code never pronounced on this question and that there are no known precedents about it….
>
> Deductive arguments are not likely to be of much use in this situation. Suppose someone tries to settle the problem deductively by arguing: "All cases of cheating violate the honor code; all cases of writing bad checks are cases of cheating; therefore, all cases of writing bad checks violate the honor code." Although this argument is valid, it does not succeed in proving its conclusion. If we were dubious about whether the conclusion is true, then we are pretty sure to be at least equally dubious about the minor premise. Here the deduction commits the fallacy of begging the question. No purely deductive line of reasoning is likely to settle this problem.
>
> Nor are inductive arguments likely to help much. Whichever conclusion we want to establish—that writing a bad check is, or is not, a violation of the honor code—in either case the conclusion does not embody predictive conjectures about future experience that go beyond what is already known. Reaching a conclusion about whether the student is guilty certainly is not the same as predicting how he is going to be treated. Nor is it the same as predicting what observable consequences his behavior is going to have. Such predictions would be inductive, but they are not what we are seeking. No purely inductive line of reasoning is sufficient here, for the conclusion is not of the inductive sort.
>
> What sort of reasoning would be appropriate to this problem? Someone would be making a helpful and relevant contribution to the discussion who reasoned as follows: "Lying and cheating are indisputably offenses against the honor code. Now, writing a bad check is like falsely stating that you have money in the bank. Also, writing a bad check is very like cheating, for you persuade the merchant to accept the check in exchange for merchandise

---

that of any rival conclusion based on the same premises. Plausibility is introduced in Sect. 3.3.2.4.

[10] John Wisdom, Barker's former teacher, makes a similar point about case-by-case (analogical) reasoning in the law (1969, p. 149, 158).

> by deceptively suggesting that the check is good. Since writing a bad check is so like lying and cheating in these respects, it therefore resembles them also in being a violation of the honor code." At the heart of this reasoning are the analogies between writing a bad check on the one hand and lying and cheating on the other hand. The whole argument essentially depends upon these analogies—the argument is a good argument if and only if these are good analogies.
>
> Unlike deductive reasoning, this sort of reasoning does not claim to be demonstrative. At best, the truth of the premises gives us only some good reason for accepting the conclusion. Also, as we saw, this sort of reasoning differs from induction—the conclusion being argued for does not embody predictive conjectures going beyond what the premises say. (2003, pp. 225–227)

Is Barker's example convincing? I think it is not. The reasons, very briefly, are as follows. The quotation's reference to inductive conclusions embodying "predictive conjectures about future experience that go beyond what is already known" is unduly restrictive, first of all. Why restrict induction to future experience? Barker himself characterizes induction without this restriction in the same work: inductive arguments have conclusions "embodying empirical conjectures about the world that do not follow deductively [from] what its premises say" (2003, p. 181). According to this more general description, empirical conjectures about past experience such as why Caesar crossed the Rubicon would also qualify as inductive. But then why restrict induction to empirical conjectures? Even mathematics employs nondeductive reasoning (Franklin 1987), and such reasoning is inductive in the sense of being nondemonstrative (cf. Carnap 1945, p. 72; Black 1967, p. 169; Skyrms 1986, pp. 6–15; Adams 1998, p. 70; Audi 2004, p. 129).

Second, Barker's analogy between writing bad checks, on the one hand, and lying and cheating, on the other, involves *classes* of actions. But there is no a priori reason to expect all or no bad check passings to be lies or cheats, and every a priori reason to think that some are and some are not. Contrast an intentional and manipulative passing of a bad check with a case where a student writes a bad check in good conscience, relying on her bank's mistaken statement of sufficient funds. The proper focus, then, is whether a *specific action*, this student's passing of this bad check, is a member of the class of lies or cheats.

Third, a respectable argument for Barker's case would therefore be something like the following:

An intentionally deceptive stated message is a lie (Bok 1979, p. 14).

This student's signing this bad check is an intentionally deceptive stated message.

Thus this student's signing this bad check is a lie.

Other arguments might be brought to bear as well, of course, but let us examine this one. The argument is deductively valid, and its empirical fulcrum is the truth of the second premise. Did the student intend to send a deceptive stated message or not? Empirical indicators of intent might well be present: the student's confession, the testimony of anyone party to the plan, a past history of writing bad checks, evidence of the bank's failure to credit a prior deposit to the student's account due to negligence, computer error, a strike, etc. If these indicators are present, they could serve as premises supporting a conclusion about the student's intent. If they are not present, we might still draw a conclusion about the student's intent based on general knowledge of correlations between bad check passing and deceptive intent.

In either case, the reasoning is inductive in nature. I conclude, then, that Barker's purported *sui generis* example can be handled through the use of standard inductive and deductive arguments.

Still, Barker's conclusion might be supported by a second objection. It might be admitted that many analogies are inductive while still maintaining that many are not, for many are abductive rather than inductive. In response, I suggest that we distinguish logical and functional views of argumentation. The logical approach is to classify an argument according to the degree of support the conclusion receives from the premises. But the functional approach keys on how the argument is used. Peirce's trichotomy of deductive, inductive, and abductive arguments is (usually) functional, for instance. Abduction is the first step of scientific reasoning; it advances a hypothesis. Induction is the last step; it uses experiment to verify a deductive consequence of the hypothesis ([1901] 1931–1958, 7.218; [1902a] 1931–1958, 2.96). One Peircean example treats 'This is an ex-priest' first as the conclusion of an abductive argument to explain a surprising conjunction of features and then, after the "experiment" of getting the man to remove his hat to confirm that he is tonsured, as the conclusion of an inductive argument (1902b).

The functional difference between abductive and inductive arguments is large, but the logical difference is small. That this man is an ex-priest follows with some probability from premises describing the initial set of features (Peirce does not name them), and it follows with some greater probability from the initial set plus the premise on tonsure. The difference is one of degree, as Peirce seems to recognize ([1878] 1986, pp. 326–327). Since both conclusions have a conditional probability less than 1, both arguments contrast with demonstrative arguments, whose conclusions have a conditional probability of 1. Both arguments are therefore inductive in the standard logical sense of being nondemonstrative. We can recognize that an argument is functionally abductive, then, and at the same time logically inductive. The foregoing claim that analogy is inductive should be understood in this logical sense.

I suggest that the burden of further proof lies with anyone who wishes to claim that argument by analogy is not inductive. In the meantime, I will appeal to the aforementioned standard for cogent induction. Since argument by analogy is inductive in Barker's sense of nondemonstrative reasoning, a cogent argument by analogy must have all true premises and a conclusion at least as probable on the evidence as any rival conclusion based on the same evidence. Accepting this standard, then, would gain us a principled distinction between good and bad analogies. Applying this standard would weed out poor analogies from the start, preventing them from crowding good seed.

2.4.3 Analogy and Inductive Strength

Let us now attend to the details of this proposal. Arguments by analogy offer difficulties, but no peculiarly analogical difficulties, in determining the truth of their premises. A moral property like honesty supervenes on factual properties, as illustrated in

Sect. 2.3.2. What makes moral properties seem puzzling, I suggest, is that they are disguised relations. When we say that one building is taller than another, we expect the buildings to be visible but not something else called 'taller than'. Similarly, when we say that Washington's response to his father is honest, the actions of father and son would be visible but not something else called 'honest'. The difference is that the explicitly relational surface grammar of 'taller than' naturally draws our attention to the relata, whereas the monadic surface grammar of 'honest' tempts us to search for the corresponding monadic property. There is none. The honesty of an action is a relation among the action's factual properties (cf. Railton 2003).

By contrast with the relative straightforwardness of the truth condition, arguments by analogy do present special difficulties over form. How might we go about applying the condition of inductive strength? Arguments by analogy are built around the relation of similarity, so intuitively it would seem that the relata of inductively strong analogies are somehow more similar than dissimilar, while those of inductively weak analogies are somehow more dissimilar than similar. But putting this intuition to work would require some sort of similarity metric.

To find one, I propose that we consider those logics developed along lines sketched out by Wittgenstein (1922, § 5.15–5.156) and Waismann (1930–1931). Carnap (1952) made the decisive step forward, and his work has served as the basis for later advances by Hintikka (1966), Carnap (1971, 1980), Pietarinen (1972), Hintikka and Niiniluoto (1976), Kuipers (1978, 1984), Niiniluoto (1981), Skyrms (1991, 1993), and Festa (1997), among others. In Carnap's mature work (e.g., 1942, pp. 96–97, 1945, pp. 73–75), the concept of range is a semantic concept explicable as the set of models in which a given sentence (or conjunction of sentences) is true. Suppose we call such models the sentence's *alethic models*.

The relationship between the alethic models of an argument's premises and those of its conclusion shows the probability of the conclusion on the evidence of the premises. There are two types of cases. If the alethic models of the conclusion include all the alethic models of the premises, the conclusion follows from the premises with probability 1, and the argument is deductively valid. On the other hand, if the alethic models of the conclusion do not include all of the alethic models of the premises, the conclusion follows with some probability less than 1, and the argument is not deductively valid. For example, if 3/4 of the premises' alethic models are included in those of the conclusion, the probability of the conclusion given the premises is 3/4.

One result of Carnap's critique of classical probability was his λ-continuum of inductive methods (1952). Given any method of this continuum, the degree of confirmation of a singular hypothesis on the evidence lies within an interval bounded by an empirical factor and a logical factor. The empirical factor is the evidence e_Q, the ratio of the n_Q favorable instances of some strongest property Q to the total number n of instances examined. The logical factor is equal to relative width, which is very roughly the coverage of a property relative to the totality of properties the language admits. Less roughly, for a first-order language with identity recognizing a finite number m of logically independent primitive properties, there are $2^m = K$ strongest properties in the language. Any property that can be picked out in the

language is either a strongest property or equivalent to a disjunction of strongest properties. If the property is a strongest property, its relative width is $1/K$. If the property is not a strongest property, it is equivalent to a disjunction of w strongest properties and its relative width is w/K.

Exactly what point of this interval represents degree of confirmation—or probability, as I shall say—is determined by taking a weighted mean of the empirical and logical factors. Different λ-methods use different logical weights, that is, different specifications of the parameter λ, which can take values from 0 to ∞ inclusive. Now suppose that we have evidence e_Q and that λ can vary with K but not with n_Q and n. Then, for any method of the continuum, the conditional probability p of the hypothesis h_Q that the next individual will have a strongest property Q is given by Eq. 2.1.

$$p(h_Q \mid e_Q) = \frac{n_Q + \frac{\lambda(K)}{K}}{n + \lambda(K)}. \tag{2.1}$$

Equation 2.1 allows for uncountably many λ-methods, but Carnap's favorite was c*, where $\lambda(K)=K$. For consistency with our probabilistic terminology, I will refer to this method as 'p*'. p*'s representative function[11] expresses the probability of the hypothesis h_Q on the evidence e_Q as in Eq. 2.2.

$$p^*(h_Q \mid e_Q) = \frac{n_Q + 1}{n + K}. \tag{2.2}$$

The methods of the λ-continuum are problematic in several ways, but the crucial shortcoming for our purposes is their handling of analogy. For example, where $K=4$, p* assigns the perfect analogy A_1 a probability of 2/3, which seems reasonable enough, while the imperfect analogy A_2 receives a probability of 1/2, which also seems reasonable enough—until we notice that its property analogy has been completely overlooked. That is, since A_2's conclusion 'Gb' has a probability of 1/2, the other possible conclusion, '$-Gb$', receives the same probability. But that is to consider the conjunction of the disanalogous properties FG and $\overline{FG}$ just as likely as that of A_2's analogous properties FG and $\bar{F}G$.

The λ-continuum has been superseded by several systems: Hintikka's α-λ continuum (1966), which extends the λ-continuum to improve the handling of inductive generalization; Carnap's Basic System (1971, 1980), which extends the λ-continuum by including predicates of unequal as well as equal widths; and Hintikka and Niiniluoto's K-dimensional system (1976), which axiomatizes a substantial portion

[11] Representative functions are so-called because they determine all other values within the system. Carnap replaced the term 'characteristic function' of (1952) with 'representative function', and his later usage is followed here.

of the α-λ continuum.[12] But all these systems have the same difficulty with singular analogy (Welch 1999). Various remedies have been proposed.[13] That to be pressed into service here originated with Kuipers (1984) as a counter-proposal to one by Niiniluoto (1980, 1981), who subsequently endorsed it (1988, p. 287).

Kuipers observes that we can view Eq. 2.1 as the application of the straight rule to n_Q real empirical instances of the strongest property Q and $\lambda(K)/K$ virtual logical instances of Q (1984, pp. 68–69). Why not then account for analogy by analogy with these virtual logical instances? That is, why not add virtual analogical instances of Q to factor in the relative similarities of properties?[14] Let each strongest property Q be associated with $\alpha_Q(e)$ virtual analogical instances that represent Q's similarity to the properties of the evidence. When $n \geq 1$, $\alpha_Q(e)$ is >0, but when $n=0$, the absence of evidence requires that $\alpha_Q(e)=0$. In addition, let $\alpha(e)$ virtual analogical instances represent the summation of similarities that all strongest properties have to the properties of the evidence. The ratio $\alpha_Q(e)/\alpha(e)$ would then indicate Q's portion of total similarity to the evidence. This ratio could therefore be added to Eq. 2.1 as an analogy factor comparable to the empirical and logical factors. Like the empirical and logical factors, the various analogy factors sum to 1. Where $0<\lambda<\infty$, Eq. 2.1 would become Eq. 2.3.[15]

$$p(h_Q \mid e_Q) = \frac{n_Q + \dfrac{\lambda(K)}{K} + \alpha_Q(e)}{n + \lambda(K) + \alpha(e)}. \tag{2.3}$$

Accordingly, the representative function for p* would be adapted for the new method p** as in Eq. 2.4.

$$p^{**}(h_Q \mid e_Q) = \frac{n_Q + 1 + \alpha_Q(e)}{n + K + \alpha(e)}. \tag{2.4}$$

Since the number of possibilities for analogy factors is unlimited, how could we determine the appropriate number of virtual analogical instances? Niiniluoto has described a natural way of measuring degrees of resemblance among strongest

[12] On the relation between the α-λ continuum and the K-dimensional system, see Kuipers (1978, p. 262).

[13] For the α-λ continuum, see Hintikka (1968, p. 228, 1969, pp. 28–33) and Pietarinen (1972, pp. 91–99). For the K-dimensional system, see Niiniluoto (1980, 1981, 1988), Spohn (1981), Costantini (1983), and Kuipers (1984).

[14] My development of this idea differs somewhat from Kuipers'.

[15] The resulting systems are unusual in that they are not indifferent to the order in which predicates are instantiated, thereby violating the axiom of individual symmetry upheld by Carnap (1952, p. 14, 1963, p. 975) and others (e.g., Maher 2000, p. 64). Nevertheless, the probabilities obtained from the various orders of instantiating predicates all converge to the same point (Kuipers 1984, p. 76). For those unwilling to give up the axiom of symmetry, steps toward a satisfactory treatment of analogy may be found in the work of Skyrms (1993) and Festa (1997).

properties (1981, pp. 12–14).[16] Where d_{uv} is the number of primitive properties not shared by the strongest properties Q_u and Q_v, their degree of resemblance r can be expressed by Eq. 2.5.

$$r_{uv} = \frac{1}{1 + d_{uv}}. \tag{2.5}$$

Given primitive properties F and G, for example, Eq. 2.5 determines the degrees of resemblance between the strongest property FG on the one hand and FG, $F\bar{G}$, $\bar{F}G$, and $\overline{FG}$ on the other to be 1, 1/2, 1/2, and 1/3 respectively.

Equation 2.5 affords a particularly simple way of determining appropriate analogy factors. Suppose initially that the evidence manifests just one strongest property. Where Q_u is this strongest property and Q_v is the strongest property of the hypothesis h_Q, let the value of r_{uv} be $\alpha_Q(e)$. Then, where Q_u is once again the strongest property of the evidence, total similarity $\alpha(e)$ would be given by Eq. 2.6.

$$\sum_{v=1}^{K} r_{uv}. \tag{2.6}$$

If the strongest property of the evidence is FG, then the analogy factors for FG, $F\bar{G}$, $\bar{F}G$, and $\overline{FG}$ would be $\frac{1}{7/3}$, $\frac{1/2}{7/3}$, $\frac{1/2}{7/3}$, and $\frac{1/3}{7/3}$ respectively. The factors are expressed in unreduced form to highlight the conceptual links with Eqs. 2.5 and 2.6.

In more complicated cases where more than one strongest property appears in the evidence, $\alpha_Q(e)$ is just the sum of the values of r_{uv} for each property Q_u of the evidence and the property Q_v of the hypothesis h_Q. For $\alpha(e)$ we note that the value of Eq. 2.6 for any strongest property of the evidence equals the value of the same equation for any other strongest property of the evidence, though the individual values of r are distributed differently. Hence where the number of strongest properties instantiated by the evidence is i, $\alpha(e)$ is expressed generally by Eq. 2.7. Examples emerge in Sect. 2.5.2 below.

$$i\sum_{v=1}^{K} r_{uv}. \tag{2.7}$$

Applying p** along these lines reflects the property analogies that unmodified p* does not. As we have seen, p* with $K = 4$ allots probabilities of 1/2 to both A_2's more similar conclusion 'Gb' and the less similar conclusion '$-Gb$'. Under the same assumptions, however, p** with analogy factors of $\frac{1/2}{7/3}$ for $\bar{F}G$ and $\frac{1/3}{7/3}$ for $\overline{FG}$ assigns probabilities of 9/17 (about .53) to A_2's more similar conclusion 'Gb' and 8/17 (about .47) to the less similar conclusion '$-Gb$'. This is not an iso-

[16] Kuipers (1984, p. 67, 73–74) and Niiniluoto (1988, pp. 279–280) offer alternative measures.

lated instance. p** is sensitive to property analogy wherever p* is not. We can use it, therefore, to estimate the probability on the premises of any singular analogical conclusion whatever.[17]

Although I have limited myself to p* for ease of illustration, any of an infinite number of alternative methods can be property-sensitized in the same way. Yet we brush up against a well-known difficulty in doing so: there are, after all, so many of these methods. Since different methods give different values, how do we know which one to choose? This is indeed a problem, but it seems not to have been noticed that there are situations where this embarrassment of methods does not matter at all. The reason is this: knowing merely that one conclusion is more probable than its rivals is sometimes enough; exactly how much more may be superfluous information. Suppose that to be the case with A_2, for example. Yet even though the probabilities of '*Gb*' and '$-$*Gb*' on A_2's premises vary from method to method, their comparative relations do not. '*Gb*' in this context is always more probable than its rival '$-$*Gb*'.

In these cases, probability can profitably be compared to temperature. There are alternative temperature scales, but since jumping from one to another preserves relations of hotter than and colder than, the main thing is to pick a scale and stick with it. Similarly, if all we need to know is which conclusion is more probable, the choice of a method is relatively unimportant. One such situation is described in Sect. 2.5.2.

To conclude this section, let us consider the bearing of these methods on non-monotonic reasoning. Take the stock argument about Tweety, who has become something of a non-monotonic celebrity:

Birds can fly.
Tweety is a bird.
Thus Tweety can fly.

Now if we add the premise 'Tweety is a penguin' to the original premises, we get the conclusion 'Tweety cannot fly'. The conclusions of the initial argument (call it Inference 1) and the augmented argument (Inference 2) are plainly contradictory.

This example appears to cover three basic cases. (1) The first premise might mean 'All birds can fly'. If so, this premise would be false, and we would reject both Inference 1 and Inference 2. (2) The first premise might mean 'Most birds can fly', and we might know that Tweety is a penguin in drawing Inference 1. Then Inference 1 would violate the requirement of total evidence, and we would therefore reject it. (3) The first premise might mean 'Most birds can fly', and we might not know that Tweety is a penguin in drawing Inference 1.

Case 3) covers three subsidiary cases. (3a) When drawing Inference 1, we might know that penguinhood is relevant but just not know how to classify Tweety. Then, when we learn that Tweety is a penguin, we would reject Inference 1. (3b) When drawing Inference 1, we might know that penguinhood is relevant but falsely believe that Tweety is not a penguin. Then, when we discover that Tweety is a penguin, we

[17] Carnap (1963, p. 75, 973–974) came to view these methods as approximations, and they are so regarded here.

would reject Inference 1. (3c) When drawing Inference 1, we might not know that penguinhood is relevant. Then, upon discovering that this background assumption is false, we would reject Inference 1.

Of these multiple possibilities, 1), 3b), and 3c) are quickly decided by appeal to the truth condition on cogent argumentation, and 2) just as quickly by the requirement of total evidence. The remaining case, 3a), is the only one directly relevant to the inductive methods under discussion. The application, I suggest, should be as follows. The conclusions of both Inference 1 and Inference 2 should be understood as following with some probability from their respective premises. The representative function for the method of choice, p** for example, could provide the probability that Tweety can fly given that we do not know whether or not she is a penguin (Inference 1). It could also determine the probability that she can fly given that we know that she is a penguin, and hence the probability that she cannot fly based on the same evidence (Inference 2). The problem is basically one of updating probabilities, and the representative function could supply the prior and posterior values for the probability transition.

2.5 Applying the Standard to Ethical Analogies

Section 2.4 proposed a standard for inductive cogency, situated analogy within the sphere of induction, and outlined a measure of inductive strength for analogical argumentation. The task of the present section is to put these ideas to work. I will attempt to illustrate how they could guide choices between arguments by analogy with phenomenally contradictory conclusions. Success in guiding such choices would mean that disagreements over core classification with vague predicates would be resolvable in principle. Resolving such a disagreement would amount to a reduction of the predicate's vagueness. Though the case study in Sect. 2.5.2 focuses on an ethical predicate, the vagueness of nonethical predicates could be reduced in fundamentally the same way.

2.5.1 Clash Points

We alluded to the early Platonic dialogues in Sect. 2.3, and we must briefly return. Despite their abstract definitional concerns, these dialogues are rooted in a practical problem that is never far beneath the surface. The *Euthyphro* is particularly explicit.[18] In the dialogue's discussion of piety, Socrates identifies two types of disagreement: those that cause hatred and social discord, and those that do not (7b–d). Those that do not include differences over number, size, and weight; those that do concern "the just and the unjust, the beautiful and the ugly, the good and the bad" (7d). These socially divisive disagreements are predominantly moral. However,

[18] The ensuing discussion builds on Welch (1997, pp. 1018–1021).

Socrates gets Euthyphro to see that people do not disagree about moral questions as such; they agree that the wrongdoer should be punished, for example, though they may disagree about whether someone is a wrongdoer (8c–d). These disagreements are disputes "about each action…. Some say it is done justly, others unjustly" (8e).

These socially divisive disagreements have a plainly identifiable root: the vagueness of terms like 'just'. The *Euthyphro* is built upon just this sort of occurrence. Euthyphro says his prosecution of his father is pious, but his family says it is not; Socrates' friends claim that Socrates' actions are pious, but Meletus counters that they are not. The term 'pious' is evidently vague. Structurally similar disagreements over whether an individual action is just or courageous or honest are just as troublesome in our own time as they were in Plato's, and they stem just as clearly from the vagueness of moral terms.

I will refer to disagreements of this concrete and socially divisive sort as *clash points*. Given their attendant social problems, clash points raise urgent normative questions: In such cases, is there a rational way of choosing sides? Provided the interlocutors are willing to give reasons for their views, there is. According to the analogy thesis, reasons for these clashing classifications must ultimately be arguments by analogy, and, as we have seen in Sect. 2.4, good analogies can be distinguished in principle from bad ones. If either of the analogies has a false premise, the analogy should be rejected. But if there is agreement over the truth of the premises, the disagreement over the conclusion must be rooted in conflicting views of the similarity between the controversial act and its prototype(s). That is a resolvable disagreement. Similarity and dissimilarity come in degrees, and quantitative inductive logics like p** provide metrics for measuring them.

The procedure is to form what John Kemeny has called the "minimal language" (1963, p. 722), the simplest language containing the singular and general terms of both premises and conclusion, and determine the conditional probability of the conclusion on the premises with the help of the representative function. If one of the analogies has the form of A_2, for instance, where *a* is the prototype, *b* is the disputed act, and *G* is the controverted moral property, we could argue that due to measurable degrees of similarity, we have more reason than not to assign *G* to *b* (details in Sect. 2.4.3). The predicate '*G*' would continue to be vague, but it would no longer be vague at the clash point.

2.5.2 The Grain Merchant

Suppose we try this out on an example that is complex enough to model real-world difficulties. In *De officiis*, Cicero presents the case of the grain merchant:[19]

> But there are occasions… which frequently arise when there is an apparent conflict between expediency and moral right; in such cases one must take a close look and see whether the conflict is a real one or whether it can be resolved. This category includes questions of the

[19] I follow the Higginbotham translation (Cicero [45–44 B.C.] 1967) except for rendering *frumentum* as 'grain' rather than 'corn'.

> following kind: if, for example, an honest merchant has brought a great quantity of grain from Alexandria to Rhodes at a time when the Rhodians are suffering from great famine and the price of grain is high, and if he knows that more merchants have set sail from Alexandria and has seen their ships on his way sailing in the direction of Rhodes laden with grain, is he to tell the Rhodians or keep quiet and get the best price he can for his cargo? We can assume that he is a wise and honest man, and can for the purposes of our discussion take it that he would not conceal the fact from the Rhodians if he thought it dishonest, but he is in doubt about its honesty. ([45–44 B.C.] 1967, III.12.50, pp. 153–154)

Cicero indicates that the case was a staple of Stoic moral discourse, and that it was a point of contention between Diogenes of Babylon and his disciple Antipater. Diogenes argued that since the grain merchant has not been asked whether the other ships are on the way, to say nothing about them is consistent with honesty. Antipater, on the other hand, thought silence dishonest. Diogenes and Antipater have gone the way of all flesh, but their disagreement has not; it can be revived in almost any contemporary audience. The term 'honest' is vague, and what we have is a genuine clash point.

To have any prospects for resolving it rationally, the disputants must agree about three things. There must be some consensus on prototypes, first of all, on certain actions as bearers of the problem predicate. This is no special requirement, however; agreement on the premises is indispensable for reaching consensus through any kind of inference, inductive or deductive. So let us assume that, as is sometimes the case, the disputants are looking back at the case from a later point in time—our own, for example. Let us say that while they have several disagreements about honesty, there is one (to keep it simple) point of agreement: the Weemsian prototype (Sect. 2.3.2). They agree that it was honest.

Just as important as agreement on a clear positive instance of the predicate is agreement on a clear negative.[20] Having agreed on the Weemsian prototype, we might suppose that George is asked whether he cut down the cherry tree but that he answers differently: "I can't tell a lie, Pa; you know I can't tell a lie. I did *not* cut it down with my hatchet." This action is clearly dishonest, and our knowledge that it is constitutes part of our evidence. Hence it should be included as a premise of the argument.

The final matter for agreement is what features of the case are morally relevant. Like all induction, moral induction requires selecting out, from all the properties of the objects under discussion, those that are relevant to the question.[21] These are the only properties that figure in the minimal language, and they are the primitive properties that make up the K strongest properties for determining relative width. Suppose, then, that the disputants agree that, with one exception, the template of

[20] I am indebted to Jesse Hughes for spotting a difficulty in an earlier version of this argument.

[21] Though the issue of relevance deserves more attention than I can give it here, I expect both widespread agreement and occasional disagreement over what counts as morally relevant. Such disagreement need not be unresolvable; our goals determine what is relevant. But the point for the moment is that 'relevant' is a vague term. Just as disagreement over borderline cases of a vague term may prevent determination of deductive soundness, disagreement over borderline cases of relevance may do the same for inductive cogency.

Sect. 2.3.2 identifies the morally relevant features. The exception is feature 1, the filial relation, which is counted out as irrelevant. Consequently, the features agreed to be relevant are: *g*'s belief that *p* (call this property *B* for short); *g*'s selfish desire that *f* not come to believe *p* (property *D*); *f*'s asking *g* whether *p* is true (property *A*); and *g*'s conveying to *f* that *p* is true (property *C*).

Where property *H* is honesty, *a* is the Weemsian prototype, *b* is the aforementioned negative instance where George conveys misinformation, and *c* is the problem case of the merchant, the analogical argument leading to Antipater's conclusion that the grain merchant's silence is not honest can be represented as:

A_3:

$Aa \wedge Ba \wedge Ca \wedge Da \wedge Ha.$
$Ab \wedge Bb \wedge -Cb \wedge Db \wedge -Hb.$
$-Ac \wedge Bc \wedge -Cc \wedge Dc.$
Hence $-Hc$.

Once the argument is explicitly set out, determining whether 'Hc' or '$-Hc$' is the better conclusion is straightforward. The question of whether the grain merchant's case is more similar to the clear positive or the clear negative case is actually evident by inspection:

Positive:	$ABCD$
Grain merchant:	$\bar{A}B\bar{C}D$
Negative:	$AB\bar{C}D$

In fact, the greater similarity of the merchant's case to the clear negative is easily quantified. Availing ourselves of Eq. 2.5, Niiniluoto's measure of resemblance between strongest properties, we have

$$\mathrm{r}(ABCD, \bar{A}B\bar{C}D) = 1/3$$
$$\mathrm{r}(AB\bar{C}D, \bar{A}B\bar{C}D) = 1/2.$$

The process of quantification can be carried a step further still by recalling the methods of Sect. 2.4. Because the disputants believe the argument's premises to be true, the clash over the conclusion must stem from divergent estimates of the support provided by the premises to the conclusion. Suppose we use p** to mediate. Since there are five primitive properties, there are $2^5 = 32$ strongest properties. The premises assert the strongest properties $ABCDH$ and $AB\bar{C}D\bar{H}$, and the rival conclusions (together with the third premise) imply the strongest properties $\bar{A}B\bar{C}DH$ and $\bar{A}B\bar{C}D\bar{H}$. To represent the relevant similarities, we can rely on Eqs. 2.5 and 2.7. They determine analogy factors of $\frac{1/4}{21} + \frac{1/2}{21} = \frac{3/4}{21}$ for $\bar{A}B\bar{C}D\bar{H}$ and $\frac{1/3}{21} + \frac{1/3}{21} = \frac{2/3}{21}$ for $\bar{A}B\bar{C}DH$. Then p** assigns a value of 21/41 (about 0.512) as the probability of A_3's conclusion on its premises and a probability of 20/41 (about

0.488) to the rival conclusion that the merchant's silence is honest. Marginally, then, and relative to the Washingtonian evidence, the grain merchant's silence is more dishonest than honest. The extension of 'honest' could therefore be clarified by subtracting the action of the grain merchant's silence. This would be one pass in the inductive molding of the concept of honesty.

Whatever else we may want to say about this conclusion, I think it accords well with at least one of our convictions about the case. Since the rival conclusions are intuitively very close, an assignment of .80, say, to one of them would clearly have been wrong. Nevertheless, this conclusion may not jibe with all of our convictions. It should do so only if our convictions are based on exactly the same evidence. The conclusion that the grain merchant's silence is honest can be shown to follow from a different evidential base. But it does not follow from this one.

As we noted above in connection with A_2, any of p**'s property-sensitized cousins will give slightly different probabilities for A_3's conclusion. Knowing which are the exact values would indeed be welcome. But here we are fortunate; that is beside the point. To resolve disagreements over clash points like the grain merchant's, all we really need to know is which conclusion is more probable, and that has been accomplished. It is an elementary matter to show that the greater probability of '$-Hc$' relative to 'Hc' on the premises of A_3 is invariant across these methods.

Up to this point our discussion of the grain merchant has tacitly assumed that the relevant predicates have equal logical weight. This assumption need not always hold, and it can be jettisoned at will. The key is to take a hint from Carnap's Basic System, which accommodates unequal as well as equal logical weights (1971, 1980). Each strongest property Q can be correlated with $\lambda_Q > 0$ virtual logical instances. The sum of the various λ_Q is λ, which represents total logical weight. Hence each strongest property can be outfitted with a logical factor λ_Q/λ that expresses its share of the system's logical weight. Like empirical and analogy factors, these logical factors must sum to 1. The result is the generalization of Eq. 2.3 (Kuipers' replacement for the representative function of the λ-continuum) expressed by Eq. 2.8.

$$p(h_Q \mid e_Q) = \frac{n_Q + \lambda_Q + \alpha_Q(e)}{n + \lambda + \alpha(e)} . \tag{2.8}$$

In the case of the grain merchant, for example, we might regard any strongest property that includes the combinations ACH or $A\bar{C}\bar{H}$ as particularly significant. We might be willing to allot more logical weight to the 8 strongest properties in which these combinations appear than to the remaining 24 properties. Instead of the single virtual logical instance employed under p**'s assumption of equal logical weight, we might apportion 1.5 logical instances to the more significant properties and 0.5 instances to the less significant ones. Since 8 "heavy" properties and 24 "light" ones would require a total of 24 virtual logical instances, each heavy property would have a logical factor of $\frac{3/2}{24}$ and each light property a logical factor of $\frac{1/2}{24}$. Like the analogy factors, the logical factors are expressed in unreduced form in order to make the number of their virtual instances transparent.

Using these logical factors and the same analogy factors as before, Eq. 2.8 determines the probability of A_3's conclusion '$-H$' on its premises to be 15/29 (about 0.517) and that of the rival conclusion 'H' to be 14/29 (about .483). Note that the probability of '$-H$' is slightly higher when the strongest property $\bar{A}B\bar{C}D\bar{H}$ receives .5 logical instances than when it receives 1 such instance (using p** above). Under our assumptions, Eq. 2.8 applied to A_3 is a decreasing function that approaches .5 as a limit when $\lambda_Q \to \infty$.

2.6 Conclusion

Understanding classification as proposed in this chapter has three principal advantages. The first is that it saves the phenomena. The phenomena identified in Sect. 2.1 are established classifications of an object as green, say, or an action as compassionate. Seeing classification as analogy permits us to explain these phenomena as analogical conclusions that meet the standard for inductively cogent arguments by analogy outlined in Sects. 2.4.1–2.4.3. The contradictories of these conclusions, on the other hand, fail conspicuously to meet this standard.

Secondly, the same approach that saves the phenomena furnishes a piecemeal solution to the problem of vague predicates. Recourse to the standard of inductive cogency permitted a solution to the vagueness of 'honest' in the case of the grain merchant and furnished an instance of inductive molding, as we saw in Sect. 2.5.2. I hasten to add, however, that appeal to inductive cogency can provide no more than an in-principle solution to the problem of vagueness. Even if correct, this solution cannot be expected to transmute acrimony into harmony at the world's clash points. The reasons can be gleaned immediately from two sets of considerations derived from the deductive special case.

The first difficulty is suggested by how little incidence a proof procedure for first-order logic has in our courts, congresses, and corporate boardrooms. The underlying causes were identified long ago by Socrates: the will to power and wealth of the interlocutors. These motives would sabotage any appeal to the standard for inductive cogency as well. But where debates are animated by the collaborative spirit of Socrates' conversation with Crito, it is possible to address the problem of vagueness by invoking this standard. The problem, of course, is that debates are rarely so motivated, even in philosophy.

Yet a second set of difficulties would remain even if we can count on the irenic spirit of Socratic dialogue. As with all logical formalisms, applying the standard for cogent arguments by analogy can be arduous. First, as we have noted, consensus through inference cannot be achieved without prior agreement on the premises. In the deductive case, vagueness, ambiguity, and lack of empirical information may create disagreement over the truth of an argument's premises, and the result can be disagreement over the soundness of the argument. The same snags can wreak the same havoc with inductive cogency. They can, in extreme cases, even impede consensus over prototypes. Second, inability to agree on which features are morally

relevant can actually prevent the premises from being formulated in a mutually acceptable way. One person might formulate premises with one set of relevant features, and another might insist on a divergent set. Finally, just as the deductive validity of very complex arguments can be determined in principle but only with difficulty in practice, the same is true of inductive strength. The evidence for a given conclusion may be quite complex. Nevertheless, the inductive methods discussed in this chapter are complete: for any noncontradictory premises and conclusions that are formulable in the language for which the method is defined, the probability of the conclusion given the premises can be determined in principle (Carnap 1952, pp. 16–18, 30–32).

Despite these cautionary notes, the theoretical difficulties attending the standard for cogent arguments by analogy are not insuperable. The same point holds for ethical and nonethical analogy alike. The focal predicate in the case of the grain merchant is ethical, but the vagueness of nonethical predicates can be reduced piecemeal through the same procedure of inductive molding. Though this is a hypothesis subject to further investigation, the formality of the present approach already provides strong confirmation. For the conclusion of any argument isomorphic to A_3 would be more probable than a contradictory conclusion based on the same premises regardless of whether the constituent predicates are ethical or not. In addition, this inductive procedure has the cognitive virtue of full coherence with our habitual criterion for sound deductive arguments, making it a natural candidate for a principled approach to vague predicates.

The third advantage of treating classification as analogy is that it provides a badly-needed corrective to one form of Aristotelian intuitionism. Aristotle thought we just *see* that we should be angry with a certain person in a certain way for a certain length of time. The decision, he claimed, "rests with perception" (*aisthēsis*), and perception is the work of the faculty of intuition (*noûs*).[22] Aristotle is disturbingly complacent about these intuitions:

> Hence any one who is to listen intelligently to lectures about what is noble and just and, generally, about the subjects of political science must have been brought up in good habits. For the facts are the starting-point, and if they are sufficiently plain to him, he will not need the reason [for the facts] as well; and the man who has been well brought up has or can easily get starting-points. (*Nicomachean Ethics* 1095b5–9)

The view that the decision rests with perception has proved to be attractive to moral particularists.[23] One reason for this is that the view does capture the psychological assurance we feel in moments of righteous anger, for instance. I grant that psychological description has its place. But it also has its limits. One limit of the view that the decision rests with perception is that it is critically toothless. If there is anything that experience teaches us, it is that mistakes are possible even in these moments of

[22] *Nicomachean Ethics* 1109b14–23, 1143a35–b5. Cf. Ross (1930, pp. 29–30, 41–42), Wiggins (1980, pp. 235–237), and Nussbaum (1990, pp. 54–104).

[23] For example, McDowell (1979, pp. 331–350), Nussbaum (1990, pp. 37–40, 54–105), and Dancy (1993, p. 50). However, T. H. Irwin maintains that Aristotle was not a moral particularist (2000).

high intuition. The person of "good habits"—Aristotle himself, for instance—might just *see* that an instance of slavery is just. We just happen to *see* it otherwise.

Italicizing the word 'see' is not an argument. Some way of adjudicating between rival intuitions is needed, and the standard for cogent arguments by analogy provides one. Aristotle's belief that slavery is just was not without rational support, as we saw at the outset of Sect. 2.4. He could have defended it by analogy with prototypically just cases of parents governing their children, for instance. And he could have claimed that the inferiority of slave to master is analogous to that of child to parent because both are due to biologically imposed limitations on reason. The slave, he remarked, has "no deliberative faculty at all" (*Politics* 1260a12–13; cf. 1254b20–24). But this is where the analogy breaks down, of course; any deliberative ineptness attributable to slaves as a class was imposed by nurture, not nature. Aristotle's argument by analogy includes a false premise, therefore. Because it includes a false premise, it fails to meet the standard of inductive cogency.

References

Adams, Ernest. 1998. *A primer of probability logic*. Stanford: CSLI Publications.

Aristotle. 1984. *Nicomachean Ethics*. In *The complete works of Aristotle,* ed. Jonathan Barnes, vol. 2, 1729–1867. Princeton: Princeton University Press.

Aristotle. 1984. *Politics*. In *The complete works of Aristotle,* ed. Jonathan Barnes, vol. 2, 1986–2129. Princeton: Princeton University Press.

Audi, Robert. 2004. Reasons, practical reason, and practical reasoning. *Ratio* 17:110–149.

Barker, Stephen F. 2003. *The elements of logic*. 6th ed. New York: McGraw-Hill.

Black, Max. 1967. Induction. In *The encyclopedia of philosophy,* ed. Paul Edwards, vol. 4, 169–181. New York: Macmillan.

Blackburn, Robin. 1997. *The making of new world slavery: From the Baroque to the Modern 1492–1800*. London: Verso.

Bok, Sissela. 1979. *Lying: Moral choice in public and private life*. New York: Vintage Books.

Carnap, Rudolf. 1942. *Introduction to semantics*. Cambridge: Harvard University Press.

Carnap, Rudolf. 1945. On inductive logic. *Philosophy of Science* 12:72–97.

Carnap, Rudolf. 1952. *The continuum of inductive methods*. Chicago: University of Chicago Press.

Carnap, Rudolf. 1963. Probability and inductive logic, My basic conceptions of probability and induction. In *The philosophy of Rudolf Carnap,* ed. Paul A. Schilpp, 71–77, 966–979. La Salle: Open Court (London: Cambridge University Press).

Carnap, Rudolf. 1971. Inductive logic and rational decisions, A basic system of inductive logic, part 1. In *Studies in inductive logic and probability,* eds. Rudolf Carnap and Richard C. Jeffrey, vol. I, 5–31, 33–165. Berkeley: University of California Press.

Carnap, Rudolf. 1980. A basic system of inductive logic, part 2. In *Studies in inductive logic and probability,* ed. Richard C. Jeffrey, vol. II, 7–155. Berkeley: University of California Press.

Cicero. 45–44 B.C. *De officiis*. English edition: Cicero. 1967. *On moral obligation: A new translation of Cicero's De officiis* (trans: Higginbotham, John). London: Faber.

Costantini, Domenico. 1983. Analogy by similarity. *Erkenntnis* 20:103–114.

Dancy, Jonathan. 1993. *Moral reasons*. Oxford: Blackwell.

Davidson, Donald. 1982. Rational animals. *Dialectica* 36:318–327. In *Actions and events: Perspectives on the philosophy of Donald Davidson,* eds. Ernest Lepore and Brian McLaughlin, 473–480. Oxford: Basil Blackwell, 1985.

Festa, Roberto. 1997. Analogy and exchangeability in predictive inferences. *Erkenntnis* 45:229–252.
Franklin, James. 1987. Non-deductive logic in mathematics. *The British Journal for the Philosophy of Science* 38:1–18.
Goldman, Alan H. 2002. *Practical rules: When we need them and when we don't*. Cambridge: Cambridge University Press.
Haack, Susan. 1978. *Philosophy of logics*. Cambridge: Cambridge University Press.
Hintikka, Jaakko. 1966. A two-dimensional continuum of inductive methods. In *Aspects of inductive logic,* eds. Jaakko Hintikka and Patrick Suppes, 113–132. Amsterdam: North-Holland.
Hintikka, Jaakko. 1968. Induction by enumeration, Induction by elimination and reply. In *The problem of inductive logic,* ed. Imre Lakatos, 191–216, 223–231. Amsterdam: North-Holland.
Hintikka, Jaakko. 1969. Inductive independence and the paradoxes of confirmation. In *Essays in honor of Carl G. Hempel,* ed. Nicholas Rescher, 24–46. Dordrecht: D. Reidel.
Hintikka, Jaakko, and Ilkka Niiniluoto. 1976. An axiomatic foundation for the logic of inductive generalization. In *Formal methods in the methodology of empirical sciences,* eds. Marian Przełęcki, Klemens Szaniawski, and Ryszard Wójcicki, 57–81. Dordrecht: D. Reidel (Wrocław: Ossolineum).
Irwin, T. H. 2000. Ethics as an inexact science: Aristotle's ambitions for moral theory. In *Moral particularism,* eds. Brad Hooker and Margaret Little, 100–129. Oxford: Clarendon Press.
Jeffrey, Richard C. 1985. Animal interpretation. In *Actions and events: Perspectives on the philosophy of Donald Davidson,* eds. Ernest Lepore and Brian McLaughlin, 481–487. Oxford: Basil Blackwell.
Jonsen, Albert R., and Stephen Toulmin. 1988. *The abuse of casuistry: A history of moral reasoning*. Berkeley: University of California Press.
Kemeny, John G. 1963. Carnap's theory of probability and induction. In *The philosophy of Rudolf Carnap*, ed. Paul A. Schilpp, 711–738. La Salle: Open Court (London: Cambridge University Press).
Kuipers, Theo A. F. 1978. On the generalization of the continuum of inductive methods to universal hypotheses. *Synthese* 37:260–272.
Kuipers, Theo A. F. 1984. Two types of inductive analogy by similarity. *Erkenntnis* 21:63–87.
Lakoff, George. 1987. *Women, fire, and dangerous things: What categories reveal about the mind*. Chicago: University of Chicago Press.
Las Casas, Bartolomé de. [1561] 1994. *Historia de las Indias*. Rpt. in *Obras completas*, vol. 5. Madrid: Alianza.
Mackie, J. L. 1977. *Ethics: Inventing right and wrong*. Harmondsworth: Penguin.
Maher, Patrick. 2000. Probabilities for two properties. *Erkenntnis* 52:63–91.
Margolis, Eric, and Stephen Laurence, eds. 1999. *Concepts: Core readings*. Cambridge: The MIT Press.
McDowell, John. 1979. Virtue and reason. *The Monist* 62:331–350.
Mill, John Stuart. [1872] 1973–1974. *A system of logic, ratiocinative and inductive*. Rpt. in Collected works of John Stuart Mill, vols. 7–8. Toronto: University of Toronto Press (London: Routledge & Kegan Paul).
Niiniluoto, Ilkka. 1980. Analogy, transitivity, and the confirmation of theories. In *Applications of inductive logic,* eds. L. Jonathan Cohen and Mary Hesse, 218–234. Oxford: Oxford University Press.
Niiniluoto, Ilkka. 1981. Analogy and inductive logic. *Erkenntnis* 16:1–34.
Niiniluoto, Ilkka. 1988. Analogy and similarity in scientific reasoning. In *Analogical reasoning,* ed. David H. Helman, 271–298. Dordrecht: Kluwer.
Nussbaum, Martha C. 1990. *Love's knowledge: Essays on philosophy and literature*. New York: Oxford University Press.
Ong, Walter J. 1982. *Orality and literacy: The technologizing of the word.* London and New York: Routledge.
Peirce, Charles S. [1878] 1986. Deduction, induction, and hypothesis. In *The writings of C. S. Peirce,* vol. 3, 323–338. Bloomington: Indiana University Press.

Peirce, Charles S. [1901] 1931–1958. The logic of drawing history from ancient documents. In *Collected papers of Charles Sanders Peirce,* eds. C. Hartshorne, P. Weiss, and A. Burks, 89–164. Cambridge: Harvard University Press.

Peirce, Charles S. [1902a] 1931–1958. Partial synopsis of a proposed work in logic. In *Collected papers of Charles Sanders Peirce,* eds. C. Hartshorne, P. Weiss, and A. Burks, 42–66. Cambridge: Harvard University Press.

Peirce, Charles S. 1902b. Application for support for his logic. MS L75 version 1, Memoir 19: On arguments. http://www.cspeirce.com/menu/library/bycsp/l75/ver1/l75v1-06.htm#m19. Accessed 30 April 2014.

Peirce, Charles S. [c. 1906] 1931–1958. Answers to questions concerning my belief in God. In *Collected papers of Charles Sanders Peirce,* eds. C. Hartshorne, P. Weiss, and A. Burks, 340–355. Cambridge: Harvard University Press.

Pietarinen, Juhani. 1972. *Lawlikeness, analogy, and inductive logic*. Amsterdam: North-Holland.

Plato. 1997. *Euthyphro.* In *Complete works*, ed. John M. Cooper, 1–16. Indianapolis: Hackett.

Putnam, Hilary. 1983. *Realism and reason. Philosophical papers*, vol. 3. Cambridge: Cambridge University Press.

Railton, Peter. 2003. *Facts, values, and norms*: *Essays toward a morality of consequence*. Cambridge: Cambridge University Press.

Reiss, Spencer. 1990. The last days of Eden. *Newsweek,* 3 December, 44–46.

Rosch, Eleanor. 1973. Natural categories. *Cognitive Psychology* 4:328–350.

Rosch, Eleanor. 1978. Principles of categorization. In *Cognition and categorization*, eds. Eleanor Rosch and Barbara B. Lloyd, 27–48. Hillsdale: Lawrence Erlbaum Associates. Rpt. in *Concepts: Core readings,* ed. Eric Margolis and Stephen Laurence, 189–206. Cambridge: The MIT Press (1999).

Rosch, Eleanor. 1983. Prototype classification and logical classification: The two systems. In *New trends in conceptual representation: Challenges to Piaget's theory?* ed. Ellin Kofsky Scholnick, 73–86. Hillsdale: Lawrence Erlbaum Associates.

Rosch, Eleanor, and Carolyn B. Mervis. 1975. Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology* 7:573–605.

Rosch, Eleanor, et al. 1976. Basic objects in natural categories. *Cognitive Psychology* 8:382–439.

Ross, W. D. 1930. *The right and the good*. Oxford: Clarendon Press.

Schroeder, Mark. 2009. Hybrid expressivism: Virtues and vices. *Ethics* 119:257–309.

Seneca. [63–65] 1920. *Epistulae morales*. Partial English edition: Seneca. *Epistles 66–92* (trans: Gummere, Richard M.). Cambridge: Harvard University Press.

Skyrms, Brian. 1986. *Choice and chance*. 3rd ed. Belmont: Wadsworth.

Skyrms, Brian. 1991. Carnapian inductive logic for Markov chains. *Erkenntnis* 35:439–460.

Skyrms, Brian. 1993. Analogy by similarity in hyper-Carnapian inductive logic. In *Philosophical problems of the internal and external worlds: Essays on the philosophy of Adolf Grünbaum,* eds. John Earman, Allen I. Janis, Gerald J. Massey, and Nicholas Rescher, 273–282. Pittsburgh: University of Pittsburgh Press.

Sorabji, Richard. 1993. *Animal minds and human morals: The origins of the Western debate*. London: Duckworth.

Sorensen, Roy A. 1988. *Blindspots*. Oxford: Clarendon Press.

Spohn, Wolfgang. 1981. Analogy and inductive logic: A note on Niiniluoto. *Erkenntnis* 16:35–52.

Tye, Michael. 1990. Vague objects. *Mind* 99:535–557.

Von Wright, Georg Henrik. 1963. *The varieties of goodness*. London: Routledge & Kegan Paul.

Waismann, Friedrich. 1930–1931. Logische analyse des wahrscheinlichkeitsbegriffs. *Erkenntnis* 1:228–248. English edition: Waismann, Friedrich. 1977. A logical analysis of the concept of probability. In *Philosophical papers,* 4–21. Dordrecht: D. Reidel.

Weems, Mason. [1809] 1962. *The life of Washington*. 9th ed. Cambridge: The Belknap Press of the Harvard University Press (Rpt).

Welch, John R. 1994. Science and ethics: Toward a theory of ethical value. *Journal for General Philosophy of Science* 25:279–292.

Welch, John R. 1997. Analogy in ethics: Pragmatics and semantics. In *The role of pragmatics in contemporary philosophy*, eds. Paul Weingartner, Gerhard Schurz, and Georg Dorn, vol. II, 1016–1021. Kirchberg am Wechsel: Die Österreichische Ludwig Wittgenstein Gesellschaft.
Welch, John R. 1999. Singular analogy and quantitative inductive logics. *Theoria* 14:207–247.
Welch, John R. 2007. Vagueness and inductive molding. *Synthese* 154:147–172.
Wiggins, David. 1980. Deliberation and practical reason. In *Essays on Aristotle's ethics,* ed. Amélie Oksenberg Rorty, 221–240. Berkeley: University of California Press.
Williams, Bernard. 1985. *Ethics and the limits of philosophy*. Cambridge: Harvard University Press.
Williamson, Timothy. 1994. *Vagueness*. London: Routledge.
Wisdom, John. 1969. *Philosophy and psycho-analysis*. Oxford: Basil Blackwell.
Wittgenstein, Ludwig. 1922. *Tractatus logico-philosophicus*. London: Routledge & Kegan Paul.
Wittgenstein, Ludwig. 1953. *Philosophical investigations*. Oxford: Basil Blackwell.

Chapter 3
Comparative Decision Theory

Abstract To prepare this work's subsequent discussions of morally instrumental and morally teleological discourse, Chap. 3 addresses the issue of theory choice. It surveys four views of how to choose a theory: probabilism, falsificationism, decision theory, and virtue epistemology. The chapter argues that each of these approaches has serious debilities, but those of decision theory are less debilitating than the rest. It therefore proposes a decision-theoretic approach to theory choice of any kind—moral and non-moral alike. However, attempts to apply decision theory to real-world problems confront a well-known difficulty: the exceptionally heavy information load the theory imposes on users. Bayesian decision theorists take the probability and utility functions that underlie expected utilities to determine sharp numeric values. Real-life decisions, however, must usually be made without nearly as much information. To ameliorate this difficulty, the chapter introduces a version of comparative decision theory. Because this form of decision theory requires no more than a bare minimum of comparative values for plausibilities and utilities, it can be widely applied. The chapter concludes with a consideration of three decision-theoretically foundational issues: transitivity, independence, and suspension of judgment.

3.1 Toward a Realistic Decision Theory

The previous chapter dealt with the first of three strata to be explored in this study: morally phenomenal discourse. The next two chapters address the remaining strata: morally instrumental and morally teleological discourse. Before doing so, however, we need to explore the issue of theory choice, for decisions about how to deploy morally instrumental and morally teleological language cannot be made independently of moral theory. Theory choice, in the sense to be employed here, is the act of selecting one theory rather than another for a cognitive purpose such as explanation or prediction.[1] This chapter aims to show that theory choice of any sort can be

[1] Theory choice and theory acceptance are distinct but intimately related acts. Whereas theory choice is selecting a theory for a cognitive purpose such as explanation or prediction, theory acceptance can be defined as the belief that a theory is true or as the mental state expressed by asserting a theory (cf. Maher 1993, p. 130, 133). One often chooses a theory because one accepts it. But since one might choose a theory for one purpose yet choose a rival for another, it is possible

J. R. Welch, *Moral Strata,* Theory and Decision Library A:, 49,
DOI 10.1007/978-3-319-08013-0_3,

carried out with the help of a novel form of decision theory.[2] The following chapters then enlist this version of decision theory as a guide to choice among moral theories.

Though decision theory has been criticized on various theoretical grounds (Elster 1979; Simon 1982; Levi 1986; Slote 1989; Elster 2000), its principal practical difficulty is the exceptionally heavy information load it imposes on users. In a vast number of situations, the numeric requirements for putting it to work are overly demanding. Strict Bayesian decision theorists take the probability and utility functions that underlie expected utilities to determine sharp numeric values (de Finetti 1937; Savage 1972). Real-life decisions, however, must usually be made without nearly as much information. Hence, the objection goes, decision theory is just inapplicable to the messy business of real-world decision making. Though real-world decision making is primarily that of human agents, it also includes decisions by software agents that may not be appropriately programmable with point-valued probabilities and utilities (Chu and Halpern 2008, pp. 4–5, 25).

Considerations like these have spurred repeated bids for a more realistic theory of decision. Some have tried to retrench by dropping back from point-valued to interval-valued functions (Kyburg 1979; Kaplan 1996).[3] Others recommend limiting a set of epistemically possible probability distributions to a realistic subset whose members are epistemically reliable (Gärdenfors and Sahlin 1982). Still others attempt to rescale decision-theoretic recommendations for ideal agents down to our actual, nonideal case (Weirich 2004; Pollock 2006).

This chapter also angles for a more realistic decision theory. It does so as a cousin to the interval-based approach of the previous paragraph, but it carries interval's retrenchment strategy to its outer limit. Just as point values require more precision than interval values, interval values require more precision than comparative values. What might we achieve by applying decision theory with a bare minimum of comparative values for probabilities and utilities?

The answer, I suggest, is 'more than you might expect'. To flesh out the details of this answer is the purpose of this chapter. It introduces a comparative version of decision theory that is really just a refinement of the reasoning that scientists and non-scientists alike manage in the course of ordinary decision making. To avoid raising unrealistic expectations, I ask the reader to keep in mind that the comparative decision theory outlined below is not completely general; like standard forms of decision theory, it does not always render a verdict. But it does offer reasonable verdicts in an overwhelming majority of cases. This claim will be focused on the limited context of theory choice.

---

to choose a theory without accepting it either in the sense of believing that it is true or in the sense of being disposed to assert it.

[2] Sects. 3.1, 3.3.1–3.5.3, and 3.6.3 draw heavily on Welch (2011, pp. 147–148, 149–169, and 159–160 respectively).

[3] Early work on interval probability functions includes Kyburg (1961), Good (1962), and Levi (1974, with further references on p. 407). Advocates of this approach include Good's less extreme Bayesians (1965, p. 10), today's robust Bayesians who take each probability to lie within an interval of values (Berger 1984; 1985).

The chapter's proposal is deployed in five stages. Rival strategies for dealing with the problem of theory choice are evaluated in Sect. 3.2. Second, the basic concepts required for implementing a decision-theoretic strategy are recapped and adapted to the special needs of theory choice in Sect. 3.3. The concept of relative disutility is then introduced and generalized in Sect. 3.4 to establish an isomorphism between information and utility. The results of Sects. 3.3–3.4 are brought to bear on comparative decision theory in Sect. 3.5. Finally, several foundational issues touched on in the course of the chapter are treated in more detail in Sect. 3.6. Even though the discussion does not range beyond theory choice, it presents comparative decision theory in terms sufficiently general to invite application in other contexts as well.

3.2 How to Choose a Theory

How to go about choosing a theory is a general problem that can be approached in at least four ways.[4] The oldest by far is probabilism. Probabilists recommend that those faced with competing theories choose the most probable of the lot. The intuition at probabilism's core is traceable at least as far back as the New Academy to Carneades' doctrine of the probable (Sextus Empiricus [c. 200] 2005, 1.159–1.189, pp. 34–39). Carneades argued that since there is no certain knowledge, we must rely on probability instead. But probability comes in degrees; hence the art of living consists in basing action on the highest available probabilities. This is the idea reflected in Joseph Butler's celebrated dictum that probability is "the very guide of life" ([1736] 1995, p. 5). It also coincides with the Rule of Maximum Probability, a rule for action that Carnap considered (and rejected): act as though the event with the highest probability is certain (1962, p. 255). More theoretically, Bruno de Finetti acknowledged points of contact between the philosophy of the New Academy and his version of probabilism ([1931] 1989, p. 220 n. 6). In elaborating this approach, he quoted Poincaré with approval: "On this account all the sciences would only be unconscious applications of the calculus of probabilities; to condemn this calculus would be to condemn science entirely" ([1931] 1989, p. 173).[5]

Probabilism was Karl Popper's *bête noire*, and our second approach was his riposte. Informative theories are not probable, he claimed; to the extent that they are informative, they are improbable.[6] Scientific theories should take the form of universal laws, and these laws have zero probability. Indeed, no amount of evidence could ever budge this probability from zero. All mathematical attempts to assign

[4] An earlier version of Sect. 3.2 appeared in Welch (2013, pp. 318–320).

[5] As noted, the topics of theory choice and theory acceptance are intimately related. An acceptance rule based on high probability is discussed in Festa (1999, § 3).

[6] Popper's claim has to be carefully qualified. Prior probability and information are inversely related, but posterior probability and information are not (Hintikka and Pietarinen 1966, pp. 106–107; Hintikka 1970a, pp. 8–9).

non-zero probabilities to universal laws lead to counter-intuitive results, Popper declared. But his fundamental objection is the Humean skeptic's point that to assign a non-zero probability to a universal law is to make an inductive leap; it is to commit the fallacy that piecemeal observation can verify universal statements. Hence it is not probability, he claimed, but falsifiability that is the mark of scientific theory. But what if there are competing falsifiable theories? Popper recommended that we choose the best-corroborated, that is, the theory that has most successfully resisted attempts to refute it. Formally, corroboration is defined for a hypothesis h based on evidence e and background knowledge b as a function of conditional probabilities p as in Eq. 3.1 (Popper 1983, p. 240).[7]

$$\frac{p(e \mid hb) - p(e \mid b)}{p(e \mid hb) - p(eh \mid b) + p(e \mid b)}. \tag{3.1}$$

A third approach to the problem of theory choice is decision-theoretic.[8] Its early advocates recommended it for decisions that are both practical (Carnap 1962, pp. 264–279; 1963, p. 74; 1971, pp. 8–9) and cognitive (Hempel [1960] 1965, pp. 73–79). Since it is applicable in principle to any kind of decision, it is applicable in particular to decisions about theories. Classical decision theory focuses on individual acts as units of choice,[9] and choosing a theory is an individual act. The trick is to assimilate theory choice to decisions under risk. Decisions under risk require two specialized functions. One is a probability function that assigns probabilities to all relevant states attributable to the world; the other is a utility function that assigns utilities to all outcomes of combining the acts under consideration with relevant states. The probability and utility functions permit one to estimate the expected utility of a given act (details in Sect. 3.3.1 below). Decision theory recommends that agents maximize expected utility; that is, that they choose an act whose expected utility is at least as great as that of any alternate act.

The last approach to theory choice to be canvassed here is the attempt to specify cognitive virtues that good theories possess and other theories do not—or do not to the same degree. Proponents of this virtue-epistemological approach include Thomas Kuhn, who touts accuracy, consistency, broad scope, simplicity, and fruitfulness (1977, pp. 321–322); W. V. Quine and Joseph Ullian, who single out conservatism, modesty, simplicity, generality, and refutability (1978, Chap. 5); Bas van Fraassen, who favors the pragmatic virtues of mathematical elegance, simplicity, great scope, completeness, unifying power, and explanatory power as well as the epistemic virtues of empirical adequacy, empirical strength, and consistency (1980, pp. 87–88); and William Lycan, who lists simplicity, testability, fertility, neatness, conservativeness, and generality (or explanatory power) (1998, § 1).

[7] For earlier attempts to define corroboration, see Popper (1959, App. *ix, p. 400; 1969, p. 58 n. 24).

[8] For accessible introductions, see Resnik (1987) and Peterson (2009).

[9] John L. Pollock argues that decision theory is properly applied to plans, that is, groups of acts (2006, Chap. 10).

Unsurprisingly, there are problems with each of these approaches. Virtue epistemology suffers from a number of debilities. To begin with, there is little consensus on just what the cognitive virtues are. Despite slight overlap, the Kuhn, Quine-Ullian, van Fraassen, and Lycan lists coincide explicitly only with respect to simplicity.[10] In addition, some of the virtues appear to be compounds constructed from subsidiary virtues. Predictive accuracy, for example, has at least five probabilistic components, and accuracy with respect to one need not ensure accuracy with respect to others (Eells 2000). Equally troubling is the well-known fact that the virtues are vaguely characterized; repeated attempts have been made to clarify the notion of simplicity, for example (Gavroglu 1989, pp. 547–548; Forster and Sober 1994; Richmond 1996; Kieseppä 1997; Sober 2002; Chang and Leonelli 2005, pp. 687–689; Baker 2007, pp. 194–196). Most seriously, perhaps, practitioners of this approach make no attempt to rank the virtues. As a result, they offer no guidance on how to make trade-offs among virtues (Kuhn 1970a, p. 262). Suppose that a given theory is simple but has little explanatory power, while one of its rivals has considerable explanatory power but is not simple. Should we opt for simplicity or explanatory power? Without answers to questions like these, we lack rational criteria for any nontrivial theory choice.

The case against probabilism has been prosecuted by Popper, and I believe he was right to this extent: probability is not all there is to the cognitive game. In particular, high probability is insufficient for rational theory choice.[11] Consider the choice between a working astronomical theory and a platitudinous alternative. Imagine that the working theory has had mixed success in predicting solar eclipses on specific dates and times while the platitudinous theory supports only imprecise predictions of eclipses at unspecified future times. The probability of the platitudinous theory would be much higher than that of the working theory, but astronomers would rationally choose to employ the working theory. In fact, if high probability were all that mattered, we could achieve our cognitive goals with a conservative gambit. We could play it safe, minimizing cognitive risk at every turn by refusing to affirm anything except the evidence (Levi 1967, p. 57). We would not end up with a theory, but we would have evidence statements with extremely high probability. That should satisfy probabilists. It would not satisfy theorists. Theorists are motivated by two cognitive passions: to know the truth and to obtain information.[12] To focus on high probability alone amounts to an obsession with obtaining truth at the expense of gaining information.

[10] Generality figures only in the Quine-Ullian and Lycan lists, though Kuhn's broad scope and van Fraassen's great scope may come to the same thing. Both Kuhn and van Fraassen include consistency. Quine-Ullian and Lycan concur on conservatism. Van Fraassen and Lycan both mention explanatory power, though Lycan adds that explanatory power "is just a wider, more global simplicity" (1998, § 1).

[11] High probability appears to be neither necessary nor sufficient for the related act of rational acceptance (Maher 1993, pp. 133–139, 163–180).

[12] This is a widely held view. See, e.g., Maher (1993, p. 208), Kuipers (2000, p. 93, 207), and Huber (2008, p. 92). James ([1897] 1979, pp. 24–27), Levi (1967, pp. 57–58), and Eells (2000, p. 205) express views that are similar.

Unfortunately, Popper's approach has grave difficulties in turn. A well-known objection motivated by Duhem-Quine holism urges that scientific theories are not tested in isolation. At a minimum, they are tested together with background assumptions and statements of initial conditions. Hence what is falsified is not the theory but the conjunction of theory, background assumptions, and initial conditions. Now suppose that we must act on the knowledge that something is wrong with our conjunction. Will corroboration guide us? Popper himself insists that "Our corroboration statements have no predictive import" (1974, pp. 1029–1030). Hence we cannot rationally predict that future data will be consistent with the theory, say, but inconsistent with background assumptions or initial conditions. A consistent Humean skeptic has no rational basis for choosing one of these possibilities over another. But this is practically devastating, as Wesley Salmon has pointed out (1981, pp. 115–125). Suppose that the suspect theory is the theory of gravitation—as in fact it is in light of the anomalous accelerations of the Pioneer, Galileo, and Ulysses space probes (Anderson et al. 1998). Then there would be no reason not to jump off the top of the Eiffel Tower, as Lakatos complained (Lakatos and Feyerabend 1999, p. 352 n. 203). But this, I submit, is all the reductio that any position needs.[13]

Then what about decision theory? Two lines of criticism are standard. The first takes issue with the decision-theoretic directive to maximize expected utility. Though this anti-maximizing approach can take different forms (Elster 1979; Levi 1986; Slote 1989; Pollock 2006), the best-known appears to be Herbert Simon's advocacy of satisficing (1982). According to Simon, the goal of maximizing expected utility is unreachable under real-world conditions. The practical strategy is to satisfice; that is, to aim at results that are good enough relative to some threshold of expected utility. The second criticism is that decision theory requires an unattainable level of numerical precision. Standard forms of decision theory call for probability and utility functions that determine sharp numerical values like 1/3 or 4/5. Understandably, many conclude that decision theory is therefore impractical, applicable above all in the rarefied worlds of textbooks (cf. Irvine 2006, p. 112).

With respect to the first criticism, the debate between maximizers and satisficers rages on (Byron 2004). Friends of maximization have argued that satisficing is rational only if optimific (Byron 1998); that we satisfice locally only to maximize globally (Schmidtz 2004; Narveson 2004); and that to satisfice rationally is a conceptual impossibility (Dreier 2004). Others maintain that maximization of expected utility forms the core of rationality (e.g., Jeffrey 1983, pp. 210–211; Larmore 2008, pp. 101–102). At the very least, then, maximization is a defensible strategy, even if not fully vindicated. But I will pursue this point no further in order to concentrate on the issue of impracticality.

[13] It is symptomatic that Popper concedes that after all he may need a "whiff" of induction: "there may be a 'whiff' of inductivism here. It enters with the vague realist assumption that reality, though unknown, is in some respects similar to what science tells us or, in other words, with the assumption that science can progress towards greater verisimilitude" (1974, pp. 1192–1193). This is hardly surprising given that even mathematics cannot get by without non-deductive reasoning (Franklin 1987).

To weigh the charge of impracticality, suppose we distinguish what we would like from decision theory and what we would need from it for practical decision making. Naturally, we would like to have a precise numerical value for the expected utility of each act under consideration. But we need to know only which act maximizes expected utility. This knowledge could plainly be acquired from probability and utility functions that generate point values. It might also be obtained from less precise functions that yield interval values. Conceivably, however, it might be gleaned from still less precise functions that return comparative probabilities and utilities. The response to the objection, then, is to relax the numerical demands of decision theory by dropping back from cardinal to interval or even comparative functions. This strategy is deployed in Sects. 3.3.2–3.5.3 below. If successful, it would bring decision theory more in line with the constraints of real-world decision making.

Let us sum up. The problems with the virtue-epistemological, probabilist, and falsificationist approaches are deeply debilitating. But decision theory's maximization strategy is at least defensible, and the problem of its numerical exigencies, while serious, can be ameliorated.

More positively, there is "nothing wrong with the standard theory" of decision-theoretic treatments of risk (Morton 1991, p. 81). In fact, there are two things quite right about it. The first is the role that decision theory allots to probability. This role strikes a happy medium, I think, between two toxic extremes: the Popperian minimum, which ignores probability altogether, and the probabilist maximum, which relies on probability alone. The second point in favor of decision theory is its inclusiveness. Many of the points emphasized by the other three approaches to theory choice are absorbed by decision theory. Like probabilism, decision theory stresses probability. Like falsificationism, decision theory warns that probability is not the whole story. Consistent with virtue-epistemological approaches that stress virtues such as consistency, simplicity, and coherence, decision theory can take these virtues into account as indicators of probability (cf. Lewis 1946, p. 346). And compatible with virtue-epistemological approaches that emphasize virtues like explanatory power, generality, accuracy, broad scope, completeness, and fertility, decision theory that is formulated in terms of information outcomes (as in Sect. 3.3.2.3 below) comfortably subsumes these virtues, the point of which is to obtain information. In my judgment, then, decision theory would appear to be the least imperfect of the four approaches—the best of a bad lot. I propose it, therefore, as a guide for decisions concerning theory choice.

This proposal can be seen as a simultaneous development of two lines of thought. One was pioneered by advocates of decision theory for cognitive choice, including Hempel (1960), Levi (1967; 1984), Hintikka (1970a), Kaplan (1981; 1996), Maher (1993), and others. The other is the comparative approach to evaluating scientific theories championed by thinkers like Kuhn (1970b, p. 77, 147), Laudan (1984, p. 29), Salmon (1990, p. 329, 331), and Sober (1999, p. 58).[14] But the comparative

[14] There are also intriguing connections to a contrastive account of knowledge (Morton and Karjalainen 2003; Schaffer 2004).

decision theory to be sketched in Sect. 3.5 diverges from both these lines of thought. It deviates from earlier decision-theoretic proposals because it requires no more than minimal precision in specifying probabilities and utilities. And it departs from the aforementioned comparative proposals, which make no reference to decision theory.

3.3 Decisions under Risk

People face decision problems. These problems are customarily divided into two classes: decisions under ignorance, which do not treat probabilities, and decisions under risk, which do. Since theory choices can be looked at as decisions under risk, decisions under ignorance will be ignored in this study. The basic concepts required for decisions under risk are reviewed in Sect. 3.3.1 and adapted to the context of theory choice in Sect. 3.3.2.

3.3.1 The Basics

Decisions under risk can be conceptualized by drawing on seven basic concepts: acts, states, outcomes, probability, utility, order, and decision rule. Though the decision-theoretic uses of these concepts will be all too familiar to most readers, I comment briefly on each in order to motivate adjustments to be proposed in the following section. The comments are based loosely on the work of von Neumann and Morgenstern (1953), Savage (1972), and Jeffrey (1983).

3.3.1.1 Acts

Acts can be thought of as propositions (Jeffrey 1983, pp. 83–85). A reluctant worker torn between going to work and staying home can be viewed as wavering between two propositions, only one of which she can make true. To choose an act is to make a proposition true. The number of acts under consideration by an agent at a given moment is assumed in this chapter to be finite.

3.3.1.2 States

Like acts, states can be understood as propositions. More specifically, states can be taken as propositions about features attributable to the world. The agent takes these features to be relevant to her decision. Propositions of interest to the reluctant worker might be that her supervisor would notice her absence and that her supervisor

would not. These propositions happen to describe concrete features of the world, but propositions relevant to other decisions can be abstract. Some abstract propositions are relevant to logic, which is "concerned with the real world just as truly as zoology, though with its more abstract and general features" (Russell 1919, p. 169). Other abstract propositions are mathematical, such as the proposition that the Goldbach conjecture is true, or ethical, like the proposition that the highest good is the greatest happiness of the greatest number (Sect. 5.4). Whether abstract or concrete, states are here taken to be finite.

3.3.1.3 Outcomes

Like acts and states, outcomes can be construed as propositions. Following a suggestion by Savage, Jeffrey suggests thinking of them as news items (1983, pp. 82–83). These news items result from pairing acts and states. In the case of the reluctant worker, for instance, the act of staying home and the state in which the supervisor notices might result in the news item that the supervisor docks the worker's pay. Since acts and states are finite in number, so are their outcomes.

3.3.1.4 Probability

A probability measure represents an agent's beliefs about states that can be attributed to the world. More specifically, a probability measure μ can be defined as a function from propositions about attributable states of the world to the unit interval. μ satisfies the usual axioms for nonnegativity, tautology, and finite additivity. The axioms can be stated for propositions q and r as follows:

Pr1. If q is contradictory, $\mu(q) = 0$.
Pr2. If q is tautologous, $\mu(q) = 1$.
Pr3. If q and r are mutually exclusive, $\mu(q \vee r) = \mu(q) + \mu(r)$.

Judgments about probabilities are expressed with varying degrees of sharpness. A probability may be characterized numerically as a point like 1/6 or as an interval such as 1/4–1/3. Yet Keynes claimed that "not all probabilities are numerical" ([1921] 2004, p. 65). This claim was criticized in its day by Jeffreys (1931, p. 223), who wanted to keep the link between probability and numbers as tight as possible. But even if all probabilities are ultimately numerical, epistemic limitations sometimes force us to talk about them with less than numerical precision. Probabilities can be described nonnumerically, either in qualitative terms like 'beyond a reasonable doubt' or in comparative terms such as 'greater than', 'equal to', or 'less than'. The probabilistic uses of both qualitative and comparative terms are amply documented in discussions of law, scientific theory, questions of conscience, documental authenticity, insurance rates, and much else (Franklin 2001).

3.3.1.5 Utility

A utility function represents the agent's preferences for possible outcomes due to choice. Typically, a utility function υ maps propositions about possible outcomes onto the real numbers. Utilities have been variously axiomatized, but the groundbreaking formulation was von Neumann and Morgenstern's axioms for preference over lotteries (1953, pp. 26–27). A particularly transparent version of these axioms due to Jensen requires completeness (full comparability and transitivity), continuity, and independence (1967, pp. 171–173).

3.3.1.6 Order

Classic formulations of decision theory assume that beliefs, desires, and preferences for acts have representations that are totally ordered. Because a probability measure represents an agent's beliefs by mapping them onto the unit interval, any belief in the probability measure's domain must be less probable, as probable, or more probable than any other. Analogously, whenever a utility function maps the agent's desires onto the reals, any desire in the function's domain must be less than, equal to, or greater than any other (e.g., von Neumann and Morgenstern 1953, p. 26). In much the same way, an agent's preference for a given act is assumed to be less than, equal to, or greater than that for any other act (e.g., Savage 1972, p. 18). The net result of these assumptions is to exclude the possibility of incomparable beliefs, desires, and preferences for acts.

3.3.1.7 Decision Rule

The six previous concepts can be employed to delimit a decision problem. A decision rule proposes a way of solving a decision problem. The classic decision rule for decisions under risk is to choose an act that maximizes expected utility. Let us say that we are considering the performance of an act a. Relevant to the decision are a finite number of attributable states of the world $s_1, s_2, \ldots, s_n$ and corresponding state probabilities $\mu(s_1 \mid e)$, $\mu(s_2 \mid e), \ldots, \mu(s_n \mid e)$ based on the evidence e. In addition, performance of a when exactly one of $s_1, s_2, \ldots, s_n$ obtains would produce outcomes $o_1, o_2, \ldots, o_n$ respectively, and these outcomes have utilities $\upsilon(o_1), \upsilon(o_2), \ldots, \upsilon(o_n)$. Then the expected utility EU of a given e can be defined by Eq. 3.2.[15]

$$EU_{a,e} = \sum_{i=1}^{n} \upsilon(o_i)\mu(s_i \mid e). \tag{3.2}$$

[15] Expectation can be defined for the infinite case by replacing summation with integration. See, for example, Pollock (2006, pp. 16–17).

Maximization of expected utility might be interpreted as a description of how people evaluate acts or as a norm about how people ought to evaluate them. This is an instance of the great divide between descriptive decision theory, which concerns how people go about choosing, and normative decision theory, which treats how people ought to go about choosing. Decision theorists like Savage (1972, pp. 19–20, 97) and Jeffrey (1983, pp. 166–167) take the normative route, and this study follows suit.

3.3.2 Adapting the Basics

Though I will continue to assume that acts, states, and outcomes are both propositional and finite, there is much about the decision-theoretic framework sketched in Sect. 3.3.1 that is either insufficient or unsatisfactory for a decision-theoretic treatment of theory choice. Hence the present section adapts this framework for use as a guide in choosing a theory. It does this by filling out and adjusting six of the seven foregoing concepts; the remaining concept, which is that of probability, is generalized as the concept of plausibility.

3.3.2.1 Acts

The acts at the heart of this chapter are acts of choosing a theory. To choose a theory, in the sense employed here, is to use one theory rather than another for a cognitive purpose such as explaining, predicting, or simply understanding some range of phenomena. The act of choosing a theory indicates some degree of belief in the theory, even if only the belief that the theory does explain, predict, or facilitate understanding of phenomena in the theory's domain. This degree of belief may, but need not, be high. Because degree of belief may change over time, a theory choice made at one time may be reviewed and revoked at a later one. Choosing a theory need not commit the agent to further acts such as communicating the decision to others, undertaking a research program, or betting the house on the theory's truth. These further acts would have to be considered independently, each in its own set of circumstances.

3.3.2.2 States

The states of the world that are relevant to theory choice are those posited by the theories under consideration, whatever they happen to be. These states can be phenomenal, instrumental, or teleological. As a practical matter, however, not all states posited by rival theories actually form part of the decision-theoretic matrix. Rather, the crucial states are chosen with an eye to contrast, for the states that matter in choosing a theory are those where the theories diverge. The propositions associated

with these states may be contradictories, such as whether the reluctant worker's supervisor would notice her absence or not, or contraries, such as whether the tempting holiday is in the spring or the fall.

3.3.2.3 Outcomes

The outcomes of theory choice can be quite varied, of course, with ramifications of economic, legal, political, military, social, aesthetic, and other sorts. Outcomes such as these are classified as noncognitive (or pragmatic) to distinguish them from cognitive outcomes like information gained and information lost. Where the relevant outcomes are purely cognitive, the decision can be classified as cognitive; where the situation is mixed, with relevant outcomes that are both cognitive and noncognitive, the decision is partly cognitive; and where the relevant outcomes are purely noncognitive, the decision is noncognitive. My concern in these pages is purely cognitive decisions (cf. Hempel [1960] 1965, pp. 75–76).

To say that a possible outcome of theory choice is information makes the question 'What is information?' decision-theoretically relevant. Some explications are statistical (Shannon 1948); others are semantic (Hintikka and Pietarinen 1966); still others are pragmatic (Levi 1984). For the limited purposes of this work, I will operate with a notion of information as reduction of uncertainty. This notion is sufficiently generic to be admitted by statistical (Floridi 2004, p. 198), semantic (Hintikka 1970b, p. 264), and pragmatic (Levi 1984, p. 65) points of view. Uncertainty, I will assume, is uncertainty about attributable states of the world. Take, for example, the proposition that impact from an asteroid or comet had nothing to do with the extinction of the dinosaurs—the most common view among paleontologists as late as 1988 (Bryson 2003, p. 195, 199). Luis and Walter Alvarez's hypothesis that just such an impact was responsible for the dinosaurs' extinction, once the impact of Comet Shoemaker-Levy 9 on Jupiter was observed and the Chicxulub impact crater identified, eliminated the no-impact possibilities for most paleontologists. The Alvarez hypothesis is therefore informative.

3.3.2.4 Plausibility

The problem of the numerical exigencies of decision theory was noted in Sect. 3.1. In an attempt to solve this problem, some talk of eliciting personal probabilities from betting preferences. To regard one state as more probable than another is to prefer to bet on one state rather than another for a valued prize (Fishburn 1986, p. 335). But the process of eliciting probabilities is plagued with difficulties. It can be enormously complex, first of all; to elicit even one's own probabilities can be a daunting task (Savage 1971, p. 795). Secondly, the procedure depends on having well-defined preferences to begin with, but this condition is frequently unsatisfied (Gilboa 2009, pp. 130–132). Finally, even when the condition is satisfied to a degree, the results can be quite imprecise. Savage frankly admitted the difficulty:

> The postulates of personal probability [used in Savage's decision theory] imply that I can determine, to any degree of accuracy whatsoever, the probability (for me) that the next president will be a Democrat. Now, it is manifest that I cannot really determine that number with great accuracy, but only roughly…. [A]s is widely recognized, all the interesting and useful theories of modern science, for example, geometry, relativity, quantum mechanics, Mendelism, and the theory of perfect competition, are inexact…. (1972, p. 59)

As an alternative to the minefield of probability elicitation, I propose that we replace probability with plausibility in the context of theory choice. Whereas a probability measure may map propositions about attributable states to numbers in the unit interval, a plausibility measure can map propositions about attributable states to members of any partially ordered set (Friedman and Halpern 1995). In this highly general conception, a plausibility function returns values that are bounded by nonnumeric limits $\top$ and $\bot$, where $\top$ represents the maximum plausibility and $\bot$ the minimum plausibility (Friedman and Halpern 1995). For any plausibility value x, therefore, $\bot \leq x \leq \top$. Though plausibility values are sometimes limited to the special case of the unit interval (e.g., Klir 2006, p. 166), they need not be numeric at all. Propositions about states can be mapped to qualitative plausibilities like high, low, intermediate, unlikely, nearly certain, and the like as long as they are partially ordered.

For our purposes, then, a plausibility measure π can be taken to map propositions about attributable states to plausibility values. π satisfies the following requirements for propositions q and r (Chu and Halpern 2004, pp. 209–210):

Pl1. If q is contradictory, $\pi(q) = \bot$.
Pl2. If q is tautologous, $\pi(q) = \top$.
Pl3. If q implies r, $\pi(q) \leq \pi(r)$.[16]

Comparing Pl1–3 to the earlier Pr1–3 for probability shows that the plausibility measure π generalizes the probability measure μ; probability is therefore a special case of plausibility. In fact, plausibility so defined appears to be the most general of the current approaches to representing uncertainty. Other standard representations, including Dempster-Shafer belief functions, possibility measures, and ranking functions, all turn out to be special cases of plausibility (Halpern 2003, Chap. 2).

3.3.2.5 Utility

The foregoing remarks about the paucity of reliable numeric probabilities also apply to utilities. Regardless of whether utilities are ultimately numerical or not, epistemic limitations sometimes force us to express them nonnumerically. As a matter of empirical fact, utilities are often expressed with varying degrees of quantitative sharpness. The utility of an outcome may be expressed numerically, either as a number or as an interval, but it can also be expressed nonnumerically, in qualitative

[16] I have adapted Chu and Halpern's set-theoretic statement to the propositional idiom.

terms like ‘good’, ‘bad’, or ‘outstanding’ (Halpern 2003, p. 165) or comparative terms such as ‘greater than’, ‘equal to’, or ‘less than’.

In the general case, decision-theoretic utility represents the evaluation of an outcome due to choice. In the special context of theory choice, therefore, utility would represent the evaluation of an outcome due to theory choice. As noted above, the outcomes of theory choice in purely cognitive decisions are information outcomes. Any theory that is more informative than another should receive, to that extent, a higher utility. Consequently, we have a norm for our normative version of decision theory: in the context of theory choice, utility should be proportional to information. This norm will prove useful in Sect. 3.4.2.

3.3.2.6 Order

The idea that preferences are totally ordered is doubtful in the extreme. Even von Neumann and Morgenstern were uneasy about it: “It is very dubious, whether the idealization of reality which treats this postulate as a valid one, is appropriate or even convenient” (1953, p. 630). Others have gone further, declaring it descriptively and normatively mistaken:

> Of all the axioms of utility theory, the completeness axiom [which posits a total order] is perhaps the most questionable. Like others of the axioms, it is inaccurate as a description of real life; but unlike them, we find it hard to accept even from the normative viewpoint. Does ‘rationality’ demand that an individual make definite preference comparisons between *all* possible lotteries (even on a limited set of basic alternatives)? For example, certain decisions that our individual is asked to make might involve highly hypothetical situations, which he will never face in real life; he might feel that he cannot reach an ‘honest’ decision in such cases. Other decision problems might be extremely complex, too complex for intuitive ‘insight’, and our individual might prefer to make no decision at all in these problems…. Is it ‘rational’ to force decisions in such cases? (Aumann 1962, p. 446; cf. Ok 2002; Ok et al. 2004)

I share these concerns; in fact, I would like to amplify them.

Suppose we are concerned with the conditional probabilities of two attributable states: s_1 and s_2. Although we can sometimes affirm that s_1 is more probable, equally probable, or less probable than s_2, sometimes we cannot. Even if we assume that, in the final analysis, any state is probabilistically comparable to any other (cf. Jeffreys 1961, p. 16), prior to the final analysis we may not be able to perform the comparison. The probability that Mexico City’s population will exceed 22,000,000 by 2022 and the probability that the first card drawn from this old and probably incomplete deck will be a heart surely defy any attempt to order them in a reasonable way.[17] For all practical purposes, then, probabilities are partially ordered (cf. Keynes [1921] 2004, pp. 29–30, 34).

Yet the cognitive situation may be worse than mere absence of numeric probabilities. An agent may be unable to say whether the plausibility of Mexico City’s population reaching a certain mark is greater than, equal to, or less than the plausibility

[17] The example is adapted from Fishburn (1986, p. 339).

of drawing a heart from a possibly incomplete deck, for example. At the crucial point of decision, some states turn out to be plausibilistically incomparable. At such points, therefore, plausibilities are partially ordered.

Like the plausibility of a state, the utility of an outcome may be viewed as greater than, equal to, or less than another. Regrettably, such comparisons cannot always be made. In a well-known example, Dewey and Tufts describe the value conflict of a citizen who wants to be loyal to his country but opposes his country's war (1932, pp. 174–175).[18] Even if these alternatives are comparable in some deep sense, epistemic limitations may prevent the citizen from carrying out the comparison. For all practical purposes, utilities are partially ordered.

To express the appropriate order relations, I will take the nonstrict comparative term '$\preccurlyeq$' as primitive. The $\preccurlyeq$ relation establishes a partial order; that is, it is reflexive, antisymmetric, and transitive.

When flanked by plausibility values, '$\preccurlyeq$' can be read as 'is less plausible than or equally plausible to' or 'is not more plausible than'. The following plausibility relations are immediately definable in terms of it, conditional plausibility ('|'), conjunction ('$\wedge$'), and negation ('$-$'):

$$\text{Infraplausibility } [\pi(s_1|e) < \pi(s_2|e)] =_{\text{df}} [\pi(s_1|e) \preccurlyeq \pi(s_2|e)] \wedge -[\pi(s_2|e) \preccurlyeq \pi(s_1|e)]$$
$$\text{Supraplausibility } [\pi(s_1|e) > \pi(s_2|e)] =_{\text{df}} -[\pi(s_1|e) \preccurlyeq \pi(s_2|e)] \wedge [\pi(s_2|e) \preccurlyeq \pi(s_1|e)]$$
$$\text{Equiplausibility } [\pi(s_1|e) = \pi(s_2|e)] =_{\text{df}} [\pi(s_1|e) \preccurlyeq \pi(s_2|e)] \wedge [\pi(s_2|e) \preccurlyeq \pi(s_1|e)]$$
$$\text{Incomparability } [\pi(s_1|e) \,||\, \pi(s_2|e)] =_{\text{df}} -[\pi(s_1|e) \preccurlyeq \pi(s_2|e)] \wedge -[\pi(s_2|e) \preccurlyeq \pi(s_1|e)].$$

This approach affords the philosophical advantage of neatly distinguishing equiplausibility from incomparability, which may not be possible when a strict relation like infraplausibility or supraplausibility is taken as primitive.

Like Savage, who used his primitive '$\leq$' to mean 'is not preferred to' for acts and 'is not more probable than' for events, I will use the primitive '$\preccurlyeq$' in different settings and allow its associated values to determine its sense. In addition to the plausibilistic usage just described, '$\upsilon(o_1) \preccurlyeq \upsilon(o_2)$' can be read as 'the utility of outcome o_1 is no greater than the utility of outcome o_2' and '$PE(a_1) \preccurlyeq PE(a_2)$' as 'the plausibilistic expectation of act a_1 is no greater than the plausibilistic expectation of act a_2' (plausibilistic expectation is defined below in Sect. 3.3.2.7). Taking these nonstrict relations as primitive, we can define relations of utility and plausibilistic expectation that are structurally parallel to infraplausibility, supraplausibility, equiplausibility, and incomparability.

3.3.2.7 Decision Rule

The preceding paragraphs amount to a retrofitting of decision problems under risk for theory choice. But decision problems plead for decision rules. What decision rule would be appropriate for theory choice?

[18] This example is discussed in Levi (1986, pp. 1–2).

To suggest an answer to this question, we begin by noting a historical process succinctly summarized by Savage. In discussing Daniel Bernoulli's advocacy of maximizing expected utility, he remarks:

> Between the time of Ramsey and that of von Neumann and Morgenstern there was interest in breaking away from the idea of maximizing expected utility…. This trend was supported by those who said that Bernoulli gives no reason for supposing that preferences correspond to the expected value of some function, and that therefore much more general possibilities must be considered. Why should not the range, the variance, and the skewness, not to mention countless other features, of the distribution of some function join with the expected value in determining preference? The question was answered by the construction of Ramsey and again by that of von Neumann and Morgenstern…; it is simply a mathematical fact that, almost any theory of probability having been adopted and the sure-thing principle [Savage's second postulate] having been suitably extended, the existence of a function whose expected value controls choices can be deduced. (1972, pp. 96–97)

The "mathematical fact" to which Savage refers was expressed in his representation theorem. Let the bare structure of his result be outlined as follows for a preference relation $\leq_A$ over a set A containing a finite number of acts $a_1, a_2, \ldots, a_n$ and expected utility EU (Sect. 3.3.1.7, Eq. 3.2):

$$a_1 \leq_A a_2 \text{ iff } EU(a_1) \leq EU(a_2).$$

What Savage showed, therefore, is that preferences for one act over another can be represented by comparative relations between expected utilities.[19]

Savage's "mathematical fact" has been clarified to an extent that he probably could not have foreseen. Francis Chu and Joseph Halpern have demonstrated that Savage's representation theorem has a plausibilistic generalization (2008, pp. 12–13). Similar to our treatment of EU, let us assume attributable states of the world $s_1, s_2, \ldots, s_n$ and corresponding state plausibilities $\pi(s_1 \mid e), \pi(s_2 \mid e), \ldots, \pi(s_n \mid e)$ based on the evidence e. We also assume that performing act a when exactly one of $s_1, s_2, \ldots, s_n$ obtains would produce outcomes $o_1, o_2, \ldots, o_n$ respectively, and that these outcomes have utilities $\upsilon(o_1), \upsilon(o_2), \ldots, \upsilon(o_n)$. In addition, we employ Chu and Halpern's notion of generalized expected utility, which is defined for an expectation domain $D = (U, P, V, \oplus, \otimes)$, where U is a set of utility values ordered by a reflexive binary relation $\lesssim_u$; P is a set of plausibility values ordered by a binary relation $\preccurlyeq_p$ that is reflexive, antisymmetric, and transitive; V is a set of expectation values ordered by a reflexive binary relation $\lesssim_v$; the multiplication-like operation $\otimes$ maps $U \times P$ to V; and the addition-like operation $\oplus$ maps $V \times V$ to V (2004, pp. 209–211, 2008, pp. 6–10). Then the generalized expected utility GEU of a given e can be expressed by Eq. 3.3.

$$GEU_{a,e} = \bigoplus_{i=1}^{n} \upsilon(o_i) \otimes \pi(s_i \mid e). \tag{3.3}$$

[19] Paul Samuelson's observation on Savage's theory (along with Ramsey's and de Finetti's) is still worth noting: "it is important to realize that this is a purely ordinal theory and the same facts can be completely described without using privileged numerical indicators of utility or probability" (1952, p. 670 n. 1).

Drawing on this notion of generalized expected utility and a preference relation $\leq_A$ over a set A containing a finite number of acts $a_1, a_2, \ldots, a_n$, Chu and Halpern show that

$$a_1 \leq_A a_2 \text{ iff } GEU(a_1) \leq GEU(a_2).$$

That is, preferences for acts can be represented by comparative relations between generalized expected utilities. However, the extreme generality of Chu and Halpern's approach blocks a plausibilistic analogue of Savage's uniqueness results whereby $\leq_A$ determines a unique probability measure and a utility function unique up to positive affine transformations (Chu and Halpern 2008, pp. 13–14).

The decision rule that corresponds to *GEU* is to maximize generalized expected utility. Chu and Halpern show that this rule is universal in the sense that it determines the same ordinal rankings as any decision rule that satisfies a trivial condition. The condition is that the rule weakly respect utility—roughly, that act preferences track outcome utilities for all constant acts (2004, p. 216, 219, 226–227). Constant acts have outcomes that are independent of states of the world (Savage 1972, p. 25).

I propose to adopt an instance of Chu and Halpern's decision rule as the main decision rule for theory choice. Let $\mathcal{D} = (\mathbf{U}, \mathbf{P}, \mathbf{T}, \mathbf{V}, \oplus, \otimes)$ be an expectation domain whose elements are defined as follows. **U** is the set of utility and disutility values $\{U, u, -u, -U\}$ such that $-U < -u < u < U$. **P** is the set of plausibility values $\{p, P\}$ such that $p < P$. $\otimes$ is the multiplication operation that maps $\mathbf{U} \times \mathbf{P}$ to **T**. **T** is therefore the set of product values $\{-UP, -uP, -Up, -up, up, uP, Up, UP\}$ to be used in calculating plausibilistic expectation. These values are ordered according to the following specifications:

(i) positive values: $up < uP < UP$; $up < Up < UP$; $uP \parallel Up$[20]
(ii) negative values: $-UP < -uP < -up$; $-UP < -Up < -up$; $-uP \parallel -Up$
(iii) mixed values: for all $x, y \in \mathbf{T}$, $(x < 0 \wedge y > 0) \rightarrow (x < y)$.

$\oplus$ is the addition-like operation that maps $\mathbf{T} \times \mathbf{T}$ to **V**. This operation, which is commutative, is defined for all $x, y \in \mathbf{T}$ and their absolute values $|x|$, $|y|$ as follows:

(i) $(x < 0 \wedge y < 0) \rightarrow ((x \oplus y) = -)$.
(ii) $(x < 0 \wedge y > 0 \wedge |x| < |y|) \rightarrow ((x \oplus y) = +)$.
(iii) $(x < 0 \wedge y > 0 \wedge |x| = |y|) \rightarrow ((x \oplus y) = 0)$.
(iv) $(x < 0 \wedge y > 0 \wedge |x| > |y|) \rightarrow ((x \oplus y) = -)$.
(v) $(x > 0 \wedge y > 0) \rightarrow ((x \oplus y) = +)$.

The operation remains undefined for the sums '$Up \oplus -uP$' and '$-Up \oplus uP$' since the absolute values of the addends are incomparable. Finally, **V** is the set of plausibilistic expectation values $\{-, 0, +\}$ ordered in the obvious way so that $- < 0 < +$.

The reader will have noticed that the values in the expectation domain $\mathcal{D}$ are both coarse and sparse. They have been chosen to reflect the merely comparative

[20] '||', which was defined in Sect. 3.3.2.6 for incomparable plausibilities, utilities, and plausibilistic expectations, is used analogously here. A product such as *uP* can be thought of as the plausibilistic expectation of an act relative to a single state.

discriminations that condition most real-life decision making. Judgments that the plausibility of one state is greater than that of another, for example, or that the utility of one outcome is equal to that of another, are not very precise. But they are precise enough to ground the comparative decision theory detailed in Sect. 3.5.

Relative to $\mathfrak{D}$, we assume attributable states $s_1, s_2, \ldots, s_n$ and a plausibility function π that maps propositions describing these states onto the values of **P**. In addition, we assume outcomes $o_1, o_2, \ldots, o_n$ and a utility function υ that maps propositions about these outcomes onto the values of **U**. Then the plausibilistic expectation *PE* of an act *a* given evidence *e* can be defined by Eq. 3.4.

$$PE_{a,e} = \bigoplus_{i=1}^{n} \upsilon(o_i) \otimes \pi(s_i \mid e). \tag{3.4}$$

Note that the definiens of Eq. 3.4 is typographically identical to that of Eq. 3.3, but their meanings are quite different; *PE* is a highly specific instance of *GEU*.

The decision rule that corresponds to *PE* would be to maximize plausibilistic expectation. This decision rule generalizes the rule based on *EU*, the standard definition of expected utility, in two directions at once: from probability to plausibility and from total to partial order. This rule is employed below in Sect. 3.5.

3.4 Relative Disutility

The comparative decision theory outlined in this chapter relies on the concept of relative disutility, present in all but name in a proposal by Jaakko Hintikka and Juhani Pietarinen (1966). Section 3.4.1 introduces and adapts the notion of relative disutility, and Sect. 3.4.2 generalizes it for theory choice. As the reader will observe, relative disutility is not meant to be applied indiscriminately; rather, it should be used only when reliable numeric utilities are not available.

3.4.1 The Hintikka-Pietarinen Proposal

Hintikka and Pietarinen proposed that the epistemic utility of a hypothesis *h* and its contradictory $-h$ be treated as follows:

> If *h* is true, the utility of his [the theorist's] decision is the valid information he has gained.… If *h* is false, it is natural to say that his disutility or loss is measured by the information he lost because of his wrong choice between *h* and $-h$, i.e., by the information he would have gained if he had accepted $-h$ instead of *h*. (1966, pp. 107–108; cf. Hintikka 1970a, p. 16)

This proposal for the utilities *u* of *h* and $-h$ can be summed up in Table 3.1, where s_h and s_{-h} are states of the world posited by *h* and $-h$ respectively.

Hintikka-Pietarinen utility assignments are highly suggestive. They suggest the possibility of generalization to include all cognitive options, not just hypotheses. They are also consistent with the view that cognitive choice is a two-person

Table 3.1 The Hintikka-Pietarinen proposal

	s_h	s_{-h}
h	$u(h)$	$-u(-h)$
$-h$	$-u(h)$	$u(-h)$

zero-sum game played by a truth-seeking self and nature (Hintikka 1983, p. 3). And they are strongly analogous to regret values for the minimax regret rule in decisions under ignorance (Peterson 2009, pp. 49–50). Regret, in fact, is a form of disutility.

To characterize Hintikka-Pietarinen utility assignments more precisely, we need to distinguish between intrinsic and relative utility.[21] The intrinsic utility of some outcome is its utility considered in itself, without reference to other utilities. By contrast, the relative utility of an outcome is its utility compared to another utility. The Hintikka-Pietarinen disutilities of $-u(-h)$ from choosing h when s_{-h} holds and $-u(h)$ from choosing $-h$ when s_h holds are relative disutilities.

Which type of utility is more appropriate: intrinsic or relative? In the context of theory choice, I submit that intrinsic utilities are inappropriate. Take the case of a false theory—Newtonian dynamics, for example. A plausible intrinsic utility for such a theory is zero. But the relative utility of Newtonian dynamics would vary with context. If Newton's theory is being compared to Buridan's impetus theory, its relative utility would normally be positive; but if it is being compared to relativistic dynamics, its relative utility would typically be negative. To invariably assign a utility of zero to cognitive options that are false would misdescribe the mechanics of cognitive choice, I think. For if the utilities of all false options are zero, one false option could not be reasonably preferred to another; yet one false option can be reasonably preferred to another; hence the utilities of all false options are not zero. The utility of a cognitive option, like so much else, depends on what is on the menu.

Though the comparative decision theory presented here draws on the Hintikka-Pietarinen proposal for epistemic utility, it differs from Hintikka and Pietarinen's approach in a number of nontrivial ways. Here I will mention only one. Hintikka and Pietarinen were concerned with the binary case of contradictory hypotheses, but many cognitive options are not so neatly related. The relevant options are often contraries instead of contradictories. Hence the comparative decision theory below generalizes Hintikka and Pietarinen's approach to cover more typical cases of cognitive choice where the options in play may be contraries as well as contradictories. This generalization is carried out in the following section.

3.4.2 Generalizing the Hintikka-Pietarinen Proposal

Initially, some terminology: total outcome, shared outcome, and unique outcome. An act's total outcome is its full set of consequences. An act's shared outcome is

[21] An analogous distinction can be drawn for probability. "Whether or not a given sentence is accepted depends not so much on its total probability taken in isolation, but on that probability as compared to the probabilities of the alternative hypotheses being considered" (Levi 1967, p. 98).

any part of its total outcome that is also obtainable by performing another act under consideration. An act's unique outcome is any part of its total outcome that is not obtainable by performing another act under consideration. A simple example: if one act pays off with a lottery ticket and a theater ticket while another act results in the same lottery ticket and a concert ticket, the lottery ticket is the shared outcome of both acts; the theater ticket is the unique outcome of the first act; and the concert ticket is the unique outcome of the second act.

To extend the concept of relative disutility to cognitive options that may be contraries as well as contradictories, we note that the information outcomes of choosing theories t_1 and t_2 may overlap or not. If they do not, the total outcome of an act is identical to its unique outcome. In such cases, the situation is only slightly more complicated than that envisioned by Hintikka and Pietarinen above. The choice of t_1 will result in the gain of any utility u_1 provided by that theory or the loss of any utility u_2 provided by the rival t_2. Conversely, the choice of t_2 will lead to the utility gain u_2 or the utility loss $-u_1$. The further complication is that, unlike the Hintikka-Pietarinen scenario, t_1 and t_2 may be contraries and therefore not jointly exhaustive. Hence there is a possibility of a third theory t_3. But we can deal with this third theory provided we can deal with the first two. That is, suppose that we bring the comparative decision theory outlined in these pages to bear on the choice between t_1 and t_2 and that t_2 turns out to be the winner. Then we can repeat the procedure for t_2 and t_3. As before, the outcomes of choosing the theories may or may not overlap. If they do not, we proceed as in this paragraph; if they do, we proceed as in the next one.

If the information outcomes do overlap, total outcome cannot be identical to unique outcome. Nor can total outcome be identical to shared outcome when the theories involved are contraries; contrary theories assure unique information outcomes. Let t_1 and t_2 be contrary theories with shared information outcomes. Since the shared outcome would be obtained regardless of whether we choose t_1 or t_2, it could not provide a reason for choosing one theory over the other. The principle of independence thus applies: "if two acts have the same consequences in some states, then the person's preferences regarding those acts should be independent of what that common consequence is" (Maher 1993, p. 10; cf. Behn and Vaupel 1982, p. 315).[22] In such cases, the principle of independence authorizes ignoring shared outcomes and choosing on the basis of unique outcomes alone. Now suppose that a unique outcome provided by t_1 has utility u_1 and a unique outcome provided by t_2 has utility u_2. Hence if we choose t_1, we miss out on any utility provided by t_2 but not by t_1; we could have obtained this utility by choosing t_2 instead. The disutility

[22] Maher takes independence to be a requirement of rationality "when the preferences are relevant to a sufficiently important decision problem, and where there are no rewards attached to violating… independence" (1993, p. 12, 63–83).

Table 3.2 Theory choice with information outcomes

	s_1	s_2
t_1	I	$-i$
t_2	$-I$	i

of this choice would be $-u_2$. Conversely, if we choose t_2, the disutility of this choice would be the loss of any utility uniquely provided by t_1. This disutility is $-u_1$.

In summary, the information outcomes of choosing t_1 and choosing t_2 are either entirely disjoint, in case they do not overlap, or have partial outcomes that are disjoint, in case they do overlap. In the latter case, the principle of independence licenses choice based on disjoint partial outcomes alone. Together, the two cases permit us to generalize the Hintikka-Pietarinen proposal to include cognitive options that are contraries as well as contradictories. For even if two theories are contraries, choice of one theory foregoes whatever unique outcome would result from choice of the other. Hence any utility attaching to an unrealized unique outcome would be lost as well. Consequently, if the unique outcome of choosing one theory has utility u, the relative disutility of the outcome of choosing the rival theory is $-u$.

Suppose we are faced with a choice between theory t_1, which posits state s_1, and theory t_2, which posits state s_2. If true (or justified), t_1 would yield unique information outcome I and t_2 unique information outcome i. Thus mistaken choices of t_1 and t_2 would result in the loss of i and I respectively. The choice can be summarized in Table 3.2.

Where $I > i$, the ranking of these outcomes according to degree of information would be:

$$I$$
$$i$$
$$-i$$
$$-I.$$

For utilities $U > u$, the Hintikka-Pietarinen utility scale parallels this information ranking at every point:

$$U$$
$$u$$
$$-u$$
$$-U.$$

This utility scale evidently complies with the norm that utility be proportional to information (Sect. 3.3.2.5), and it does so in the simplest possible way: information and utility turn out to be isomorphs.

Table 3.3 The basic binary case

Case	Plausibility	Utility
1	<	<
2	<	>
3	<	=
4	>	<
5	>	>
6	>	=
7	=	<
8	=	>
9	=	=

3.5 Comparative Decision Theory

Relying on the conceptual framework outlined in Sects. 3.3–3.4, we can now indicate how comparative decision theory could be applied in choosing a theory.[23] In its simplest and most decisive form, theory choice is the selection of one of two rival theories. Section 3.5.1 treats the binary case in its most basic form; Sect. 3.5.2 amplifies this treatment to include incomparable plausibilities and utilities; and Sect. 3.5.3 extends the binary strategy to the finite general case. I assume that the numeric data that would make the application of standard forms of decision theory feasible are unavailable.

3.5.1 The Basic Binary Case

We begin with the case of rival theories t_1 and t_2. Suppose that we can describe the relations between the states posited by these theories in terms of infraplausibility (<), supraplausibility (>), and equiplausibility (=) plus the corresponding utility relations as defined in Sect. 3.3.2.6. In other words, we assume for the moment that all plausibilities and utilities are comparable. (Incomparable plausibilities and utilities are treated in Sect. 3.5.2.) Evidently, there are nine possible cases. These cases are summarized in Table 3.3. To facilitate focus, I have adopted a kind of shorthand where '<' in the plausibility column, for example, abbreviates '$\pi(s_1|e)<\pi(s_2|e)$', which says that the plausibility of state s_1 posited by t_1 given the total evidence e is less than that of state s_2 posited by t_2 given e. Similarly, '<' in the utility column stands for '$\upsilon(o_1)<\upsilon(o_2)$', which says that the utility of outcome o_1 from choosing t_1 is less than that of outcome o_2 from choosing t_2.

These cases are conveniently divided into groups, the first of which is characterized by the relative uniformity of cases 1 and 5. Let us take case 1 as representative. If we adopt the conventions that 'U' and 'u' express higher and lower utility

[23] An alternative approach is explored by Ted Lockhart, who relies on ordinal probability rankings, second-order probabilities, integration to calculate average values, and the principle of indifference to address moral questions (2000, pp. 62–66, 71–72).

as before while '*P*' and '*p*' stand respectively for higher and lower plausibility, then t_1 promises utility u with plausibility p and disutility $-U$ with plausibility P. On the other hand, t_2 offers utility U with plausibility P and disutility $-u$ with plausibility p. To choose between the theories, we invoke the decision rule of maximizing plausibilistic expectation. Where plausibilistic expectation is defined by *PE* (Sect. 3.3.2.7, Eq. 3.4), the plausibilistic expectation of t_1 is

$$PE_1 = up \oplus -UP = -.$$

Similarly, *PE* provides the plausibilistic expectation of t_2:

$$PE_2 = -up \oplus UP = +.$$

Because $u < U$ and $p < P$, the plausibilistic expectation of t_1 is negative while that of t_2 is positive. By these lights, t_1 is inferior to t_2. A parallel argument for case 5 shows that t_2 is inferior to t_1.

A second group is comprised of cases 3, 6, 7, and 8, where one relation is = and the other is either < or >. In any choice with two relevant criteria and a tie with respect to one of them, the common-sense approach is to let the other criterion tip the scale. Say that there are two candidates for an academic post and two relevant criteria: prior publications and teaching ability. In the judgment of the hiring committee, the candidates are equal as teachers but unequal as scholars. Then the appointment should go to the better scholar. Case 3 illustrates this kind of thinking for theory choice. Here t_1 yields utility U with plausibility p and disutility $-U$ with plausibility P; for t_2, the plausibilities are reversed. According to *PE*, then, the plausibilistic expectation of t_1 is

$$PE_1 = Up \oplus -UP = -,$$

while that of t_2 is

$$PE_2 = -Up \oplus UP = +.$$

Because the theories are tied with respect to utility, the decision boils down to plausibility. That is, the ratio formed by the two plausibilistic expectations is determined entirely by their comparative plausibilities. This value is positive for t_2 and negative for t_1. On these grounds, then, t_2 is superior to t_1.

Case 9, which is akin to the cases of the previous group, constitutes a third group all by itself. Here not just one but both relations are =. *PE* would then give the plausibilistic expectation of both options as

$$PE = UP \oplus -UP = 0.$$

This result may seem surprising if we interpret '$PE = 0$' in intrinsic terms; it may seem that there is no advantage at all to adopting either option. But this impression is dispelled once we recall that we are working in relative terms. '$PE = 0$' means that there is no comparative advantage for either option—a tie.

A fourth group is formed by cases 2 and 4, whose physiognomy is mixed in the sense that each theory compares favorably in one respect but unfavorably in the

other. Consider case 2, for instance. Where t_1 promises more utility with less plausibility, its plausibilistic expectation would be

$$PE_1 = Up \oplus -uP,$$

while the plausibilistic expectation of t_2 would be

$$PE_2 = -Up \oplus uP.$$

Here, at last, we have run out of steam. As noted in Sect. 3.3.2.7, the $\oplus$ operation is undefined in both cases. There is no way to compare these plausibilistic expectations without knowing how much more desirable t_1 is and how much more plausible t_2 is. Nevertheless, these two cases will receive further discussion toward the end of this section.

One advantage of the comparative approach of the preceding paragraphs is that the difference between indifference and indecision is entirely transparent. Where the result is indifference (case 9), we have a good decision-theoretic reason to choose either theory. But where the outcome is indecision (cases 2 and 4), we have no decision-theoretic reason to choose at all.

Indifference is a form of suspending judgment.[24] To affirm 't_1 or t_2', for instance, is to suspend judgment concerning t_1 and t_2. But it is also to form a disjunctive judgment, structurally comparable to the disjunctive solutions proposed in the literature on moral dilemmas (Greenspan 1983, pp. 117–118; Gowans 1987, p. 19; Zimmerman 1996, p. 209, 220–221). To appreciate the work that disjunctive judgments do, take the four basic possibilities for binary choice of any kind: option 1, option 2, both option 1 and option 2, neither option 1 nor option 2. Now consider a disjunctive judgment about Sophie's choice, where a sadistic SS doctor attempts to force Sophie to choose which of her two children will survive (Styron 1979, Chap. 15, p. 589). Sophie clearly rejects the neither option. And she *might* feel compelled to accept the terms imposed by the doctor and reject the both option as well; she might think she should choose one child or the other even if she has no reason for choosing one over the other. Similarly, a theorist who forms the judgment 't_1 or t_2' (as in case 9) has already rejected the neither option. And since contrary and contradictory theories cannot be true (or justified) at all the same points, the theorist *might* feel compelled by circumstances to choose one of them even though she has no reason to choose it over its rival. Both Sophie and the theorist would have made disjunctive judgments that exclude two of the four basic options: neither and both.

We can now summarize the results for the basic binary case in Table 3.4. Of the nine cases listed there, six result in a unique recommendation (cases 1, 3, 5, 6, 7, 8); one results in a disjunctive recommendation (case 9); and two result in no recommendation (cases 2, 4). Consequently, seven of nine cases can be resolved in comparative terms.

Cases that result in no recommendation, where plausibility pulls in one direction while utility pulls in another, may seem tantalizingly close to resolution. We might wonder whether some additional decision-theoretic technique might supplement the

[24] The option of suspending judgment is discussed further in Sect. 3.6.3.

Table 3.4 The basic binary case with resolutions

Case	Plausibility	Utility	Resolution
1	<	<	t_2
2	<	>	no decision
3	<	=	t_2
4	>	<	no decision
5	>	>	t_1
6	>	=	t_1
7	=	<	t_2
8	=	>	t_1
9	=	=	t_1 or t_2

approach outlined in this chapter. Robert Behn and James Vaupel, for instance, apply decision theory by relying on "preference-probabilities" (1982). A preference-probability is an indifference probability, that is, the probability of obtaining the best outcome in a hypothetical reference gamble that makes the decision maker indifferent between an uncertain event and the reference gamble. Preference-probabilities function like utilities in calculations of expected utility. Behn and Vaupel manage to show that a reasonable decision can be made in some situations through simple comparison of the relevant preference-probabilities. Some of these comparison-friendly situations are double-risk dilemmas characterized by a choice between two alternatives, each of which entails a risk (1982, pp. 127–129). Similarly, theory choice as conceptualized in this chapter requires a choice between two theories, each of which entails the risk of utility loss.

Unfortunately, the resemblance between double-risk dilemma and theory choice turns out to be superficial. Double-risk dilemma in its general form entails a riskier choice that might result in the best or the worst possible outcome and a less risky choice with one or more intermediate outcomes. Theory choice in cases 2 and 4 lacks this contrast between more and less risky alternatives. The best outcome (U) is a possible outcome of one choice, while the worst outcome ($-U$) is a possible outcome of the other.

This difference is enough to spoil the comparative fun. To see this, let us take case 2 as representative and try the experiment of setting the expressions for the plausibilistic expectation of t_1 and t_2 equal to each other:

$$Up \oplus -uP = uP \oplus -Up.$$

This expression simplifies readily to

$$Up = uP$$

and states the condition under which the choice between t_1 and t_2 would be a decision-theoretic toss-up. Similarly,

$$Up > uP$$

would call for the choice of t_1 and

$$Up < uP$$

for the choice of t_2. All these expressions suffer from comparative symmetry in the sense that the greater utility that recommends t_1 is offset by the greater plausibility that favors t_2. In other words, simple comparison of plausibilities and utilities has taken us as far as it can. Cases like 2 and 4 have no general comparative solution.

Nevertheless, solutions are sometimes available on a case-by-case basis. Take a decision about whether or not to attend an academic conference, for example. Although serious accidents can happen at home, venturing out onto the highways and airways of the conference circuit most likely increases the plausibility of a serious accident. Hence plausibilistic considerations would encourage us to stay home, yet many of us decide not to do so. Why? Because we take the far greater utility of a conference to outweigh the slightly greater plausibility of a travel accident. On the other hand, if we were to judge the plausibility of a travel accident to be significantly higher than the plausibility of an accident at home, we would choose to stay home. This is a rational choice provided we assign a normal (very great) disutility to serious bodily harm. The writer once opted out of a conference in Turkey because he estimated the plausibility of war spilling over from northern Iraq to Turkey to be decidedly higher than the plausibility of an accident at home.

A second example of single-case resolution can be taken from the context of theory choice. Consider the negation of a theory t_1, where t_1 is a conjunction of descriptions of states attributed to the world.[25] By DeMorgan's theorem, this anti-theory would be a disjunction of negations that could be summarized as 'There is something wrong with t_1'. The plausibility of this anti-theory would normally be higher than that of t_1, yet it is nearly impossible to imagine circumstances where we would choose the anti-theory as a theory. We might affirm it as a truth, of course, just as I might affirm the truth that I am not a giraffe or not a pumpkin, but no one would confuse this with a theory. The anti-theory is so abysmally uninformative that we immediately identify it as theoretically trivial. In cases like these, then, we would almost invariably opt for t_1's much greater epistemic utility despite its lower plausibility.

3.5.2 The Full Binary Case

The prior presentation of the basic binary case ignored the possibility of incomparable plausibilities and utilities. To take these possibilities into account, we need to supplement the discussion of decision rules in Sect. 3.3.2.7. For once incomparability forces the door, our decision rule based on *PE* does not suffice. We have assumed that the $\preccurlyeq$ relation establishes a partial order on utilities, for example. Hence one utility may be neither greater than, equal to, or less than another; the two utilities may be incomparable. The unhappy result: in such cases, *PE* cannot be applied.

[25] Thanks to David Schrader for bringing this case to my attention.

Suppose that $?_1$ and $?_2$ are incomparable utilities and p_1 and p_2 are comparable plausibilities. Then the products in

$$(?_1 \otimes p_1) \oplus (?_2 \otimes p_2)$$

are incomparable, and the summation cannot be carried out.[26] Like *EU* (Sect. 3.3.1.7, Eq. 3.2), *PE* requires that outcome utilities be comparable. For analogous reasons, *EU* demands comparable state probabilities and *PE* comparable state plausibilities. But comparability of all utilities, probabilities, and plausibilities would appear to hold, if at all, *sub specie aeternitatis*.

Given that we must operate *sub specie temporis*, what would be a rational approach to these incomparabilities? Since they cannot be wished away, I am afraid the options are stark: to abandon decision theory whenever incomparability rears its head, or to adapt the decision rule to the situation. I would not object to abandoning decision theory provided we have a viable alternative. Unfortunately, I do not know what that would be. I conclude, then, that our best hope is to adapt the decision rule to the situation.

The intuition underlying my proposed adaptation can be introduced as follows. Suppose that an opening in Buddhist philosophy is to be filled by one of two specialists: one has published exclusively on philosophy of logic in Hindi; the other, exclusively on ethics in Mandarin. The departmental chair, who must make the decision, is utterly unable to read either language and lacks contacts with the relevant expertise. From the chairperson's point of view, the candidates' publications are not comparable, but their teaching abilities are comparable and unequal. If publications and teaching are the only relevant criteria, the chair should hire the better teacher.

If we generalize this intuition, we end up with something like the following norm: Where just two criteria are relevant to a choice but one is somehow inapplicable, we must fall back on the other criterion. To apply this norm to theory choice, we would have to consider two cases. The first is where utilities are comparable while plausibilities are not; the other, where plausibilities are comparable but utilities are not. I will refer to the first case as 'utility-comparable' and to the second as 'plausibility-comparable'. In both cases, the strategy is to ignore the incomparable values, since nothing useful can be extracted from them, and rely on the comparable values instead.

In the utility-comparable case, the expectation domain $\mathcal{D}$ would become $\mathcal{D}_u = (\mathbf{U}, \mathbf{P}, \mathbf{T}, \mathbf{V}, \oplus, \otimes)$ with elements defined as follows. **U**, **V**, and $\oplus$ have the same meanings as for *PE* (Sect. 3.3.2.7, Eq. 3.4). **P** is the set of incomparable plausibility values $\{p_1, p_2, \ldots, p_n\}$. $\otimes$ is defined so that incomparable plausibilities become right-identity elements; that is, for all $u \in \mathbf{U}$ and all $p \in \mathbf{P}$, $u \otimes p = u$. As a consequence,

[26] "In nonquantitative cases the principle to maximize utility may not apply because options' utilities may not exist. The absence of crucial probabilities and utilities may prevent computing them according to principles of expected utility analysis" (Weirich 2004, p. 59).

$\mathbf{T}=\mathbf{U}$. *PE* then reduces to the definition of utility-comparable expectation (*UCE*) in Eq. 3.5.

$$UCE_{a,e} = \bigoplus_{i=1}^{n} \upsilon(o_i). \tag{3.5}$$

The corresponding decision rule would be to maximize utility-comparable expectation.

In the plausibility-comparable case, the expectation domain $\mathfrak{D}$ would become $\mathfrak{D}_p=(\mathbf{U}, \mathbf{P}, \mathbf{T}, \mathbf{V}, \oplus, \otimes)$, defined analogously to $\mathfrak{D}_u$. **P**, **V**, and $\oplus$ retain their original meanings. **U** is the set of incomparable utilities and disutilities $\{u_1, u_2, \ldots, u_n, -u_1, -u_2, \ldots, -u_n\}$. $\otimes$ is defined so that incomparable utilities are left-identity elements and incomparable disutilities are negative left-identity elements; that is, for all $u, -u \in \mathbf{U}$ and all $p \in \mathbf{P}$, $u \otimes p = p$ and $-u \otimes p = -p$. Therefore, $\mathbf{T}=\{-P, -p, p, P\}$, ordered in the obvious way. Accordingly, *PE* contracts to the definition of plausibility-comparable expectation (*PCE*) in Eq. 3.6.

$$PCE_{a,e} = \bigoplus_{i=1}^{n} \pi(s_i \mid e). \tag{3.6}$$

The associated decision rule would be to maximize plausibility-comparable expectation.

My full proposal for decision rules thus amounts to this: for fully comparable situations, maximize plausibilistic expectation (*PE*); for utility-comparable situations, maximize utility-comparable expectation (*UCE*); and for plausibility-comparable situations, maximize plausibility-comparable expectation (*PCE*). Note that all three senses of expectation are special cases of Chu and Halpern's *GEU* (Sect. 3.3.2.7, Eq. 3.3).

Let me try to sum up the rationale for these rules as concisely as possible. The rationale is pragmatic, and it can be articulated around one belief, one desire, and two hard facts. The belief is that classical decision theory, for all its elegance and power, does not afford a realistic approach to theory choice. The desire is to move decision theory in a direction that will remedy this situation. The hard facts are numeric poverty and incomparability. Numeric poverty is the lack of reliable numbers for probabilities and utilities that characterizes most real-life decision making. Incomparability is the occasional but persistent inability to reasonably determine whether one probability, plausibility, or utility is greater than, equal to, or less than another. Together, these facts drive the shift from *EU* to *PE*. *PE* admits nonnumeric representations of beliefs and desires and, because it countenances partial orders, recognizes the reality of incomparability.

But, as we have seen, incomparability incapacitates *PE*; if two utilities, say, are incomparable, the summations of products it mandates cannot be performed. Hence the two options mentioned above: abandon decision theory in cases of incomparability, or adapt *PE* to the situation. The option of abandoning decision theory is doubly prohibitive: it runs directly counter to the project of developing a realistic decision-theoretic treatment of theory choice, and it leaves us high and dry without an alternative. Consequently, I think we should adapt *PE* to the situation. Granted, the special-case rules based on *UCE* and *PCE* are far from ideal. But their distance from the ideal does not mean *they* are faulty; rather, it reflects a defective *situation*.

Table 3.5 The full binary case

Case	Plausibility	Utility
1	<	<
2	<	>
3	<	=
4	<	||
5	>	<
6	>	>
7	>	=
8	>	||
9	=	<
10	=	>
11	=	=
12	=	||
13	||	<
14	||	>
15	||	=
16	||	||

The rationale for using them is basically no different than that for using the rule based on *PE*. The rationale for all three can be summed up in three words: 'Use comparable data!'

Possible objections to *UCE* and *PCE* are addressed in Sect. 3.6.3. The point for the moment is simply to illustrate how to apply them. Let us assume that the relations between the states posited by theories t_1 and t_2 can be described in terms of infraplausibility (<), supraplausibility (>), equiplausibility (=), and incomparability (||) plus the corresponding utility relations as defined in Sect. 3.3.2.6. Then there are 16 possible cases, which are summarized in Table 3.5. As in Tables 3.3 and 3.4, '<' in the plausibility column, for example, abbreviates '$\pi(s_1|e)<\pi(s_2|e)$', which says that the plausibility of state s_1 posited by t_1 given the total evidence e is less than that of state s_2 posited by t_2 given e. Similarly, '<' in the utility column stands for '$\upsilon(o_1)<\upsilon(o_2)$', which says that the utility of outcome o_1 from choosing t_1 is less than that of outcome o_2 from choosing t_2.

In addition to the groups already described for the basic binary case, we have a new group composed of cases 13, 14, and 15. These cases are characterized by a plausibility relation that is || and a utility relation that is not ||. These cases are then utility-comparable. If we apply the decision rule based on *UCE*, this amounts to ignoring the incomparable plausibilities and basing the decision on utility alone. In case 13, for example, the utility-comparable expectation of t_1 would be

$$UCE_1 = u \oplus -U = -,$$

while that of t_2 would be

$$UCE_2 = -u \oplus U = +.$$

Since the utility-comparable expectation of t_1 is negative while that of t_2 is positive, the choice would be t_2.

Table 3.6 The full binary case with resolutions

Case	Plausibility	Utility	Resolution
1	<	<	t_2
2	<	>	no decision
3	<	=	t_2
4	<	||	t_2
5	>	<	no decision
6	>	>	t_1
7	>	=	t_1
8	>	||	t_1
9	=	<	t_2
10	=	>	t_1
11	=	=	t_1 or t_2
12	=	||	t_1 or t_2
13	||	<	t_2
14	||	>	t_1
15	||	=	t_1 or t_2
16	||	||	no decision

Cases 4, 8, and 12 constitute an additional group in which the utility relation is || and the plausibility relation is not ||. The members of this group are plausibility-comparable. If we apply the decision rule based on *PCE* to case 4, the plausibility-comparable expectation of t_1 would be

$$PCE_1 = p \oplus -P = -$$

and that of t_2 would be

$$PCE_2 = -p \oplus P = +.$$

Evidently, therefore, since the plausibility-comparable expectation of t_2 is positive and that of t_1 is negative, we would opt for t_2.

Only case 16 remains. Like cases 2 and 5, it results in no decision, but it does so for a different reason. In cases 2 and 5, the decision-theoretic machinery breaks down. In case 16, however, the machinery cannot even start up; since both plausibility and utility are incomparable, decision theory has no grist for its mill. Here it can say nothing at all.

We are now in a position to summarize the results of our discussion in Table 3.6. Of the 16 cases, ten result in a unique recommendation; three result in a disjunctive recommendation; and three result in no recommendation. Hence 13 cases are resolvable in comparative terms; only three are not.

3.5.3 The Finite General Case

We have assumed from the beginning that the number of acts open to the agent at a given moment is finite. Consequently, the number of candidate theories under

consideration by the agent at a given moment is finite. This implication may appear false in light of the frequent observation that, at any given moment, there are an infinite number of theories from which to choose. I grant that there may be an infinite number of epistemically possible theories at any given moment, but there are not an infinite number of epistemically promising ones, that is, theories regarded as serious candidates by experts in the field (cf. Gärdenfors and Sahlin 1982, p. 366; Laudan 1984, p. 28). Famously, there were two serious candidates in the field of gravitational physics at the time of the solar eclipse of 1919: Newton's and Einstein's. That this was no exception is borne out by the history of science. The number of serious candidates at any given moment appears to be always, or almost always, finite and small. A theory with a parameter having a large—perhaps infinite—number of possible values is not viewed as a large number of theories; it is regarded as a single theory with a large number of parametric versions. The general theory of relativity is one theory despite many possible values for the curvature of space. Hence if we are always, or almost always, faced with a small number of serious candidate theories, the binary case is critical. For if it is possible to choose between t_1 and t_2 such that t_2, say, is the winner, then it is also possible in principle to hold a run-off between t_2 and any theory t_3—and so on successively.

This assumes that preference among theories is a transitive relation. The transitivity of preference is routinely affirmed by decision theorists (Savage 1972, p. 18; Jeffrey 1983, pp. 144–145; Maher 1993, p. 60), yet this affirmation has been repeatedly challenged. Before discussing some of these challenges, I would like to state the two following claims. Even if theory preference should turn out to be intransitive, first of all, the comparative approach to binary theory choice outlined in Sects. 3.5.1–3.5.2 would still go through, for transitivity is not an issue where only two theories are concerned. Hence the comparative route is always open for any two theories we care to evaluate. The second claim is that the assumption that theory preference is transitive, if properly understood and suitably employed, does in fact hold. The main consideration is to restrict transitive inference to the same sense of 'preference'. That is, we need to avoid equivocation.

To see the damage that equivocation can wreak, let us consider a relatively transparent instance. Max Black attempted to show that intransitive preferences can be rational by instancing job candidates A, B, and C who are rated for expertise, congeniality, and intelligence on a scale of 1–3, where 3 is high (1985). Their scores for each characteristic in the order mentioned are as follows:

A: 3, 2, 1
B: 1, 3, 2
C: 2, 1, 3.

Given these scores, an employer would prefer A to B (for expertise), B to C (for congeniality), and C to A (for intelligence). Hence, it appears, transitivity is violated but not rationality.

That transitivity is violated is a mere appearance, however, thrown up by simple equivocation. We have cycled from preferred-for-expertise to preferred-for-congeniality to preferred-for-intelligence. Jumbling these three senses of preference

together can create problems in much the same way as mixing binary, decimal, and hexadecimal numerals. Decision-theoretic preference is not preference for one characteristic and then another; it is preference overall. "I am concerned with preference *all things considered*, so that one can prefer buying a Datsun to buying a Porsche even though one prefers the Porsche qua fast (e.g., since one prefers the Datsun qua cheap, and takes that desideratum to outweigh speed under the circumstances)" (Jeffrey 1983, p. 225). If an employer were to have an *overall* preference for A to B, B to C, and C to A, we would have a genuine violation of transitivity. But we would also have a violation of rationality. The issue of transitivity is discussed in greater depth in the following section.

3.6 Shoring Up the Foundations

Readers making an initial pass at the conceptual framework of this chapter might opt to skip the present section, which is intended to provide a closer look at several foundational issues. Section 3.6.1 returns to the issue of transitivity, which was assumed in the discussion of the finite general case (Sect. 3.5.3). Section 3.6.2 treats the principle of independence, which figured in the generalization of the Hintikka-Pietarinen proposal (Sect. 3.4.2). Finally, Sect. 3.6.3 explores the possibility of suspending judgment, which was mentioned in passing in the exposition of the basic binary case (Sect. 3.5.1).

3.6.1 Transitivity

Transitivity has been challenged by counterexample—specifically, by attempting to exhibit intransitive yet rational preferences. One such attempt is due to R. I. G. Hughes (1980), who considers a voter who prefers candidates A to B to C on the basis of their policies. From the point of view of honesty, however, the voter's preferences are reversed: C to B to A. Now the voter might be such that when "in terms of honesty, two candidates are less than some critical distance apart, he neglects the difference between them; on the other hand, should they be widely separated the question of integrity becomes paramount, overriding all other considerations" (1980, pp. 132–133). Hence if the "distances" with respect to honesty between A and B and between B and C are less than the critical threshold while the distance between A and C is greater than this threshold, the voter would prefer A to B, B to C, yet C to A, thereby violating transitivity.

Hughes' criteria are nonlinear in the sense that small differences "do not matter much (or not at all) to the chooser, but as they add up…, their importance… might suddenly be much bigger" (Baumann 2005, p. 237). The impact of nonlinearity on theory choice has been studied by Peter Baumann, who claims that nonlinearity and multiplicity of criteria are individually necessary and jointly sufficient for some intransitive but rational preferences among theories (2005, p. 238). I submit, though,

that one advantage of applying decision theory to theory choice is that the criteria of plausibility and utility are not, in Baumann's sense, nonlinear. Nonlinearity admits reversal of priorities: the explanatory power of a theory, say, could be weightier than simplicity below a certain threshold, while simplicity could count more above this threshold. But as one can verify from the definition of plausibilistic expectation (*PE* in Sect. 3.3.2.7, Eq. 3.4), neither plausibility nor utility carries more weight than the other. Hence the reversal of priorities permitted by nonlinearity cannot take place. In addition, since decision-theoretic preference is ultimately based on the single criterion of plausibilistic (or probabilistic) expectation, the multiplicity condition may not be satisfied either, as Baumann appears to recognize (2005, p. 233 n. 4).

Within the confines of comparative decision theory, transitivity can be an issue at three different levels.[27] At the most basic level, if plausibility p_1 (or utility u_1) is greater than plausibility p_2 (or utility u_2), and p_2 (or u_2) is greater than plausibility p_3 (or utility u_3), then we might infer that p_1 (or u_1) is greater than p_3 (or u_3). Such inferences, including analogous inferences with the relations of equal to, less than, and incomparable to, exhibit *factor transitivity*. At a second level, whenever a product r_1 of utility and plausibility is equal to another such product r_2 and r_2 is equal to a third such product r_3, we may conclude that r_1 is equal to r_3. Along with similar inferences involving the relations greater than, less than, and incomparable to, these inferences instantiate *product transitivity*. Finally, if the plausibilistic expectation of act a_1 is less than the plausibilistic expectation of act a_2 and that of a_2 is less than that of act a_3, then we may deduce that the plausibilistic expectation of a_1 is less than that of a_3. Such inferences, including parallel inferences with the relations of greater than, equal to, and incomparable to, display *expectation transitivity*.

Factor transitivity and product transitivity are both relatively transparent and relatively peripheral to our present concerns. They are relatively transparent because the inferences involve a single, clearly articulated relation: >, <, =, or || as already defined for factor relata (Sect. 3.3.2.6) and analogously definable for product relata. For example, if one plausibility is greater than a second and this second is greater than a third, the plausibility of the first must be greater than the third. Nevertheless, factor transitivity and product transitivity are relatively peripheral for our purposes because the extension of comparative decision theory from the binary to the finite general case appeals to expectation transitivity, not factor or product transitivity. We want to be able to infer, say, that the expectation of act a_1 is equal to that of act a_3 simply because the expectation of a_1 is equal to that of act a_2 and that of a_2 is equal to that of a_3. The inference permits a comparative evaluation of a_1 and a_3 without having to compare them directly.

Let us therefore turn to expectation transitivity. Our discussion requires attention to three types of *homogeneous groups*. *Fully comparable groups* are composed of one or more binary comparisons in which the utilities and plausibilities are comparable in terms of greater than, equal to, or less than. For example, a comparison of act a_1 and act a_2 where the relevant utilities are comparable and the relevant plausibilities are comparable would constitute a fully comparable group.

[27] The remainder of Sect. 3.6.1 and Sect. 3.6.2 appear substantially as in Welch (2012, pp. 563–571).

Table 3.7 a_1 or a_3?

	$\pi(s_1) = p$	$\pi(s_2) = P$
a_1	$\upsilon(o_1) = U$	$\upsilon(o_2) = -u$
a_3	$\upsilon(o_3) = -U$	$\upsilon(o_4) = u$

Utility-comparable groups are formed by one or more binary comparisons in which the utilities are comparable while the plausibilities are incomparable. *Plausibility-comparable groups* consist of one or more binary comparisons in which the plausibilities are comparable but the utilities are incomparable.

Fully comparable groups appear to pose no problems for transitivity. The relata, which are plausibilistic expectations (*PE*), form a homogeneous set. If we are considering acts a_1, a_2, and a_3 where $PE(a_1) < PE(a_2)$ and $PE(a_2) < PE(a_3)$, then $PE(a_1) < PE(a_3)$.

The other homogeneous groups are similarly transparent. Consider acts a_1, a_2, and a_3 with outcome utilities that are comparable but state plausibilities that are incomparable. Since this is a utility-comparable group, the appropriate decision rule is based on utility-comparable expectation (*UCE* in Sect. 3.5.2, Eq. 3.5). Let $UCE(a_1) > UCE(a_2)$ and $UCE(a_2) > UCE(a_3)$. Then, straightforwardly, $UCE(a_1) > UCE(a_3)$. Parallel remarks apply to plausibility-comparable groups.

Unfortunately, not all groups are as well-behaved. The decision rules based on *PE*, *UCE*, and *PCE* all assume homogeneity: for a given choice, all the utilities and plausibilities are fully comparable, or they are all utility-comparable, or they are all plausibility-comparable. At times, however, some utilities and plausibilities may be fully comparable while others can be either utility-comparable or plausibility-comparable. These heterogeneous decision problems are characterized by *mixed groups*.

Watch what can happen when we attempt transitive inference across mixed groups. Let acts a_1 and a_2 be utility-comparable such that $UCE(a_1) > UCE\ (a_2)$. In addition, acts a_2 and a_3 are utility-comparable such that $UCE\ (a_2) > UCE\ (a_3)$. But let a_1 and a_3 be fully comparable. Their comparison is represented by Table 3.7, where π is a plausibility function; s_1 and s_2 are relevant states; p and P are state plausibilities such that $p < P$; υ is a utility function; o_1, o_2, o_3, and o_4 are outcomes of act-state pairs; and U, u, $-u$, $-U$ are outcome utilities such that $-U < -u < u < U$. According to the analysis of Sect. 3.5.1, the comparison of a_1 and a_3 should result in no decision because the decision-theoretic verdict is split: utility considerations favor a_1, but plausibility considerations favor a_3 (cf. case 2 of Table 3.4). But, having noted that the expectation of a_1 is greater than that of a_2 and that of a_2 is greater than that of a_3, we might venture the transitive inference that the expectation of a_1 is greater than that of a_3. This would be inconsistent with the conclusion that the pairwise comparison of a_1 and a_3 results in no decision.

What has gone wrong? The answer, in a word, is equivocation. The transitive inference that the expectation of a_1 is greater than that of a_3 is fallacious. It equivocates by conflating the utility-comparable expectations of a_1 relative to a_2 and a_2 relative to a_3 with the fully comparable expectation of a_1 relative to a_3. Instead, we should conclude that even though a_1 is decision-theoretically superior to a_2 and a_2 is decision-theoretically superior to a_3, there is no decision-theoretic reason to prefer

a_1 over a_3 or vice versa—unless, of course, we are willing to place more weight on either plausibility or utility.

Clearly, then, we need to restrict expectation transitivity. The restriction is that expectation transitivity must be limited to homogeneous groups. Transitivity can be invoked if all the comparisons are fully comparable, or if they are all utility-comparable, or if they are all plausibility-comparable. In insisting on this restriction, we are merely insisting on the same sense of 'expectation' in each case. Fully comparable expectation, utility-comparable expectation, and plausibility-comparable expectation are different, though closely related, concepts. Mixing them up can generate fallacies.

These brief considerations cannot pretend to establish the transitivity of preference, of course; the issue is a large one indeed (Maher 1993, Chap. 2). Though transitivity is common to both the Anglo-American and Franco-European schools of decision theory (Fishburn 1991, p. 115) and seems to be as widely accepted as any normative principle of rational choice, it is nonetheless controversial. Recent discussion includes both formidable challenges to transitivity (Willenken 2012; Temkin 2012) and defenses of the principle in the face of these challenges (Makin 2012; Huemer 2013). Given the ongoing controversy, we might do well to heed Patrick Maher's appeal to Chairman Mao: let "a hundred flowers blossom, and a hundred schools of thought contend." That is, "Since foundational arguments have been found inadequate to settle the [transitivity] issue either way, advocates of different positions should get to work developing theories based on their preferred principles. We can then use our judgments of the resulting theories to help decide between the principles" (Maher 1993, p. 62). This is the approach of the present chapter.

3.6.2 The Principle of Independence

The treatment of overlapping outcomes in Sect. 3.4.2 relied on the principle of independence, which licenses ignoring shared outcomes and concentrating on unique outcomes. The principle of independence is controversial. It figured as one of Savage's postulates under the guise of "the sure-thing principle" (1972, p. 21), was targeted by the Allais and Ellsberg paradoxes (Allais 1953, 1979a, 1979b; Ellsberg 1961), and continues to be affirmed in one form or another by Jeffrey (1983, p. 23), Levi (1986, p. 129, 144), and Maher (1993, p. 12, 83). While a full-blown discussion would be out of place at this point, I do want to acknowledge the controversy and say just a few words about it. The discussion proceeds through three stages: a brief exploration of the Allais and Ellsberg paradoxes; a search for solid ground for the principle of independence; and a suggestion about how to port trickier cases to this solid ground.

Some adherents of the principle of independence objected to Allais' original counter-examples because they unrealistically require ordinary people to make hypothetical choices with potential payoffs of hundreds of millions of dollars (e.g., Morgenstern 1979, p. 178, 180; Amihud 1979, pp. 151–152). In response, Allais

pointed out that maximizing expected utility can also create problems in realistic situations. The paradox introduced in the following paragraph is an instance. Even though it does not rely directly on independence, I concentrate on it because of its simplicity and proximity to ordinary reasoning. The treatment of this simple example can be extended in obvious ways to more complicated scenarios that do rely directly on independence, but I will not carry out this extension here.

Allais claims that "a person who is not generally considered as irrational, faced with a single, non-renewable choice, may well take ten dollars in cash rather than gamble on an even chance of winning $22 or nothing" (1979b, p. 539). Such a person, that is, might choose the sure $10 despite the fact that the choice is not the expected utility winner, and such a choice need not be irrational. In this and other examples, Allais' point is that maximizing expected utility neglects "the impact of the greater or lesser propensity for risk-taking or security, the consequence of which is, in particular, a complementarity effect in the neighbourhood of certainty" (1979b, p. 442). In short, the maximizing approach ignores "the considerable psychological importance attaching to the advantage of certainty as such" ([1953] 1979a, p. 88).

The first thing to be noticed about this example is that it trades on a mistake: the conflation of monetary outcomes and utilities. In order to make the point that the choice of $10 is not the expected utility winner, Allais assumes outright that the utilities of the outcomes of $0, $10, and $22 are 0, 10, and 22 respectively. At least since the time of Daniel Bernoulli, however, economists have recognized that the relation between money and its utility is not necessarily linear (cf. Levi 1986, p. 142). Hence the monetary outcomes of the example may or may not have utilities of 0, 10, and 22. For the sake of the argument, however, I will assume that they do.

The idea that acts should be evaluated by their consequences has been called a decision-theoretic "pre-axiom" by Peter Hammond (1988, p. 73), who argues that this consequentialist presupposition implies a number of standard axioms, including some forms of the principle of independence. Hence non-consequentialist preferences—preferences for a state or an act, for example—can lead to violations of these axioms. An example of a state-dependent preference would be a rosy view of a state of financial crisis because the Joneses would be poorer and the task of keeping up with them less onerous (cf. Hirshleifer 1965, p. 532). Perhaps, then, the security motivating those who would choose the sure $10 is a state-dependent preference.

However, as one writer on state-dependent preferences observed, "states may vary in respect to 'nonpecuniary income'" (Hirshleifer 1965, p. 532). This is a revealing observation. If income is a decision-theoretic consequence and there are nonpecuniary incomes, then such preferences may actually be consequence-dependent rather than state-dependent. Suppose we explore this idea in the context of the Allais paradoxes (cf. Maher 1993, p. 82). Consider our example with the sure $10 from the point of view of someone for whom $10 would be extremely important—someone who would urgently need the money for first aid, for example. In such a case, we would need to include the outcomes of first aid (FA) and no first aid (−FA) in addition to the already-specified monetary outcomes. The agent's predicament can then be represented by Table 3.8.

Table 3.8 An Allais problem with outcomes

	s_1	s_2
accept the \$10	\$10, FA	\$10, FA
accept the gamble	\$22, FA	\$0, –FA

Table 3.9 An Allais problem with utilities

	s_1	s_2
accept the \$10	u	u
accept the gamble	U	$-U$

What would be the utilities of these four outcomes? If we adapt the comparative resources developed so far in this chapter, the answer is straightforward: For a utility function υ with values $U > u > -U$,

$$\upsilon(\$22,\ \mathrm{FA}) = U$$
$$\upsilon(\$10,\ \mathrm{FA}) = u$$
$$\upsilon(\$0,\ -\mathrm{FA}) = -U.$$

The third utility assignment is based on the consideration that all utility from the outcome (\$22, FA) would be lost by accepting the gamble when s_2 obtains. This is actually a variation on the theme of relative disutility introduced in Sect. 3.4.1. The original concept of relative disutility could be called 'column disutility': disutility relative to other values in the column that represent what the agent might have enjoyed if she had acted differently. By contrast, the relative disutility employed in the third utility assignment would be 'row disutility': disutility relative to other values in the row that represent what the agent might have obtained had the world been different. Note that since there is no overlap between the outcomes (\$22, FA) and (\$0, −FA), there is no call to apply the principle of independence here.

Substituting these utilities for their associated outcomes in Table 3.8 yields Table 3.9.

We can now calculate the expected utilities *EU* of the two acts by applying Eq. 3.2 (Sect. 3.3.1.7). Since the probability of each state is ½, *EU*(accept the \$10)=u, and *EU*(accept the gamble)=0. The security-conscious agent we are considering should therefore accept the \$10 (though other agents faced with other outcomes might reasonably accept the gamble). We have arrived at the secure Allais result, but we have done so by simple maximization of expected utility.

Let us turn briefly to the Ellsberg paradox (1961, pp. 653–656). Imagine an urn known to contain 30 red balls and 60 black and yellow ones; the proportion of black to yellow is unknown. One ball is to be drawn at random from the urn. You are faced with two choices. Choice 1 is between act a_1, to bet on red with payoffs of \$100 if you win and \$0 if you lose, and act a_2, to bet on black with payoffs of \$100 if you win and \$0 if you lose. Choice 1 is summarized by Table 3.10. Choice 2 is between act a_3, to bet on red or yellow with payoffs of \$100 if you win and \$0 if you lose, and act a_4, to bet on black or yellow with payoffs of \$100 if you win and \$0 if you lose. Choice 2 is summarized by Table 3.11.

Table 3.10 The Ellsberg paradox: Choice 1

	red	black	yellow
a_1 bet on red	$100	$0	$0
a_2 bet on black	$0	$100	$0

Table 3.11 The Ellsberg paradox: Choice 2

	red	black	yellow
a_3 bet on red or yellow	$100	$0	$100
a_4 bet on black or yellow	$0	$100	$100

Unlike the Allais problem we have just discussed, the Ellsberg problem directly challenges the principle of independence. According to it, the yellow column in both choices should be ignored since its payoffs in Choice 1 are the same and its payoffs in Choice 2 are the same. If that is done, Choice 1 and Choice 2 become numerically identical. Hence anyone who prefers a_1 over a_2 should also prefer a_3 over a_4. But Ellsberg reports that many people prefer a_1 to a_2 yet also prefer a_4 to a_3. Others, though fewer, prefer a_2 to a_1 yet also prefer a_3 to a_4. Both preference patterns violate the principle of independence.

I submit that the twist in the Ellsberg problem is that both Choice 1 and Choice 2 are instances of what I will call 'laminated choice'. Each choice is constituted by a superposition of two layers, but one of these layers is explicit while the other is implicit. In Choice 1, the explicit layer is fully accounted for by Table 3.10. But the implicit layer is a further choice.

This further choice hinges on the difference between definite probabilities, such as the probability of drawing a red ball in this situation, and indefinite probabilities, such as that of drawing a black ball. The acts under consideration are to bet with definite probabilities and to bet with indefinite probabilities. The possible outcomes for this choice are a better chance, a worse chance, and an unchanged chance to win the bet. States, as noted in Sect. 3.3.1.2, can be thought of as propositions, and the states relevant to the choice at hand are states that affect the world's predictability. Of particular interest in this case are the states that definite probabilities facilitate accurate prediction more than indefinite probabilities, that indefinite probabilities facilitate accurate prediction more than definite probabilities, and that neither type of probability facilitates accurate prediction more than the other. For brevity, I will refer to these states of the world as 'favors definite', 'favors indefinite', and 'favors neither'. The choice in the implicit layer of the problem can then be summarized by Table 3.12.

What I am suggesting, then, is that Choice 1 can be summarized by two decision tables: Table 3.10, which features a_1 and a_2, and Table 3.12, which features a_5 and a_6. An analogous point holds for Choice 2. These laminated choices are possible because acts are subject to multiple true descriptions. One act can be truly described as both a_1, a bet on red, and a_5, a bet with definite probabilities. The decision tables for choices defined in these alternative terms are, as it were, superposed.

Table 3.12 The Ellsberg paradox: Choice 1's implicit layer

	favors definite	favors indefinite	favors neither
a_5 bet with definite probabilities	better chance to win	worse chance to win	unchanged chance to win
a_6 bet with indefinite probabilities	worse chance to win	better chance to win	unchanged chance to win

The superposition of choices can be made more transparent by noting the distinction that Ellsberg builds his entire analysis around: Frank Knight's distinction between measurable uncertainty or risk, which can be expressed by numerical probabilities, and unmeasurable uncertainty, which cannot be ([1921] 1971, pp. 19–20). The statistical probabilities of 1/3 for a red ball and 2/3 for a black or yellow ball are measurable uncertainties, but Ellsberg thinks that the problem's residual uncertainty is not probabilistic and not measurable (1961, p. 659).

Ellsberg is right, I believe, to think that there are two types of uncertainty here, but I would develop the contrast differently. All probabilities are plausibilities, as we noted in Sect. 3.3.2.4, but not all plausibilities are probabilities. Hence there are probabilistic and nonprobabilistic plausibilities. Both are present in the Ellsberg problem: the numerical probability of states like the drawing of a red ball and the comparative plausibility of states like favoring bets made with definite probabilities. The explicit layer of the problem relies on numerical probabilities; the implicit layer, on comparative plausibilities.

Looked at in this way, the Ellsberg problem no longer appears to violate the principle of independence. Take Choice 1, understood as a superposition of the choice between a_1 and a_2 and the choice between a_5 and a_6. The choice between a_1 and a_2 cannot be made by maximizing expected utility (*EU* in Sect. 3.3.1.7, Eq. 3.2). Although *EU* of a_1 could be determined, that of a_2 could not, for *EU* requires a definite probability of drawing a black ball, and that we do not have. Although it would be possible to estimate the probability of black in various ways—by defining upper and lower probability measures, for instance (Halpern 2003, pp. 25–28)—unless we are prepared to work with multiple probability measures and to recast our decision rule to accommodate them, there is no solution at the explicit layer.

But there is a solution at the implicit layer. Given that the utilities of the outcomes of a_5 and a_6 are evenly balanced, those who would prefer a_5 can do so reasonably if and only if they hold a plausibility function π that returns these comparative plausibilities of states:

$$\pi(\text{favors definite}) > \pi(\text{favors indefinite}).$$

That is, they would choose a_5 because they believe it offers them a better chance of winning the bet, and they believe this because, in effect, they are maximizing plausibilistic expectation (*PE* in Sect. 3.3.2.7, Eq. 3.4). Given π, a_5 turns out to maximize plausibilistic expectation. Analogously, Choice 2 cannot be made by maximizing *EU* for the choice between a_3 and a_4, but it can be made by maximizing *PE* for the choice between a_5 and a_6. In both cases, those who opt for a_5 could

Table 3.13 Grounds for the principle of independence

	s_1	s_2	s_3
a_1	o_1	o_2	o_3
a_2	o_4	o_5	o_3

do so for the same plausibilistic reason. And they could do so without violating the principle of independence.

The second stage of our discussion of the principle of independence is a search for solid ground for the principle. To initiate this search, consider the usual decision-theoretic case where state probabilities are independent of acts. As a simple illustration, take a case that is structurally similar to the Ellsberg problem. The possible outcomes o_1–o_5 of acts a_1 and a_2 given states s_1, s_2, and s_3 are reflected in Table 3.13, where o_1 and o_4, on the one hand, and o_2 and o_5, on the other, are assumed to be non-identical. Since the outcomes in the s_3 column are identical, their utilities are identical too. Provided that the probability of s_3 does not vary with the choice of a_1 or a_2, the products r_3 formed by o_3's utility and s_3's probability must therefore be identical as well. Where the remaining products of utility and probability are expressed in the obvious way, the expected utility EU of the two acts would be:

$$EU(a_1) = r_1 + r_2 + r_3$$
$$EU(a_2) = r_4 + r_5 + r_3.$$

Consequently, the relative magnitude of the acts' expected utilities is independent of the products r_3—just as the independence principle says. The same point holds for plausibilistic expectation. In these cases, then, independence is no more than elementary algebra. When state probabilities do not vary with acts, the principle of independence is on entirely solid ground (cf. Jeffrey 1983, p. 23).

The final stage of our discussion of independence concerns the remaining question: What if state probabilities do vary with acts? Here, of course, independence need not hold. Michael D. Resnik describes a decision about whether or not to smoke where the relevant states are contracting lung cancer and not contracting lung cancer (1987, pp. 15–16). Evidently, the probabilities of these states do vary with the acts of smoking and not smoking. But Resnik thinks the problem should be reformulated. Since not all smokers get lung cancer, there must be some protective factor that some people have and others do not. So Resnik proposes replacing the states of getting lung cancer and not getting lung cancer with four states related to this protective factor: having the protective factor and getting lung cancer from nonsmoking causes; having the protective factor and not getting lung cancer from nonsmoking causes; not having the protective factor and getting lung cancer from nonsmoking causes; and not having the protective factor and not getting lung cancer from nonsmoking causes (1987, p. 16). The probabilities of these states do not vary with the acts of smoking and not smoking.

I agree that the problem should be reformulated, but my suggestion is different. From the point of view of the person trying to decide whether to smoke, getting lung cancer and not getting lung cancer are not states at all. They are outcomes. The relevant states, on the other hand, can be very roughly described as having a

predisposition to lung cancer and not having a predisposition to lung cancer. The probabilities of these states, like the probabilities of Resnik's states, do not vary with the acts of smoking and not smoking. Consequently, when the decision is conceptualized in these terms, the principle of independence can be applied unproblematically.

In sum, the suggestion for dealing with states whose probabilities vary with acts is to attempt to reformulate them as states whose probabilities do not vary with acts. Whether this strategy can always be employed, or if not, when it can and cannot be employed, are questions for further research.

3.6.3 Suspending Judgment

As noted in Sect. 3.5.2, the rationale for the decision rules based on *PE*, *UCE*, and *PCE* can be summed up as 'Use comparable data!'. Some may object, however, that the use of *UCE* and *PCE* would be better replaced by a policy of suspending judgment. Here I consider two forms of this objection.

The first builds on the claim that decision-theoretic choice requires comparability of both plausibilities and utilities. When these conditions are not met, therefore, we can only suspend judgment. This would amount to relying exclusively on the decision rule associated with *PE* and rejecting the special-case decision rules based on *UCE* and *PCE*.

In response, I would recall William James' distinction between forced and avoidable options ([1897] 1979, pp. 14–15). A forced option, in James' sense, is a "complete logical disjunction" such as "Either accept this truth or go without it." Here logic forces a choice of exactly one alternative. But other options are characterized by pragmatic, not logical, force: a choice must be made in order to achieve some goal. Even though it is not logically necessary to choose one of a set of screwdrivers, for instance, it might be pragmatically necessary in order to set a screw. In much the same way, theory choice can be pragmatically forced. There are times when we want to explain, to predict, or to evaluate an experiment, and without choosing a theory we would not be able to proceed. When faced with these pragmatically forced options, the sensible response is to rely on the comparable data at hand, whether plausibilities and utilities, just utilities, or just plausibilities. This is the strategy underlying the decision rules based on *PE*, *UCE*, and *PCE*.

A variant of the objection in favor of suspending judgment is that all we really care about is utility; plausibility registers in decision-theoretic choice only to the extent that it maximizes utility. Hence the parallel modifications of *PE* that generate *UCE* and *PCE* are misguided. The decision rule based on *UCE* is acceptable because it plays the utility game, but the rule associated with *PCE* should be rejected and replaced by a policy of suspending judgment.

Three observations can be offered in response. If we take our cue on decision making from *EU*, *PE*, and *GEU*, there is no mathematical justification for favoring utility over probability or plausibility. Even though probability in *EU* and

plausibility in *PE* and *GEU* play markedly different roles than utility, the products in the summations that yield mathematical expectations are made up of one part probability or plausibility and one part utility. That is, probability or plausibility receives the same mathematical emphasis as utility.

In addition, the mathematics reflects what I take to be the right response to the following scenario. Suppose that attaining a cognitive goal requires you to choose between two theories whose information outcomes under relevant states of the world are known. Try as you might, however, you simply cannot rank one outcome as more desirable, equally desirable, or less desirable than the other. Nevertheless, the states of the world posited by one theory appear to be more plausible than the states posited by its rival. You need to choose a theory; how should you proceed? I can only suggest that ignoring what is known about the theories—the relative plausibilities of the states they posit—would be epistemically imprudent. This is the gist of the decision rule based on *PCE*.

Finally, there are historical considerations that buttress the *PCE*-based decision rule. *PCE* is intimately related to what is perhaps the most ancient cognitive practice of all: probabilism (Sect. 3.2). Since those who act as probabilists are typically focused on plausibility, not probability in the strict mathematical sense, 'plausibilism' would be a more accurate description than 'probabilism' (cf. Pigozzi 2009, p. 4). So understood, plausibilism would dictate *PCE* in plausibility-comparable situations. Though these historical considerations are not conclusive in themselves, the fact that *PCE* can be grafted onto this age-old cognitive tradition hardly strikes me as trivial. In our suite of three decision rules, then, it is *PCE* that can claim bragging rights for pedigree.

To conclude this section, I note that suspension of judgment is often motivated by concerns that have no intrinsic connection with information outcomes.[28] Suppose that a theorist is faced with the choice between t_1 and t_2, that the states posited by t_1 appear slightly more plausible than those posited by t_2, and that the utilities of the information outcomes from choosing either theory are equal. On strictly cognitive grounds, the theorist should choose t_1. (This is case 6 from Table 3.4.) But suppose, in addition, that there are further grounds for choice. Say that the theorist happens to be a candidate for a Nobel prize and that being right about t_1 would not improve her chances while being wrong would ruin them. Where $u(i_1)$ and $u(i_2)$ are the utilities of information outcomes i_1 and i_2, $u(f)$ is the utility of feeling relieved at not spoiling a chance at the Nobel, $-u(s)$ is the disutility of spoiling a chance at the Nobel, and '$+u(s)$' is short for '$--u(s)$', the situation can be represented as in Table 3.14.

In these circumstances, the act of suspending judgment produces the disutility of not feeling relieved due to a right choice of theory and the utility of not spoiling a chance at the Nobel. Thus the utility of suspending judgment is $-u(f)+u(s)$ regardless of whether s_1 or s_2 obtains. Provided the utilities of information and feeling

[28] This issue is addressed in a different context in Welch (2013, p. 327).

Table 3.14 Partly cognitive suspension of judgment

	s_1	s_2
choose t_1	$u(i_1) + u(f)$	$-u(i_2) - u(s)$
choose t_2	$-u(i_1) - u(s)$	$u(i_2) + u(f)$
suspend	$-u(i_1) - u(f) + u(i_1) + u(s)$	$u(i_2) + u(s) - u(i_2) - u(f)$

relieved are relatively small and the utility of not spoiling a chance at the Nobel is relatively large, the intuitively and decision-theoretically rational choice would be to suspend judgment. Yet problems of this sort are partly cognitive, not cognitive, and therefore fall outside the purview of this study.[29]

3.7 Conclusion

How would the results of applying comparative decision theory stack up against those from standard numeric forms of decision theory? Recall that the decision rule linked to plausibilistic expectation (*PE* in Sect. 3.3.2.7, Eq. 3.4) cannot be applied to cases with either incomparable utilities (4, 8, and 12 in Table 3.6) or incomparable plausibilities (13, 14, and 15 in Table 3.6). Consequently, we derived decision rules associated with *UCE* and *PCE* (Sect. 3.5.2, Eqs. 3.5 and 3.6) for these special cases. But the standard decision rule based on expected utility (*EU* in Sect. 3.3.1.7, Eq. 3.2) fails to be applicable to these very same cases. Still, if we employ the same tactics for *EU* as we did for *PE*, we could adopt special-case decision rules analogous to *UCE* and *PCE*, and these rules would yield comparable verdicts. As a result, standard numeric forms of decision theory would determine fifteen of sixteen cases in Table 3.6—only case 16 would remain unresolved. Comparative decision theory is marginally less effective in this sense, for it would determine theory choice in thirteen of the sixteen cases. In another sense, however, comparative decision theory is much more effective, for it can frequently be applied where numeric forms of decision theory cannot. A bare minimum of comparative inputs can return verdicts where more finely-tuned forms of decision theory return nothing at all.

I conclude with the observation that comparative decision theory is not restricted to the context of theory choice. In fact, it is not restricted by context at all. It can be applied anywhere provided the utility scale has the kind of symmetry illustrated here for the problem of theory choice. The results, as we have seen, are surprisingly good odds when faced with the usual human predicament: the need to decide without enough numbers.

[29] Distinctions among cognitive, partly cognitive, and noncognitive decisions are introduced in Sect. 3.3.2.3.

References

Allais, Maurice. 1953. Fondements d'une théorie positive des choix comportant un risque et critique des postulats et axiomes de l'école américaine. *Econometrie* 40:257–332. English edition: Allais, Maurice. 1979a. The foundations of a positive theory of choice involving risk and a criticism of the postulates and axioms of the American school. In *Expected utility hypotheses and the Allais paradox: Contemporary discussions of decisions under uncertainty with Allais' rejoinder,* ed. Maurice Allais and Ole Hagen, 27–145. Dordrecht: D. Reidel.

Allais, Maurice. 1979b. The so-called Allais paradox and rational decisions under uncertainty. In *Expected utility hypotheses and the Allais paradox: Contemporary discussions of decisions under uncertainty with Allais' rejoinder,* ed. Maurice Allais and Ole Hagen, 437–681. Dordrecht: D. Reidel.

Amihud, Yakov. 1979. Critical examination of the new foundation of utility. In *Expected utility hypotheses and the Allais paradox: Contemporary discussions of decisions under uncertainty with Allais' rejoinder,* ed. Maurice Allais and Ole Hagen, 149–160. Dordrecht: D. Reidel.

Anderson, John D. et al. 1998. Indication, from Pioneer 10/11, Galileo, and Ulysses data, of an apparent anomalous, weak, long-range acceleration. *Physical Review Letters* 81:2858–2861.

Aumann, Robert J. 1962. Utility theory without the completeness axiom. *Econometrica* 30:445–462.

Baker, Alan. 2007. Occam's razor in science: A case study from biogeography. *Biology and Philosophy* 22:193–215.

Baumann, Peter. 2005. Theory choice and the intransitivity of 'is a better theory than'. *Philosophy of Science* 72:231–240.

Behn, Robert D., and James W. Vaupel. 1982. *Quick analysis for busy decision makers*. New York: Basic Books.

Berger, James O. 1984. The robust Bayesian viewpoint. In *Robustness of Bayesian analyses,* ed. Joseph B. Kadane, 63–124. Amsterdam: North-Holland.

Berger, James O. 1985. *Statistical decision theory and Bayesian analysis*. 2nd ed. New York: Springer.

Black, Max. 1985. Making intelligent choices: How useful is decision theory? *Dialectica* 39:19–34.

Bryson, Bill. 2003. *A short history of nearly everything*. New York: Broadway Books.

Butler, Joseph. 1736. The analogy of religion, natural and revealed, to the constitution and course of nature. In *The works of Joseph Butler,* ed. W. E. Gladstone, vol. I. Bristol: Thoemmes Press, 1995.

Byron, Michael. 1998. Satisficing and optimality. *Ethics* 109:67–93.

Byron, Michael, ed. 2004. *Satisficing and maximizing: Moral theorists on practical reason*. Cambridge: Cambridge University Press.

Carnap, Rudolf. 1962. *Logical foundations of probability*. 2nd ed. Chicago: University of Chicago Press.

Carnap, Rudolf. 1963. Probability and inductive logic, My basic conceptions of probability and induction. In *The philosophy of Rudolf Carnap,* ed. Paul A. Schilpp, 71–77, 966–979. La Salle: Open Court. (London: Cambridge University Press).

Carnap, Rudolf. 1971. Inductive logic and rational decisions, A basic system of inductive logic, part 1. In *Studies in inductive logic and probability,* ed. Rudolf Carnap and Richard C. Jeffrey, vol. I, 5–31, 33–165. Berkeley: University of California Press.

Chang, Hasok, and Sabina Leonelli. 2005. Infrared metaphysics: Radiation and theory-choice. Part 2. *Studies in History and Philosophy of Science Part A* 36:686–705.

Chu, Francis C., and Joseph Y. Halpern. 2004. Great expectations. Part II: Generalized expected utility as a universal decision rule. *Artificial Intelligence* 159:207–229.

Chu, Francis C., and Joseph Y. Halpern. 2008. Great expectations. Part I: On the customizability of generalized expected utility. *Theory and Decision* 64:1–36.

De Finetti, Bruno. 1931. Probabilismo. Saggio critico sulla teoría delle probabilità e sul valore della scienza. Logos 14:163–219. English edition: De Finetti, Bruno. 1989. Probabilism: A critical essay on the theory of probability and on the value of science. *Erkenntnis* 31:169–223.

De Finetti, Bruno. 1937. La prévision, ses lois logiques, ses sources subjectives. *Annales de l'Institut Henri Poincaré* 7:1–68. English edition: De Finetti, Bruno. 1980. Foresight: Its logical laws, its subjective sources. In *Studies in subjective probability,* ed. H. E. Kyburg and H. E. Smokler, 53–118. New York: R. E. Krieger.

Dewey, John, and James H. Tufts. 1932. *Ethics*. Rev. ed. In The later works, 1925–1953, ed. Jo Ann Boydston, vol. 7. Carbondale and Edwardsville: Southern Illinois University Press, 1989.

Dreier, James. 2004. Why ethical satisficing makes sense and rational satisficing doesn't. In *Satisficing and maximizing: Moral theorists on practical reason,* ed. Michael Byron, 131–154. Cambridge: Cambridge University Press.

Eells, Ellery. 2000. Prediction, probability, and pragmatics. *Canadian Journal of Philosophy* 30:183–206.

Ellsberg, Daniel. 1961. Risk, ambiguity, and the Savage axioms. *Quarterly Journal of Economics* 75:643–669.

Elster, Jon. 1979. *Ulysses and the sirens: Studies in rationality and irrationality*. Cambridge: Cambridge University Press.

Elster, Jon. 2000. *Ulysses unbound: Studies in rationality, precommitment, and constraints*. Cambridge: Cambridge University Press.

Festa, Roberto. 1999. Scientific values, probability, and acceptance. In *Incommensurability and translation: Kuhnian perspectives on scientific communication and theory change,* ed. R. Rossini Favretti, G. Sandri, and R. Scazzieri, 323–338. Cheltenham: Edward Elgar.

Fishburn, Peter C. 1986. The axioms of subjective probability. *Statistical Science* 1:335–358.

Fishburn, Peter C. 1991. Non-transitive preferences in decision theory. *Journal of Risk and Uncertainty* 4:113–134.

Floridi, Luciano. 2004. Outline of a theory of strongly semantic information. *Minds and Machines* 14:197–221.

Forster, Malcolm, and Elliott Sober. 1994. How to tell when simpler, more unified, or less *ad hoc* theories will provide more accurate predictions. *The British Journal for the Philosophy of Science* 45:1–35.

Franklin, James. 1987. Non-deductive logic in mathematics. *The British Journal for the Philosophy of Science* 38:1–18.

Franklin, James. 2001. *The science of conjecture: Evidence and probability before Pascal*. Baltimore: The Johns Hopkins University Press.

Friedman, Nir, and Joseph Y. Halpern. 1995. Plausibility measures: A user's guide. In *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence* (*UAI '95*), 175–184.

Gärdenfors, Peter, and Nils-Eric Sahlin. 1982. Unreliable probabilities, risk taking, and decision making. *Synthese* 53:361–386.

Gavroglu, Kostas. 1989. Simplicity and observability: When are particles elementary? *Synthese* 79:543–557.

Gilboa, Itzhak. 2009. *Theory of decision under uncertainty*. Cambridge: Cambridge University Press.

Good, I. J. 1962. Subjective probability as the measure of a non-measurable set. In *Logic, methodology, and the philosophy of science,* ed. Patrick Suppes, Ernest Nagel, and Alfred Tarski, 319–329. Stanford: Stanford University Press.

Good, I. J. 1965. *The estimation of probabilities: An essay on modern Bayesian methods*. Cambridge: The MIT Press.

Gowans, Christopher W., ed. 1987. *Moral dilemmas*. New York: Oxford University Press.

Greenspan, Patricia. 1983. Moral dilemmas and guilt. *Philosophical Studies* 43:117–125.

Halpern, Joseph Y. 2003. *Reasoning about uncertainty*. Cambridge: The MIT Press.

Hammond, Peter J. 1988. Consequentialist foundations for expected utility theory. *Theory and Decision* 25:25–78.

Hempel, Carl G. 1960. Inductive inconsistencies. *Synthese* 12:439–469. In *Aspects of Scientific Explanation,* 53–79. New York: The Free Press and London: Collier Macmillan, 1965.
Hintikka, Jaakko. 1970a. On semantic information. In *Information and inference,* ed. Jaakko Hintikka and Patrick Suppes, 3–27. Dordrecht: D. Reidel.
Hintikka, Jaakko. 1970b. Surface information and depth information. In *Information and inference,* ed. Jaakko Hintikka and Patrick Suppes, 263–297. Dordrecht: D. Reidel.
Hintikka, Jaakko, 1983. *The game of language: Studies in game-theoretical semantics and its applications*. Dordrecht: D. Reidel.
Hintikka, Jaakko, and Juhani Pietarinen. 1966. Semantic information and inductive logic. In *Aspects of inductive logic,* ed. Jaakko Hintikka and Patrick Suppes, 96–112. Amsterdam: North-Holland.
Hirshleifer, J. 1965. Investment decision under uncertainty—choice theoretic approaches. *Quarterly Journal of Economics* 79:509–536.
Huber, Franz. 2008. Assessing theories, Bayes style. *Synthese* 161:89–118.
Huemer, Michael. 2013. Transitivity, comparative value, and the methods of ethics. *Ethics* 123:318–345.
Hughes, R. I. G. 1980. Rationality and intransitive preferences. *Analysis* 40:132–134.
Irvine, William B. 2006. *On desire: Why we want what we want*. Oxford: Oxford University Press.
James, William. 1897. *The will to believe, and other essays in popular philosophy*. New York: Longmans, Green & Co. (Cambridge and London: Harvard University Press, 1979).
Jeffrey, Richard C. 1983. *The logic of decision*. 2nd ed. Chicago: University of Chicago Press.
Jeffreys, Harold. 1931. *Scientific inference*. Cambridge: Cambridge University Press.
Jeffreys, Harold. 1961. *Theory of probability*. 3rd ed. Oxford: Clarendon Press.
Jensen, Niels Erik. 1967. An introduction to Bernoullian utility theory: I. Utility functions. *Swedish Journal of Economics* 69:163–183.
Kaplan, Mark. 1981. A Bayesian theory of rational acceptance. *The Journal of Philosophy* 78:305–330.
Kaplan, Mark. 1996. *Decision theory as philosophy*. Cambridge: Cambridge University Press.
Keynes, John Maynard. 1921. *A treatise on probability*. London: Macmillan. (Mineola: Dover, 2004).
Kieseppä, I. A. 1997. Akaike information criterion, curve-fitting, and the philosophical problem of simplicity. *The British Journal for the Philosophy of Science* 48:21–48.
Klir, George J. 2006. *Uncertainty and information: Foundations of generalized information theory*. Hoboken: John Wiley & Sons.
Knight, Frank H. 1921. *Risk, uncertainty and profit*. Boston: Hart, Schaffner & Marx. (Chicago and London: University of Chicago Press, 1971).
Kuhn, Thomas S. 1970a. Reflections on my critics. In *Criticism and the growth of knowledge,* ed. Imre Lakatos and Alan Musgrave, 231–278. Cambridge: Cambridge University Press.
Kuhn, Thomas S. 1970b. *The structure of scientific revolutions*. 2nd ed. Chicago: University of Chicago Press.
Kuhn, Thomas S. 1977. Objectivity, value judgment, and theory choice. In *The essential tension,* 320–339. Chicago: University of Chicago Press.
Kuipers, Theo A. F. 2000. *From instrumentalism to constructive realism*. Dordrecht: Kluwer Academic Publishers.
Kyburg, Henry E., Jr. 1961. *Probability and the logic of rational belief*. Middletown: Wesleyan University Press.
Kyburg, Henry E., Jr. 1979. Tyche and Athena. *Synthese* 40:415–438.
Lakatos, Imre, and Paul Feyerabend. 1999. *For and against method,* ed. Matteo Motterlini. Chicago: University of Chicago Press.
Larmore, Charles. 2008. *The autonomy of morality*. Cambridge: Cambridge University Press.
Laudan, Larry. 1984. *Science and values: The aims of science and their role in scientific debate*. Berkeley: University of California Press.
Levi, Isaac. 1967. *Gambling with truth: An essay on induction and the aims of science*. New York: Alfred A. Knopf.

Levi, Isaac. 1974. On indeterminate probabilities. *The Journal of Philosophy* 71:391–418.
Levi, Isaac. 1984. *Decisions and revisions*. Cambridge: Cambridge University Press.
Levi, Isaac. 1986. *Hard choices: Decision making under unresolved conflict*. Cambridge: Cambridge University Press.
Lewis, C. I. 1946. *An analysis of knowledge and valuation*. La Salle: Open Court.
Lockhart, Ted. 2000. *Moral uncertainty and its consequences*. New York: Oxford University Press.
Lycan, William G. 1998. Theoretical (epistemic) virtues. In *Routledge encyclopedia of philosophy,* ed. E. Craig. London: Routledge. http://www.rep.routledge.com/article/P050. Accessed 30 April 2014.
Maher, Patrick. 1993. *Betting on theories*. Cambridge: Cambridge University Press.
Makin, Stephen C. 2012. Action individuation and deontic cycling. *Ethics* 123:129–136.
Morgenstern, Oskar. 1979. Some reflections on utility. In *Expected utility hypotheses and the Allais paradox: Contemporary discussions of decisions under uncertainty with Allais' rejoinder,* ed. Maurice Allais and Ole Hagen, 175–183. Dordrecht: D. Reidel.
Morton, Adam. 1991. *Disasters and dilemmas*. Oxford: Basil Blackwell.
Morton, Adam, and Antti Karjalainen. 2003. Contrastive knowledge. *Philosophical Explorations* 6:74–89.
Narveson, Jan. 2004. Maxificing: Life on a budget; or, if you would maximize, then satisfice! In *Satisficing and maximizing: Moral theorists on practical reason,* ed. Michael Byron, 59–70. Cambridge: Cambridge University Press.
Ok, Efe A. 2002. Utility representation of an incomplete preference relation. *Journal of Economic Theory* 104:429–449.
Ok, Efe A., Juan Dubra, and Fabio Maccheroni. 2004. Expected utility theory without the completeness axiom. *Journal of Economic Theory* 115:118–133.
Peterson, Martin. 2009. *An introduction to decision theory*. Cambridge: Cambridge University Press.
Pigozzi, Gabriella. 2009. Interview with John Woods. *The Reasoner* 3 (3): 1–4.
Pollock, John L. 2006. *Thinking about acting: Logical foundations for rational decision making*. Oxford: Oxford University Press.
Popper, Karl. 1959. *The logic of scientific discovery*. London: Hutchinson.
Popper, Karl. 1969. *Conjectures and refutations*. 3rd rev. ed. London: Routledge & Kegan Paul.
Popper, Karl. 1974. Replies to my critics. In *The philosophy of Karl Popper,* ed. Paul A. Schilpp, 959–1197. La Salle: Open Court.
Popper, Karl. 1983. *Realism and the aim of science*. London: Hutchinson.
Quine, W. V., and J. S. Ullian. 1978. *The web of belief*. 2nd ed. New York: Random House.
Resnik, Michael D. 1987. *Choices: An introduction to decision theory*. Minneapolis: University of Minnesota Press.
Richmond, Samuel A. 1996. A simplification of the theory of simplicity. *Synthese* 107:373–393.
Russell, Bertrand. 1919. *Introduction to mathematical philosophy*. London: George Allen & Unwin.
Salmon, Wesley. 1981. Rational prediction. *The British Journal for the Philosophy of Science* 32:115–125.
Salmon, Wesley. 1990. The appraisal of theories: Kuhn meets Bayes. *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association* 2:325–332.
Samuelson, Paul A. 1952. Probability, utility, and the independence axiom. *Econometrica* 20:670–678.
Savage, Leonard J. 1971. Elicitation of personal probabilities and expectations. *Journal of the American Statistical Association* 66:783–801.
Savage, Leonard J. 1972. *The foundations of statistics*. 2nd rev. ed. New York: Dover.
Schaffer, Jonathan. 2004. From contextualism to contrastivism. *Philosophical Studies* 119:73–103.
Schmidtz, David. 2004. Satisficing as a humanly rational strategy. In *Satisficing and maximizing: Moral theorists on practical reason,* ed. Michael Byron, 30–58. Cambridge: Cambridge University Press.

Sextus Empiricus. circa 200. *Adversus mathematicos*. English edition: Sextus Empiricus. 2005. *Against the logicians* (trans. and ed: Betts, Richard). Cambridge: Cambridge University Press.
Shannon, Claude E. 1948. A mathematical theory of communication. *Bell System Technical Journal* 27:379–423. In *The mathematical theory of communication,* ed. Claude E. Shannon and Warren Weaver, 29–115. Urbana: University of Illinois Press, 1949.
Simon, Herbert A. 1982. *Models of bounded rationality*. Cambridge: The MIT Press.
Slote, Michael. 1989. *Beyond optimizing: A study of rational choice*. Cambridge: Harvard University Press.
Sober, Elliott. 1999. Testability. *Proceedings and Addresses of the American Philosophical* Association 73:47–76.
Sober, Elliott. 2002. Instrumentalism, parsimony, and the Akaike framework. *Philosophy of* Science 69 (Suppl 3): S112–S123.
Styron, William. 1979. *Sophie's choice*. New York: Bantam.
Temkin, Larry. 2012. *Rethinking the good: Moral ideals and the nature of practical reasoning*. Oxford: Oxford University Press.
Van Fraassen, Bas. 1980. *The scientific image*. Oxford: Clarendon Press.
Von Neumann, John, and Oskar Morgenstern. 1953. *Theory of games and economic behavior*. 3rd ed. Princeton: Princeton University Press.
Weirich, Paul. 2004. *Realistic decision theory: Rules for nonideal agents in nonideal circumstances*. Oxford: Oxford University Press.
Welch, John R. 2011. Decision theory and cognitive choice. *European Journal for Philosophy of Science* 1:147–172.
Welch, John R. 2012. Real-life decisions and decision theory. In *Handbook of risk theory,* ed. Sabine Roeser, Rafaela Hillerbrand, Per Sandin, and Martin Peterson, 545–573. Dordrecht: Springer.
Welch, John R. 2013. New tools for theory choice and theory diagnosis. *Studies in History and Philosophy of Science* 44:318–329.
Willenken, Tim. 2012. Deontic cycling and the structure of commonsense morality. *Ethics* 122:545–551.
Zimmerman, Michael J. 1996. *The concept of moral obligation*. Cambridge: Cambridge University Press.

Chapter 4
Working with Moral Means

Abstract Chapter 4 treats the instrumental stratum of moral discourse. Since this stratum is composed of substrata that can be delimited in different ways, the chapter works up through progressively larger discursive chunks: individual sentences, inferences, and theories. The justifiability of individual sentences, whether instrumental descriptions such as 'That shows the value of courage' or instrumental directives such as 'Be honest so that people will respect you', is an initial theme. The chapter emphasizes the role of observation and inductive inference in evaluating such sentences. Inference is then treated by investigating the possibility of justifying practical inferences (including so-called practical syllogisms). The chapter maintains that practical inferences can be evaluated by appeal to the standard of inductive cogency. Finally, how to choose a moral theory on instrumental grounds is illustrated by an extended analysis of Kantian, Benthamite, and Frankenian theory as applied to Sophie's choice. The analysis leans heavily on the comparative decision theory launched in Chap. 3.

4.1 Moral Instrumentality

Every action is aimed at a good. Chapters 4 and 5 rely on this Aristotelian thesis (*Nicomachean Ethics* 1094a1–3; *Politics* 1252a1–4), for it opens up two broad avenues for evaluating actions. We can evaluate the good or evaluate the aim. That is, we might endorse an action's good or reject it, alleging that it is not a good at all or that another good should have been chosen instead. Alternatively, we might praise an action because it leads effectively to its good or blame an action because it does not; it is well or badly aimed. The first kind of evaluation is teleological; the second, instrumental.

The present chapter is devoted to instrumentality. It opens by broaching two preliminary matters that impinge on instrumentality of any sort. The first is a consequence of the fact that instruments and human life are everywhere entwined: the sheer variety of instrumentality. Instruments come in different forms, and so do our ways of talking about them. Section 4.2 sets out to identify some main kinds of each. The second preliminary is the customary understanding of instruments as means. Though this conceptual linkage is nearly inevitable, it has met with varied forms of resistance. Section 4.3 attempts to scatter this resistance.

J. R. Welch, *Moral Strata*, Theory and Decision Library A:, 49,
DOI 10.1007/978-3-319-08013-0_4,

However, the bulk of the chapter treats instrumentality of a specific sort: instrumental moral discourse. The instrumental stratum of moral discourse is composed of substrata that can be defined in different ways. The chapter works up through progressively larger discursive chunks: individual sentences, inferences, and finally theories. Individual sentences are treated in Sect. 4.4 by exploring the justification of instrumental descriptions and instrumental directives. Practical inference is the focus of Sect. 4.5, which probes the so-called practical syllogism, applies the standard of inductive cogency introduced in Sect. 2.4.1, and criticizes the Kantian view of moral reasoning. Theory receives extended treatment in Sect. 4.6, where the comparative decision theory launched in Chap. 3 without reference to morality is applied to the problem of moral theory choice.

The reader will notice that even though the chapter's center of gravity is instrumental moral discourse, the discussion advances at each step of the way—sentence, inference, theory—by going beyond the bounds of morality. I have had to take my cue from Wittgenstein: "The motto here is always: Take a *wider* look around" (1978, p. 127).

4.2 Distinguishing Means

Instrumentality is hardly monolithic. Means come in several forms, and there are various ways of talking about them. I will begin by rehearsing some distinctions among means before turning to discourse about them.

Some means are actions; others are things. Instrumental actions appear to be more fundamental than instrumental things, for in order that a wine cask, say, can count as a means, a vintner has to pour wine into it. As von Wright points out, "Instruments [understood as things] would not be means to ends, unless they were *used*, *i.e.* unless there was human action aiming at certain goals. Because of this we can say that 'means' in the sense of action is *primary* to 'means' in the sense of instrument" (1963, p. 163).

The class of instrumental actions subdivides further. Let us recall von Wright's distinction between productive and necessary means (1963, p. 165). Turning on the air conditioner, for example, is a productive means; it produces a cool room. By contrast, closing the windows may be a necessary means in hot weather, for it is a causal requirement for having a cool room. These two classes of means overlap, for some means can be both productive and necessary. This occurs whenever there is only one means to an end (1963, p. 165).

We can make this vocabulary more transparent by noting that a productive means produces a consequence; it is therefore causally sufficient for that consequence, though it may not be necessary. We note further that means that are both necessary and productive parallel causal conditions that are both necessary and sufficient. Hence the distinction between a means that is both necessary and productive and one that is merely necessary would be that the former is necessary and sufficient

while the latter is necessary but insufficient. That would give us three types of means within the class of instrumental actions: necessary (but insufficient), sufficient (but unnecessary), and necessary and sufficient.

Besides these distinctions among different sorts of means, there are different ways of talking about them. The major divide is between instrumental descriptions and instrumental directives. Instrumental descriptions aim to state the facts. Here are some familiar moral instances: virtue is necessary for happiness (Aristotle, *Nicomachean Ethics* 1098a17–19); the natural laws secure goods such as self-preservation and social peace (Aquinas 1265–1273, I–II, q. 94, a.2); a social contract improves chances for survival in a state of war (Hobbes 1651, Chap. xiii); the hedonistic calculus measures pleasure and pain (Bentham [1789] 1970, Chap. iv, pp. 38–41); the categorical imperative discloses duty (Kant 1785, Ak 5:403, 5:421). By contrast, instrumental directives are prescriptive. They include what von Wright called the three "aspects" of norms: commands, rules, and practical necessities (1963, p. 157). Commands are formulated in the imperative mood ('Formulate the maxim of your action', for instance). Rules are associated with verbs such as 'ought to', 'may', and 'must not' ('You ought to calculate the utilities of all options'). Practical necessities are expressed with verbs like 'must', 'need not', and 'cannot' ('You must be virtuous to be happy'). However, von Wright observes that ordinary language does not sharply distinguish the three types of sentence (1963, p. 158), and I will rest little weight on these secondary distinctions. The main thing is whether an instrumental sentence is a description or a directive.

4.3 Means, Ends, and Their Critics

This chapter and the next are structured by the distinction between means and ends. I take this distinction to be a central part of human rationality: Every action (as means) is aimed at a good (its end). But since this nexus of ideas has been attacked from various angles, it cannot simply be taken for granted. I will briefly address some of these challenges, therefore. I want to say something about three in particular.

The first objection is that what is an end from one point of view may be a means from another; hence the distinction between means and ends is relative. Dewey, who disliked dualisms in general and the dualism of means and ends in particular, claimed that the terms 'means' and 'end' are "two names for the same reality" ([1922] 1983, p. 28). The distinction, he thought, depends entirely on the speaker's point of view.

This objection does contain a kernel of truth: some ends are indeed also means, as Aristotle grants (*Nicomachean Ethics* 1096b18–20). The end of publishing a book may simultaneously be a means to securing tenure. But simultaneity of means and ends is not the whole truth. The action of taking bitter medicine is a means to the end of restored health. Here, as a matter of empirical fact, the means is distinct

from the end; the means is a present action while the end is a merely possible future state. In addition, we recognize ends that are not means: happiness, say, as Aristotle understood it. What we have, then, are three clearly distinguishable cases: means that are not ends, ends that are not means, and means that are simultaneously ends. There is no brief for relativism about means and ends here. We simply need to remember that the 'or' in 'means or end' is inclusive.

A second complaint is due to G. H. von Wright, who observes that the traditional division of goods into goods as means and goods as ends is inadequate and artificial (1963, pp. 11–12). This may be true if we are trying to account for the bewildering variety of uses of 'good' in English. But that is not my business here. The goods that figure in this chapter are good because they help us to achieve our ends, and the goods of the following chapter are these ends themselves. Whatever additional goods there may be, means and ends are paradigmatic goods. Moral means and ends can be investigated without making the dubious claim that all goods are means or ends.

Whereas the first two objections target the means-end distinction, a third objection accepts the distinction but refashions the bridge to instrumentality. According to this final objection, all instruments are means yet not all means are instruments. The reason: there are constitutive as well as instrumental means. Virtue, for example, might be seen as a constitutive means to happiness if it is viewed as part of happiness but not as an instrumental means to happiness.

The concept of a constitutive means is sometimes traced back to Aristotle's distinction between making and doing (Irwin 2003, pp. 61–62): "For while making has an end other than itself, action cannot; for good action itself is its end" (*Nicomachean Ethics* 1140b6–7). That is, a vintner may select instrumental means to the end of the finished wine, whereas someone might perform an ethical action for the sake of performing that very action.

I submit, however, that this analysis of ethical action fails upon further reflection. Take a glass blower who makes only for the sake of the finished glass and another who makes solely for the sake of making, for sheer love of the craft. We can distinguish the two cases in much the same way that Aristotle contrasts making and doing: one makes for an end other than making, the other makes for the sake of making. But does the second case count as a constitutive, non-instrumental means? No, it does not. In the first case, the maker makes for the sake of an external object. In the second case, the maker makes for the sake of an internal state: the experience of making. Hence there is a notable difference between the two cases, but it is a difference of end, not a difference of means. Both means are instrumental to their respective ends.

Observe now two additional pairs of cases. The first consists of a dancer who dances to win a contest and a dancer who dances for the joy of the dance. Dancing is instrumental in both cases, but what it is instrumental for is evidently different: an external state in one case and an internal state in the other. The second pair is formed by an altruist who feeds the hungry only so that the hungry get fed and an egoist who feeds the hungry solely for the experience of doing so—she feels at

peace with her conscience, say. The first action is for the sake of an external state; the second action, for an internal one. Once again, then, we have an important difference, but the difference lies in the ends, not the means. In both cases, the means are instrumental.

Reflecting on cases like these strongly suggests the following moral: the temptation to posit non-instrumental, constitutive means should be resisted by recognizing differences among ends. Some ends are things, while others are states of affairs; states of affairs are internal in some cases but external in others. These distinctions are treated somewhat more fully in Sect. 5.2.

4.4 Instrumental Moral Sentences

Talk about means is a topic for development in several directions throughout the rest of this chapter. As noted in Sect. 4.2, means that are actions are more fundamental than means that are things, and I will therefore concentrate on instrumental talk about actions. From this point on, however, the focus is instrumental language that is specifically moral. I will amplify the instrumental moral theme by working up through progressively larger linguistic units: sentences, inferences, and finally theories. Each of these units results from an organization of substrata within the instrumental stratum. How to justify moral discourse within each substratum is the main theme of our discussion. I will hew throughout to the Deweyan line that "nothing can justify or condemn means except ends, results" ([1922] 1983, p. 157). The reader should be forewarned, however, that doing justice to instrumental moral discourse will repeatedly require casting our net widely enough to capture any kind of instrumental discourse.

To begin our treatment of individual sentences, consider an instrumental moral description.[1] An observer who witnesses a pedestrian save a child from being run over by a car exclaims "That shows the value of courage." Despite its apparent simplicity, this is a rather complex sentence. It expresses at least three claims on the part of the observer: that the action was courageous; that it prevented the child's injury or death; and that the action somehow illustrates why we encourage courage. Hence the sentence has truth conditions of a complex sort. It is true if, and only if, all three subsidiary claims are true.

Little need be said about the first two claims, I think. Whether the action is courageous is a matter of classification, and its problems and prospects are conditioned by the treatment of moral classification in Chap. 2. In addition, whether the action is sufficient to prevent the injury or death of the child is a straightforward empirical question. But the third claim is more problematic. While not entirely explicit, the idea is roughly that other courageous actions produce comparably favorable results. Could we tell whether this claim is true or false? The answer is yes, at least

[1] The balance of this section is based loosely on Welch (1994, pp. 285–287).

in principle. We could test it in the same way we test statistical generalizations in the social sciences—through careful observation and inductive inference. Claims like the foregoing are empirically no different than Gresham's law, for example, in economics. What *is* different is that, while prodigious amounts of energy and ingenuity have been expended on empirical testing in the social sciences, to my knowledge empirical tests of statistical generalizations are rarely even attempted in ethics. And that, I am afraid, is a serious indictment of ethics.

At least one twentieth-century classic coincides with the viewpoint of the previous paragraph. Carl Hempel's *Aspects of Scientific Explanation* treats instrumental descriptions of value along the same general lines (note that Hempel uses the term 'judgment' instead of 'description'):

> [*I*] *f* our children are to become happy, emotionally secure, creative individuals rather than guilt-ridden and troubled souls *then* it is better to raise them in a permissive than in a restrictive fashion. A statement like this represents a *relative, or instrumental, judgment of value*. Generally, a relative judgment of value states that a certain kind of action, *M*, is good (or that it is better than a given alternative M_1) *if* a specified goal *G* is to be attained; or more accurately, that *M* is good, or appropriate, for the attainment of goal *G*. But to say this is tantamount to asserting either that, in the circumstances at hand, course of action *M* will definitely (or probably) lead to the attainment of *G*, or that failure to embark on course of action *M* will definitely (or probably) lead to the nonattainment of *G*. In other words, the instrumental value judgment asserts either that *M* is a (definitely or probably) sufficient means for attaining the end or goal *G*, or that it is a (definitely or probably) necessary means for attaining it. Thus, a relative, or instrumental, judgment of value can be reformulated as a statement which expresses a universal or a probabilistic kind of means-end relationship, and which contains no terms of moral discourse—such as 'good', 'better', 'ought to'—at all. And a statement of this kind surely is an empirical assertion capable of scientific test. (1965, pp. 84–85)

The reformulation of instrumental descriptions of value that Hempel contemplates in this passage is free of "thin" moral terms such as 'good' and 'better'. But there is no reason why these reformulations could not contain "thick" moral terms for both means ('That shows the value of courage') and ends ('The new law encourages honest reporting'). This follows from the discussion of classification in Chap. 2. Regardless of whether our classificatory predicates are moral or not, classification is ultimately carried out by analogy. A given classification can be defended via an argument by analogy, and that argument is subject to critical assessment relative to the standard of inductive cogency outlined in Sect. 2.4.1 and applied in Sect. 2.5.2.

If the foregoing manages to sketch some guidelines for instrumental moral descriptions, we can shift our focus to another type of instrumental moral sentence. A parent enjoins a child to be honest, and the child responds by demanding to know why. The parent explains that honesty is necessary for gaining the respect of others. The original injunction should thus be understood as 'Be honest so that people will respect you'. This is an instrumental moral directive.

'Is' and 'ought' are related here in a most suggestive way. We know, of course, that 'ought' in the form of 'You ought to be honest' cannot be derived from 'is' in the form of 'Honesty is necessary for gaining respect'. But 'ought' can be *justified*

by ‘is’. Why ought the child to be honest? Because, according to the parent, honesty is necessary for gaining people’s respect. This justificatory relationship is instantiated over and over again in ethical and nonethical contexts alike. Take the nonethical case of a master potter who instructs a new assistant to fire a freshly enameled pot. The assistant asks why, and the potter answers by stating that firing is necessary to fix the colors. The original instruction was thus an instrumental directive: to fix the colors, fire the pot. The relationship between the ‘ought’ of ‘You ought to fire the pot’ and the ‘is’ of ‘Firing is necessary to fix the colors’ is that the latter warrants the former.

Although an instrumental directive might be justified by another that is more general, general instrumental directives must be justified by instrumental descriptions. That the arrow of justification runs in this direction rather than the opposite is fortunate, for we are in relatively good shape with instrumental descriptions, as I argued above. Hence if instrumental descriptions ground instrumental directives, we are in relatively good shape with the directives as well. To devise reliable empirical indicators for concepts like respect would take some doing, admittedly. But the methodological ingenuity that scholars have shown in managing concepts as slippery as alienation and happiness is one indicator of what can be done.[2]

4.5 Practical Inference

The reader may have already noticed that the relation between instrumental descriptions and instrumental directives is analogous to that between the premises and conclusions of arguments. Just as a challenge to an argument’s conclusion can be met by adducing its premises, a challenge to an instrumental directive can be met by invoking an instrumental description. The analogy between the two patterns of justification can be spelled out further by relating the directive-description linkage exemplified two paragraphs above to the so-called practical syllogism.

4.5.1 The Practical Syllogism

The study of practical syllogisms appears to have originated with Aristotle, though he never actually uses the term.[3] In fact, the term ‘practical syllogism’ is a misnomer (Kenny 1979, pp. 111–114). Aristotle’s examples are not constrained by the syllogistic patterns of the *Prior Analytics*, and they typically belong to Aristotelian productive sciences like medicine rather than practical sciences such as ethics.

[2] Seminal studies on alienation include Seeman (1959), Kohn (1974), and Schacht (1994). Recent studies of happiness include Nettle (2005), Layard (2005), and Angner (2013).

[3] He does come close at *Nicomachean Ethics* 1144a30–33, however.

Hence another term would be desirable. Since our contemporary usage of 'practical' has the broader sense of being oriented toward action and this is the sense I would like to develop, I will use the term 'practical inference'.

The following passage from *Movement of Animals* contains six practical inferences, the last three of which appear to be continuous stages of an extended deliberation. I have numbered the examples within the quotation in order to facilitate reference to them below.

> [1] [F]or example, whenever one thinks that every man ought to walk, and that one is a man oneself, straightaway one walks; [2] or that, in this case, no man should walk, one is a man: straightaway one remains at rest. And one so acts in the two cases provided that there is nothing to compel or to prevent. [3] Again, I ought to create a good, a house is a good: straightaway he makes a house. [4] I need a covering, a coat is a covering: I need a coat. [5] What I need I ought to make, I need a coat: I make a coat. And the conclusion 'I must make a coat' is an action. And the action goes back to a starting-point. [6] If there is to be a coat, there must first be this, and if this then this—and straightaway he does this. Now that the action is the conclusion is clear. But the premisses of action are of two kinds, of the good and of the possible.[4]

If we examine the conclusions of these practical inferences using the terminology of Sect. 4.2, we get some interesting results. Three of the conclusions (inferences 1, 2, and 5) point to necessary means, for they are causally necessary but insufficient for their ends. Two other conclusions (inferences 3 and 4) indicate sufficient means; they are sufficient but unnecessary for their ends. Finally, one conclusion (inference 6) appears to gesture toward a means that is both necessary and sufficient, though the example is so sketchy it is hard to be sure. But if its first premise is understood as 'I ought to make a coat', its second and third premises identify things that are necessary for the coat, and the conclusion would then seem to require securing these necessities and using them in an appropriate way. That would be necessary and sufficient for the coat.

Before we can relate instrumental descriptions and directives to practical inferences, we need to pause briefly over two controversial points in the just-quoted passage from *Movement of Animals*. The first is the concluding remark that the premises of practical inference are of two kinds: "of the good and of the possible." In a classic article, D. J. Allan suggested that a premise "of the possible" concerns means to a desirable end, and a premise "of the good" states a general rule to be followed (1955, pp. 330–331). Corresponding to these two types of premises, he argued, are two types of practical inference (1955, pp. 336–337). For example, inference 1 above uses premises of the good such as 'Every man ought to walk', while inference 6 uses premises of the possible like 'If there is to be a coat, there must first be this'. However, other writers counter that both sorts of premises are found together in the same inference. Anthony Kenny, for example, points out that "all the genuinely practical premises in the examples in the *De Motu* [*Movement of Animals*] are premises 'of the good'…. What the reference to possibility means is

[4] *Movement of Animals* 701a13–25. See also 701a26–33; *Nicomachean Ethics* 1141b18–21, 1147a5–8, 29–31.

this: practical reasoning can only come to a successful end when it reaches some action which is in the agent's power, some state of affairs which he can bring about" (1979, pp. 119–120).

Allan is right to single out inference 6, which is peculiar in two respects. It is the only one that refers to means that are things, and it is the only one that does not obviously begin with a premise of the good. But the other inferences also concern means, though they deal with means that are actions rather than means that are things. And since inferences 4–6 fit together roughly in a practical analogue of a Goclenian sorites, where the conclusion of each prosyllogism forms the major premise of its episyllogism, the conclusion of inference 5 ('I must make a coat') would be the initial premise of inference 6. Inference 6 would therefore begin with a premise of the good like all the rest. I conclude that Kenny is correct that all the examples from *Movement of Animals* contain premises of the good. They are not segregated from premises of the possible, whatever Aristotle meant by that term. Thus the distinction between these two types of premises is no place to begin a typology of practical inference. More promising approaches emerge in the following paragraphs.

The second controversial point from the *Movement of Animals* passage is the claim that the conclusion of a practical inference is an action. This immediately creates a puzzle: How could premises that are sentences imply a conclusion that is an action? To address this puzzle, we need to consider the question of variety in practical inferences. We have already recognized that the conclusions of practical inferences can point to means that are sufficient, necessary, or both. But we can find further variety of form and function.

Aristotle's own examples show variety of form. Some premises use 'ought' while others use 'need' or 'must', and the number of premises varies. Other writers have identified other forms. Robert Audi, for instance, identifies four kinds of schemata instantiated in practical reasoning. Where an agent contemplates an action to achieve a specific goal, *necessary condition schemata* take the action to be necessary for achieving the goal; *sufficient condition schemata* take it to be sufficient for the goal; *sufficient reason schemata* take it to be sufficiently reasonable for the goal—probable, for instance, even if not necessary or sufficient; and *rule schemata* take it to be required by a particular rule (1989, pp. 86–87; cf. 2004, pp. 128–130).

Necessary condition schemata are of particular interest for my purposes. They have been studied by von Wright, who specifies various primary forms of practical inference and secondary forms derived from the primary ones (1983, pp. 1–17). The most prominent of his primary forms is the necessary condition schema (1983, p. 2):

> One wants to attain *x*.
> Unless *y* is done, *x* will not be attained.
> Therefore *y* must be done.

Instantiating this form might give us:

> One wants to make the hut habitable.
> Unless the hut is heated, it will not become habitable.
> Therefore the hut must be heated.

According to von Wright, this impersonal inference branches into personalized inferences in the first person:

> I want to make the hut habitable.
> Unless I heat the hut, it will not become habitable.
> Therefore I must heat the hut.

and the third person:

> Alvin wants to make the hut habitable.
> Unless Alvin heats the hut, it will not become habitable.
> Therefore Alvin must heat the hut.

This distinction between first-person and third-person practical inferences may be of some help initially, but relying on it to the extent that von Wright does is misguided, I believe. The resulting typology is patently incomplete, first of all, for there are second-person practical inferences as well as first- and third-person ones. For example,

> You want to make the hut habitable.
> Unless you heat the hut, it will not become habitable.
> Therefore you must heat the hut.[5]

Secondly, von Wright's own use of the distinction reveals that the grammatical category of person is not the essential consideration. He classifies

> I wanted to reach the train in time.
> Unless I ran, I would not have reached the train in time.
> Therefore I had to run.

as a third-person inference because "The agent is here speaking *about* himself, as it were viewing himself from the outside" (1963, p. 167). Finally, and most importantly, von Wright's development of this distinction results in a theory of practical inference that is far more complicated than necessary. He maintains that the premises and conclusions of first-person inferences are radically different from those of third-person inferences. In the third-person case,

> the premisses are the *propositions* that a certain person pursues a certain end of action and that a certain thing is a necessary means to this end. The conclusion is a third proposition, namely that the person will fail to reach some end of his action unless he does this thing. In the case of the inference in the first person the correct answer seems to be this: the premisses are a person's *want* and his *state of knowing or believing* a certain condition to be necessary for the fulfilment of that want. The conclusion is an *act*, something that this person does. Wants, states of knowing or believing, and acts are not only mutually rather different from each other. They are all of them entities of a radically different sort from propositions. It is of the essence of propositions that they are expressed by sentences…. Wants, states of knowing or believing, and acts have no analogous essential connection with *language*. Therefore the relation to language of a practical inference in the first person is in principle different from the relation to language of a practical inference in the third person. (1983, pp. 8–9)

[5] This is an instantiation of one of von Wright's forms (1963, p. 161).

But this just chops things up, splitting off non-propositional first-person inferences from third-person propositional inferences. The constituents of first-person inferences are incongruent with each other, and the complexes they form are incongruent with third-person inferences. The resulting theory is a farrago.

Far better, I believe, to aim for an ontologically unified theory of practical inference. To find one, we need to focus on distinctions of function rather than form. Elsewhere I have urged that some practical inferences are deliberative while others are reconstructive (Welch 1991, pp. 77–78). The deliberative inference identifies means to an unrealized end. The reconstructive variety explains why an agent—a person or even an animal—acted as it did. This difference of function implies a further difference, one that can serve as an independent criterion for the same distinction. This second difference is temporal. If we note the relationship between the assertion of the premises of a practical inference and the action specified by its conclusion, we note two distinct possibilities. In deliberative inferences, we have first the assertion of the premises, then the action of the conclusion. In reconstructive inferences, however, the order is reversed: first the action of the conclusion, then the assertion of the premises. Since the inference reconstructs a past or ongoing action, the action specified by the conclusion must occur before it can be explained via the premises.

Aristotle keeps both deliberative and reconstructive functions of practical inference in view. Inferences 1–6 above exemplify elementary deliberations. Yet *Movement of Animals* treats the motion of unreasoning as well as reasoning animals. Here, for instance, Aristotle makes the same practical inference do for both:

> [F]or example if walking is good for man, one does not dwell upon the proposition 'I am a man'. And so what we do without reflection, we do quickly. For when a man is actually using perception or imagination or thought in relation to that for the sake of which, what he desires he does at once. For the actualizing of desire is a substitute for inquiry or thinking. [7] I want to drink, says appetite; this is drink, says sense or imagination or thought: straightaway I drink. In this way living creatures [zōa] are impelled to move and to act, and desire is the last cause of movement, and desire arises through perception or through imagination and thought. (*Movement of Animals* 701a26–35)

In this passage, inference 7 explains actions of human beings who are not deliberating at the moment and of animals that, according to Aristotle, cannot deliberate at all. The inference attempts to explain their actions. It is reconstructive.

We are now in a position to address the puzzle about the conclusion of a practical inference: How could premises that are sentences imply a conclusion that is an action? The answer is that they cannot, and the deliberative-reconstructive distinction shows us why. For the reconstructive case, take inference 7 from the previous paragraph as an example. Since the action of drinking has already taken place, it cannot possibly reoccur later on as the conclusion of its own explanation. The reconstructive inference must have a sentence such as 'Therefore I drink' as its conclusion. Then what about the alternative, the deliberative inference? Aristotle is wrong, I believe, though not far wrong, to say that "the action is the conclusion" of the inference (*Movement of Animals* 701a23). He is not far wrong because of the near inseparability of a conclusion such as 'I ought to walk' and my walking. Aristotle is almost right when he says that "that which is last in the process of thinking is the

beginning of the action" (*On the Soul* 433a17). For that which is last in the process of thinking, the conclusion, is (part of) the action's cause.[6] Identifying the conclusion and the action, then, is a mistake, the mistake of identifying cause and effect. It confuses the logical relation between premises and conclusion with the causal relation between conclusion and action. The conclusion of a deliberative inference must be a sentence as well.

The point of taking the scenic route through the practical syllogism is that it offers insight into what might otherwise remain mysterious: why instrumental descriptions can justify instrumental directives. Let us recall the instrumental justifications of Sect. 4.4. The instrumental description 'Honesty is necessary for gaining others' respect' backed up the ethical directive 'Be honest', and the instrumental description 'Firing is necessary to fix the colors' buttressed the nonethical directive 'Fire the pot'. Both justifications, we can now see, are enthymematic practical inferences. If we fill in the blanks using one of von Wright's primary forms as a guide, we get something like this:

> You want to gain the respect of others.
> Unless you are honest, you will not gain the respect of others.
> Therefore you must be honest.

> You want to fix the colors.
> Unless you fire the pot, you will not fix the colors.
> Therefore you must fire the pot.

The conclusions of both practical inferences are sentences that enjoin necessary means to their respective ends.

4.5.2 Assessing Practical Inference

Audi takes practical reasoning to be "a kind of means-end reasoning" characterized by a major premise representing a goal, a minor premise representing a belief that a certain means would help to achieve that goal, and a conclusion representing a practical judgment that one should enact this means (1989, p. 146, 99, 2004, p. 128). Yet some practical inferences are trustworthy; others are not. How might we separate the wheat from the chaff? As with better-known forms of inference, the assessment of practical inference proceeds by attending to conditions on the inference's content and form.

This section concentrates on the question of form. To manage the complex issues that intersect here, our inquiry will be distributed along the following lines: whether any practical inference could be valid (Sect. 4.5.2.1); whether the instances of necessary condition schemata in Sect. 4.5.1 are valid (Sect. 4.5.2.2); whether other forms of practical inference are valid (Sect. 4.5.2.3); and whether any invalid forms of practical inference are nonetheless reliable (Sect. 4.5.2.4).

[6] Cf. von Wright: "Aristotle would have been *quite* right, had he said that the practical syllogism *leads up to* action. It ends, not necessarily in *doing* something, but in *setting oneself to do* something" (1963, p. 169).

4.5.2.1 Whether Any Practical Inference Could Be Valid

Whether any practical inference could be valid is the most fundamental question of the lot. Since validity in the relevant sense is the impossibility that an inference's conclusion is false while its premises are true, the question of whether a practical inference is valid cannot even be raised unless all its component sentences have truth values. But can prescriptive sentences that say that someone must do something or that someone ought to do something be true or false?

They can, quite plausibly, provided we are willing to accept von Wrightian glosses of each (1983, p. 5, pp. 21–22). If we take 'She must heat the hut' and 'She has to heat the hut' to mean 'She will fail to attain one of her ends unless she heats the hut', these sentences do indeed bear truth values. 'Ought' sentences such as 'She ought to heat the hut', which have long been recognized to refer "to actions and to actions alone" (Prichard [1912] 1949, p. 4), can be interpreted in the same way. Von Wright remarks:

> [W]e could also say that the idea of 'ought' has two main sources. The one source is in the will of a commanding agent or norm-authority. The other is a double source in ends of human action and necessary connections between things.
> In themselves, the two sources are of a rather different nature. But they are related to one another through the notion of a foundation of a norm (as a manifestation of the will of a norm-giver). Norms are frequently, perhaps one could say: normally, given for the sake of some ends. For this reason it may happen that the 'ought', which flows from a commanding will, becomes supported by the 'ought' of a technical rule and will rest on this latter 'ought' as on its foundation. (1983, p. 74; cf. 96)

In short, the 'ought' of a "commanding will" is normally grounded by the 'ought' of a "technical rule" such as 'She will fail to attain one of her ends unless she heats the hut'. The grounding 'ought' is instrumental.[7]

The 'must' and 'ought' under consideration appear in conclusions of necessary-condition practical inferences, not in premises that urge a particular end.[8] If we adopt von Wright's interpretation of such conclusions, claims that someone must or ought to act in a certain way are true or false, depending on their fidelity to the causal relations between the specified means and ends. This holds for both the nonmoral and moral case. Just as the potter's 'must' or 'ought' may be tantamount

[7] Von Wright's interpretation of terms like 'must' and 'ought' complements well-known views of other thinkers. One is Hempel's line on instrumental judgments of value, which was noted already in Sect. 4.4: "a relative, or instrumental, judgment of value can be reformulated as a statement which expresses a universal or a probabilistic kind of means-end relationship, and which contains no terms of moral discourse—such as 'good', 'better', 'ought to'—at all. And a statement of this kind surely is an empirical assertion capable of scientific test" (1965, p. 85). Another close relative is Gilbert Harman's "good-reasons" analysis of 'ought': "In this view, then, to say that *P* ought to do *D* is to say that *P* has sufficient reasons to do *D* that are stronger than reasons he has to do something else. If what you mean is that *P* morally ought to do *D*, you mean that *P* has sufficient moral reasons to do *D* that are stronger than the reasons he has to do something else" (1978, p. 112). According to Harman, both moral and nonmoral 'ought' derive their force from agents' goals and ends (1978, pp. 113–117).

[8] The truth of such premises might be established by chaining practical inferences, as discussed in Sect. 4.5.3.

to 'One of your goals will not be accomplished unless you fire the pot', the moralist's 'must' or 'ought' may amount to 'One of your goals will not be realized unless you are honest'. These conclusions are true if, and only if, one of the agent's goals is not attained or the specified action is performed. Therefore, since the major and minor premises of necessary-condition practical inferences are true or false and the conclusions of these inferences are also true or false, necessary-condition practical inferences can be evaluated for validity.

The pattern of argument just deployed can be usefully inverted. I have been arguing that because the component sentences of practical inferences have truth values, they can be valid; but Walter Sinnott-Armstrong contends that because practical inferences can be valid, their component sentences have truth values (2006, pp. 20–23). That is, since moral instances of modus ponens, say, are valid and validity means the impossibility of true premises and a false conclusion, the components of moral instances of modus ponens have truth values.

4.5.2.2 Whether the Instances of Necessary Condition Schemata in Sect. 4.5.1 Are Valid

Some careful thinkers have denied that inferences like the hut examples in Sect. 4.5.1 are logically valid. Henry Richardson, for example, objects to the impersonal form of the inference as follows:

> For the pattern to be deductively valid, the 'unless' [in the second premise] would have to stand for a relation at least as strong as the logical 'if… then': 'if the hut is not heated, then it will not become habitable.' As von Wright himself points out, however, the temperature might rise by itself (due to global warming, we might imagine). Alternatively, its owner might drink enough vodka not to feel the cold, or else buy a down comforter. (1994, p. 39)

But suppose we render the inference's second premise as Richardson proposes and understand 'must' with von Wright. Then the impersonal version of the hut inference becomes:

> One wants to make the hut habitable.
> If the hut is not heated, then the hut will not be made habitable.
> Therefore one will fail to attain an end one holds unless the hut is heated.

The second premise is a sentence type variously instantiable as sentence tokens, and these tokens will be true or false depending on the circumstances in which they are uttered. In some cases, an expedient along the lines Richardson suggests may serve to heat the hut, thereby making the premise false. In other cases, however, the premise will be true. A hut whose inhabitants have to stumble around in down comforters or stoke themselves with vodka may not meet one's criteria of habitability, and global warming may not habilitate the hut before it crumbles into dust. But whether the second premise is true or false is really beside the point. The point is validity, and the question is whether the conclusion must be true *if* the premises are true. The answer is clearly affirmative. Hence the hut inference is valid.

Admittedly, though, the hut inference is not *formally* valid. But inferences whose validity is not formal are more complex than they appear, for they have a suppressed premise (Sinnott-Armstrong 2006, p. 138). If we spell out the obvious by adding 'To make the hut habitable is to attain an end one holds' to the previous paragraph's version of the hut inference, the inference becomes formally valid. Or, better still, we could obtain the same result by simply replacing the first premise with 'To make the hut habitable is to attain an end one holds'. Analogous maneuvers could be carried out on the respect, pottery, and train inferences of Sect. 4.5.1.[9] If their conclusions are interpreted along von Wrightian lines, all are valid as they stand, and all can be made formally valid by addition or substitution of their suppressed premise.

The validity of necessary condition schemata has been attacked from another quarter, however. Georg Spielthenner argues that such schemata are generally invalid (2007, pp. 142–143). One of his counterexamples runs as follows: "Assume that I want to become the next heavyweight-boxing champion of the world. Does from this goal follow that I should start training? Clearly not, because even though training is a necessary condition for my aim, I will not become the next boxing champion anyway" (2007, p. 143). In other words, the premises that I want to become heavyweight champion and that training is necessary to become heavyweight champion could be true while the conclusion that I ought to start training could be false.

On a purely intuitive level, the conclusion 'I ought to start training' does not seem to be obviously false. After all, pursuit of an end that is unattainable, strictly speaking, may still be worthwhile provided the end can be approximated. More to the point, this conclusion is not false provided we interpret it along von Wrightian lines. That is, if 'I ought to start training' means 'I will fail to attain one of my ends unless I start training', the conclusion is logically equivalent to the conditional 'If I don't start training, then I will fail to attain one of my ends'. But this is equivalent in turn to the disjunction 'I do start training or I will fail to attain one of my ends'. So if it turns out that I will not become heavyweight champion under any circumstances, the second disjunct will always be true, thereby making the conclusion always true. Hence this is not a counterexample that shows the invalidity of a necessary condition schema.

4.5.2.3 Whether Other Forms of Practical Inference Are Valid

In spite of Spielthenner's criticism of necessary condition schemata, he contends that they can be salvaged by transforming them into necessary and sufficient condition schemata (2007, p. 147). That is, validity can be preserved by augmenting the second, instrumental premises with sufficient conditions. The first-person form of the hut inference would then be:

> I want to make the hut habitable.
> The hut will be habitable if and only if I heat it.
> Therefore I must heat the hut.

[9] Cf. Audi (1989, pp. 5–7, 13–14, 20–22, 30–31).

If we continue to interpret ‘must’ and ‘ought’ in the conclusions of such inferences along von Wrightian lines, the conclusion ‘I will fail to attain one of my ends unless I heat the hut’ is true provided the premises are true. This is then a valid inference.

In addition to affirming the validity of some necessary condition schemata, Audi takes certain rule schemata to be valid as well (Audi 1989, pp. 146–147, 2004, p. 128). Sinnott-Armstrong offers the following example of a rule schema in modus ponens (2006, p. 20):

> Lying is wrong.
> If lying is wrong, then paying your little brother to lie for you is wrong.
> Thus paying your little brother to lie for you is wrong.

Sinnott-Armstrong observes that this inference is valid, adding “All [moral] expressivists whom I know admit that [this] is a valid argument” (2006, p. 20 n. 4).

John Broome also defends the validity of some forms of practical inference. In the following example (2001, pp. 176–178), the major premise and conclusion express intentions:

> I am going to leave the next buoy to starboard.
> In order to leave the next buoy to starboard, I must tack.
> Thus I shall tack.

Once again, the premises cannot be true without the conclusion also being true.

Whereas Sect. 4.5.2.2 argued that some necessary condition schemata are valid, this section suggests expanding the catalog of valid practical inferences. So even though disagreement may persist over the validity of specific forms of practical inference, there is substantial agreement that some forms of practical inference are valid. Von Wright’s comments, steeped in the spirit of the late Wittgenstein, can serve as coda here:

> Shall we deny then that the [practical] syllogism is logically valid? This way out too has been suggested—but seems to me to be a mere evasion. We must, I think, accept that practical syllogisms are logically valid pieces of argumentation in their own right. Accepting them means in fact an enlargement of the province of logic. We cannot reduce the practical syllogisms to other patterns of valid inference. (1963, pp. 167–168)

4.5.2.4 Whether Any Invalid Forms of Practical Inference Are Reliable

Everyone recognizes that some common forms of practical inference are not deductively valid. The real target of Richardson’s critique of the hut examples, for example, is overemphasis on deductive logic in practical reasoning: “I have not meant to deny that deductive inference is ever useful in the course of practical reasoning. My point is the more modest one that practical reasoning is far from exhausted by deductive inference” (1994, p. 41). I concur.

Consider the following instance of a sufficient condition schema:

> I want to catch the bus.
> If I hurry, I will catch the bus.
> Therefore I ought to hurry.

Despite the inference's common-sense appeal, Spielthenner points out that it does not satisfy the validity condition (2007, p. 141). Since I could also catch the bus by walking leisurely at times, the conclusion that I ought to hurry would be false on those occasions. More generally,

> If I attain a good in any case or if an evil does not happen no matter what I do, it does not follow that I should do a sufficient condition for this good or should not do a sufficient condition for the evil. More specifically, assuming that the choice is based on a given set of alternatives {a, b, c}, then if more than one member of this set is sufficient for a good of the agent, it does not follow that he should fulfill one of these sufficient conditions, and if more than one member of this set is sufficient for an evil, it does not follow that an agent should omit any particular one of them. (Spielthenner 2007, p. 141)

Spielthenner's observations tell part of the story about sufficient condition schemata, but they do not tell the whole story. To fill in the blanks, we recall von Wright's gloss of 'ought' in conclusions of necessary condition schemata from Sect. 4.5.2.1. Something similar can be done for 'ought' in conclusions of sufficient condition schemata. To extrapolate von Wright's reading to sufficient condition schemata, we note that obligation, whether moral or not, can be prima facie or overall (cf. Zimmerman 1996, p. 207). In the prima facie case, we could take the conclusion of the bus inference to mean 'If I hurry, then I will attain one of my ends'. In the overall case, the conclusion can be understood as 'If I hurry, then I will attain one of my ends optimally'.

The application to the bus inference is straightforward. If the conclusion means 'If I hurry, then I will attain one of my ends optimally', it certainly does not follow from the premises, as Spielthenner explains. But if the conclusion is 'If I hurry, then I will attain one of my ends', the logical situation is different: this conclusion does follow from its premises. Provided the premises are true, the conclusion must be true as well. Hence sufficient condition schemata of the prima facie sort can be valid, as Audi points out (2004, p. 129; cf. Spielthenner 2007, p. 141 n. 6). Moreover, like the hut inference discussed in Sect. 4.5.2.2, the bus inference has a suppressed premise: 'To catch the bus is to attain one of my ends'. Adding it to the inference or substituting it for the first premise makes the inference formally valid.

Even though practical inferences like the bus inference in overall mode are not deductively valid, we might be quite willing to accept their logic and to act on the belief that their conclusions are true. The reason is that, from a practical point of view, the decisive consideration is whether the inferences are inductively cogent (cf. Spielthenner 2007, p. 144). As I argued in Sect. 2.4.1, necessary conditions for inductive cogency include a condition on the inference's content and a condition on the inference's form. The content condition is that all the premises be true. The formal condition is that the conditional probability of the conclusion be greater than or equal to that of any rival conclusion based on the same premises. If these conditions are not met, we should reject the inference.

In the assessment of practical inference, the formal component of the standard of inductive cogency permits great systematization. To see this, consider Audi's five patterns of practical reasoning (1989, pp. 146–149, 2004, pp. 128–131). These patterns, which draw on his schemata of practical reasoning, are as follows:

1. Necessity patterns follow a necessary condition schema or a rule schema and the major premise represents an overriding need.
2. Optimality patterns follow a sufficient condition schema, the major premise represents an overriding need, and the minor premise represents a belief that a certain means is the best way to satisfy the need.
3. Minimal adequacy patterns affirm the reasonableness of the conclusion given the premises: that is, believing the conclusion is at least as reasonable as believing its contradictory.
4. Standard adequacy patterns affirm greater reasonableness of the conclusion given the premises: that is, not to draw the conclusion would be unreasonable.
5. Cogency patterns affirm still greater reasonableness of the conclusion given the premises: that is, not to draw the conclusion would be extremely unreasonable.

Minimal adequacy patterns (pattern 3) are minimally adequate precisely because the conditional probability of their conclusions is only at least as great as that of any rival conclusion based on the same premises. The conclusions of standard adequacy patterns (pattern 4) have a conditional probability that is greater than that of any rival. The conclusions of cogency patterns (pattern 5) have a conditional probability that is much greater, though short of maximally greater, than that of any rival. Finally, the conclusions of necessity patterns (pattern 1) have a conditional probability that is maximally greater than that of any rival; that is, necessity patterns are valid (Audi 1989, p. 146, 2004, p. 128), which means that the conditional probability of their conclusions is 1 while that of any rival is 0.

Where, then, do optimality patterns (pattern 2) fit in? Optimality patterns are special cases of necessity patterns. A necessity pattern for a goal Φ and an act *A* might be represented as:

> I have an overriding need to Φ.
> Unless I *A*, I will not manage to Φ.
> Therefore I will fail to satisfy an overriding need unless I *A*.

Now as long as we substitute a description of a goal plus the expression ‘in the best way possible’ for ‘Φ’, we have an optimality pattern that emerges as a substitution instance of a necessity pattern. For example,

> I have an overriding need to resolve this conflict in the best way possible.
> Unless I apologize, I will not manage to resolve this conflict in the best way possible.
> Therefore I will fail to satisfy an overriding need unless I apologize.

The gain in systematicity is evident. The four fundamental patterns actually constitute a continuum of inductive acceptability. At the low end of the continuum are minimal adequacy patterns. Positions of increasing acceptability are then occupied respectively by standard adequacy, cogency, and necessity patterns. And optimality patterns fall in line as special cases of necessity patterns at the high end of the continuum.

4.5.3 Chaining Practical Inferences

Practical inferences focus on means that are actions, as the examples of Sect. 4.5.1 and Audi's description of practical reasoning in Sect. 4.5.2 show. The ends that figure in practical inferences are assumed, not justified. But since an end can be a means to a further end, an end that is assumed in one practical inference might be justified as a means to a further end in another practical inference. This process of linking practical inferences produces a chain of ends, short in some cases, long in others. A typical nonmoral example is the following:

Billy wants to meet Sally.
Unless Billy goes to the concert, he will not meet Sally.
Thus Billy needs to go to the concert.

Billy needs to go to the concert.
Unless Billy buys a ticket, he will not go to the concert.
Thus Billy needs to buy a ticket.

A moral example exemplifying the same structure is:

I want to be moral.
Unless I am just, I will not be moral.
Thus I need to be just.

I need to be just.
Unless I become less egoistic, I will not be just.
Thus I need to become less egoistic.

In principle, such chains can extend into the upper reaches of teleology. Here is an example with a strong Aristotelian flavor:

She wants to be happy.
Unless she is moral, she will not be happy.
Thus she needs to be moral.

She needs to be moral.
Unless she acquires the virtues, she will not be moral.
Thus she needs to acquire the virtues.

An alternative, non-Aristotelian justification for adopting the end of being moral is explored in Sect. 5.7 below.

4.5.4 The Kantian Alternative

The instrumental 'ought' of Sect. 4.5.2.1 covers moral and nonmoral usage. But the instrumental moral 'ought' has an imposing rival: the moral 'ought' defended by Kant. According to Kant, this instrumental 'ought' is nothing more than a hypothetical imperative. But the moral 'ought' is categorical, he claimed; it binds regardless of inclinations and consequences.

To evaluate this claim, let us recall three characteristically Kantian observations. "It is impossible to think of anything at all in the world, or indeed even beyond it,

that could be considered good without limitation except a **good will**" (1785, Ak 4:393). Second, "the highest and unconditional good alone can be found" in a good will (1785, Ak 4:401).[10] Finally, "reason is nevertheless given to us as a practical faculty, that is, as one that is to influence the *will*" (1785, Ak 4:396). Hence Kant believes that a moral description to the effect that the highest good is a good will can ground a moral directive that we ought to act with a good will.

Now juxtapose these statements with two Aristotelian observations. The first was this chapter's point of departure: every action is aimed at a good (Sect. 4.1). Recalling it here reminds us that Kantian moral action is aimed at a good will. Now the second observation: "It is debated, too, whether the choice or the deed is more essential to excellence, which is assumed to involve both; it is surely clear that its completion involves both" (*Nicomachean Ethics* 1178a34–b1). Viewing Kant's good will against the backdrop of the debate that Aristotle mentions reminds us that Kantian ethics is partisan. The relative merits of choice and deed, will and act, were already being debated in Aristotle's time, and Kant comes down squarely on the side of will: "an action from duty has its moral worth *not in the purpose* to be attained by it but in the maxim in accordance with which it is decided upon" (1785, Ak 4:399).

Questions arise in droves. Why should we set our moral sights on a good will rather than good results, given the staggering amount of suffering in the world? Why not aim for a combination of good deeds and good will, as Aristotle seems to imply? And why must a good will spring from duty, as Kant stipulates, rather than inclination? After all, a sick person might very well feel devalued, even treated as a mere means to someone else's moral quest, if the visitor is motivated by an impersonal sense of duty rather than inclination for the sick person. A good will appears to be one moral good among others. Why should it take priority over the rest?

Kant argues that the good will is superior to results and inclinations in the following passage:

> A good will is not good because of what it effects or accomplishes, because of its fitness to attain some proposed end, but only because of its volition, that is, it is good in itself and, regarded for itself, is to be valued incomparably higher than all that could merely be brought about by it in favor of some inclination and indeed, if you will, of the sum of all inclinations. Even if, by a special disfavor of fortune or by the niggardly provision of a stepmotherly nature, this will should wholly lack the capacity to carry out its purpose—if with its greatest efforts it should yet achieve nothing and only the good will were left (not, of course, as a mere wish but as the summoning of all means insofar as they are in our control)—then, like a jewel, it would still shine by itself, as something that has its full worth in itself. Usefulness or fruitlessness can neither add anything to this worth nor take anything away from it. (1785, Ak 4:394)

This argument makes a legitimate point: a good will that cannot accomplish its purpose nonetheless remains good.

[10] The description of the highest good in the *Critique of Practical Reason* is more complex: "Now, inasmuch as virtue and happiness together constitute possession of the highest good in a person, and happiness distributed in exact proportion to morality (as the worth of a person and his worthiness to be happy) constitutes the *highest good* of a possible world, the latter means the whole, the complete good, in which, however, virtue as the condition is always the supreme good, since it has no further condition above it" (1788, Ak 5:110–111).

But the argument has further ramifications. To see what they are, let us consider agency as it unfolds in time from will to act to result. Considering this process permits us to see that Kant's description of the possibility that "this [good] will should wholly lack the capacity to carry out its purpose—if with its greatest efforts it should yet achieve nothing" covers two distinct cases. In the first, a good will is present but circumstances prevent the act from being performed; hence the agent's purpose is not accomplished. In the second, a good will is present and the act is performed, but "a special disfavor of fortune" or "the niggardly provision of a stepmotherly nature" prevents the intended result; hence the agent's purpose is not accomplished. The good will of the first case would still shine by itself like a jewel, just as Kant says. But the good will *and* the act of the second case would also shine by themselves—which Kant does not say. The second case motivates the decisive point that a good act does not suffer by comparison with a good will. Both are authored by the agent, and both remain good even if their purpose is not accomplished. All the Kantian argument shows is that if the process of agency is interrupted due to circumstances beyond the agent's control, any part of the process prior to the interruption that is good remains good despite the interruption. This prior part will sometimes be the will, sometimes the will and the act. The only priority that can be claimed for the will is temporal.

I suggest, then, that a good will is one moral good among others and that other moral goods may rightly take precedence on occasion. One example is the life-saving lie to the would-be assassin of a friend in Kant's "On a Supposed Right to Lie from Philanthropy" (1797). Let the criterion of good will be the categorical imperative in the formulation of universal law and let the maxim truly describe the situation as one where lying is possible. Then the maxim for the lie would not express a good will, as Kant argued, yet the act would be good on almost any account but Kant's (Bok 1979, pp. 39–44, 114–116). Granted, the discrepancy between good will and good act in this case depends on a description of the situation that may not be optimal; compare 'a situation where lying is possible' with 'a situation where lying is the only way to save an innocent friend from assassination', for example. Most people would presumably prefer the second, more complete description. But Kant seems to have presupposed something like the first description in his discussion of the case.

Be that as it may, there are other possible divergences between good will and good act. Fred Feldman adduces several cases (1978, pp. 116–117). One involves a man who decides to withdraw all his money from the bank when the Stock Market Index reaches 1000. If the criterion of good will is the categorical imperative in the formulation of universal law, the will to perform this act would not be good, Feldman argues. The maxim 'When the Stock Market Index reaches 1000, I shall withdraw all my money from the bank' could not be consistently willed to be a law of nature, for the banking system would collapse if everyone acted on it. Yet the act is morally permissible. Similarly, a second person decides not to become a doctor. The will to act in this way would not be good either, according to Feldman. The associated maxim could not be consistently willed to be a law of nature, for rational persons would recognize the need for some people to be doctors. Yet, once again, the act is morally permissible.

As a final example, take Hume's celebrated claim that it is "not contrary to reason to prefer the destruction of the whole world to the scratching of my finger" ([1739] 1973, 2.3.3.6, p. 416).[11] Suppose that I have the power to scratch Hume's finger against his will and thereby save the world. If the criterion of good will is the categorical imperative in the formulation of the end in itself, to use this power would be to treat Hume as a mere means. It would therefore be an action motivated by a bad will. But most people would take this to be the right thing to do. This act evaluation, unlike the act evaluations of the previous two paragraphs, admittedly depends on the assumption that consequences are morally relevant, and Kant rejects this assumption. I address Kant's view of this largest of meta-ethical issues in Sect. 4.6.1 below. For the moment, I ask the reader to tolerate the stopgap measure of an argument from near consensus. The view that consequences are not morally relevant is very much a minority view, both within theoretical ethics, where it is rejected by all but strong deontologists, and a fortiori in ordinary moral reasoning. This is not a reason to reject Kant's view, of course. But it is a reason to be wary of it.

Fortunately, it is not necessary to assume that consequences have this-world relevance for a weakened version of the argument about the scratched finger to go through. For even if the proposition that consequences are not morally relevant is true, it would not appear to be a necessary truth. In some possible worlds, then, consequences are morally relevant. These possible worlds include worlds where consequences are all that is morally relevant and worlds where consequences together with other considerations are morally relevant. In at least some of these worlds where consequences matter, scratching Hume's finger would be the better choice. In these possible worlds, therefore, a good will and a good act diverge.

These three sets of cases (the life-saving lie, the bank withdrawal and non-medical career, the scratched finger) show that a good will and a good act do not always coincide. This observation is not vulnerable to the counter that Kant would make to the argument that a good will and good results do not always coincide: the goodness of the good will is moral, while the goodness of good results is not. The reason is that the act is no less the work of the moral agent, no less expressive of her moral personality, than the will. Both will and act contrast with results, which are usually due in part to circumstances beyond the agent's control.

We are now prepared to carry out a non-Kantian maneuver: Kant's defense of the good will can actually be reversed. Because we can have a will that is not good and an act that is good, good acts are not good because of the will that produces them. They are good in themselves. Even if the will from which they spring is not good, they would still shine by themselves like jewels. The goodness or badness of the will can neither add to nor take away from their worth. Note that this conclusion could be strengthened, as Kant strengthened his, by claiming that not only is the goodness of a good act independent of the goodness of a good will but that a good act is morally superior to a good will. This stronger conclusion might even

[11] Mencius accused the Taoist philosopher Yang Chu of being unwilling to pluck a single hair to benefit the whole world.

be defensible on the basis of additional evidence. Within the narrow limits of the present discussion, however, it would be just as lopsided as Kant's conclusion that a good will is morally superior to a good act. But a weaker conclusion is, I think, fully justified: the evaluative independence of will and act provides no ground for maintaining the moral priority of either one.

If this is so, then the Kantian route to morality is closed. The priority Kant confers on the good will is, I suggest, more of a personal postulate than a requirement of practical reason. Hence the 'ought' associated with the good will turns out not to be categorical after all.[12] The goodness of the good will must be weighed against that of the act and—as I will argue in Sect. 4.6.1—that of the results. Each, like a jewel, can shine by itself.

4.6 Moral Theory Choice

This chapter has explored instrumental moral language at the level of individual sentences like 'That shows the value of courage' and 'Be honest so that people will respect you' (Sect. 4.4). It has also discussed practical inference from various vantage points (Sects. 4.5.1–4.5.4). We are now in a position to move up to theory. What Dewey called "the instrumental function of theory" ([1938] 1986, p. 468) is patent in sentences like 'The theory of cultural selection explains the high incidence of albinism in Hopi Indians'. It is equally evident in the context of theory choice, where we are concerned with the relative success of rival theories as cognitive instruments. Chapter 3 addressed the problem of theory choice in general terms, without reference to particular theories, by developing a form of comparative decision theory. It is now time to specialize by bringing comparative decision theory to bear on moral theory. The problem we will address is how to choose a moral theory on instrumental grounds.

No attempt will be made here—nor could it—to discuss all moral theories. Instead, I will provisionally describe a minimal field of three theories. The problem of theory choice will then be addressed relative to this field. Limited though the field is, the theories that make it up are not chosen at random. They occupy salient positions along what I will call the continuum of consequences. At one extreme of the continuum are strong deontological theories that take no account of consequences at all. At the opposite extreme are consequentialist theories that take account of nothing but consequences. Between the two extremes are mixed deontological theories that consider consequences, which makes them akin to consequentialism, but not exclusively, which links them to strong deontologism. These positions on the continuum are represented by three standard theories I propose for discussion: Kant's strong deontologism, Bentham's act-utilitarian consequentialism, and Frankena's

[12] For a diagnosis of the cultural situation that makes Kant's categorical 'ought' an attractive option, see MacIntyre (1981, pp. 105–107).

mixed deontologism. These theories have the advantage of familiarity, which will permit us to proceed directly to evaluation without expository delay.

The argument that follows falls into three parts. The first canvasses the comparative plausibilities of the theories in the field. The second investigates their comparative utilities. The third draws on the results of the two previous parts in an attempt to establish comparative plausibilistic expectations for all three theories. I assume throughout that the choice among these theories is pragmatically forced (Sect. 3.6.3); that is, the option of suspending judgment about these theories is unsatisfactory.

4.6.1 Plausibility

We begin, then, with comparative plausibility. Recall that Sect. 3.3.2.4 took a plausibility measure π to map propositions about attributable states of the world to plausibility values. The relevant states for scientific theory choice are putative phenomena described by the theories, but what would be the relevant states for moral theory choice? I submit that two classes of states are germane to the choice of a moral theory on instrumental grounds.

The first of these classes can be located via the phenomenal stratum of moral discourse. As we noticed in Sect. 2.3.2, the presence of certain initial conditions may lead us to describe actions and people as courageous, honest, and cruel, much as initial conditions of another sort might prompt us to describe certain organisms as lepidoptera. Drawing on this phenomenal stratum, a Frankenian might describe a certain act as unjust. That the act *is* unjust is the corresponding *ontic state*. Similarly, a Kantian may describe an intention as an instance of the good will, and a utilitarian could claim that a given act produces disutility. These descriptions posit additional ontic states. In surveying the class of ontic states, let us recall that some ethical theories posit states that intersect with domains of the social sciences. Mill, for example, makes the factual claim that human nature desires "nothing which is not either a part of happiness or a means of happiness" ([1861] 1969, Chap. iv, p. 237).

Besides descriptions of putative ontic states, moral discourse includes a great deal of permission, prohibition, and obligation. What matters to discourse of this sort is that actions are—and are not merely taken to be—morally permitted, prohibited, and obligatory. That breaking a certain promise is morally permitted, charging a determinate price is morally prohibited, and telling a particular truth is morally obligatory are all examples of *deontic states*. These examples happen to be first-order deontic states that bear directly on actions in the domain of moral theory. But other deontic states are second-order; they permit, prohibit, or obligate certain procedures in moral decision making. That use of the hedonistic calculus is obligatory and that moral consideration of consequences is prohibited are meta-ethical instances of deontic states.

Talk about deontic states somehow feels different than talk about ontic states like a weight that is heavy and a gift that is generous. They are different, I claim, but the difference is not that ontic states are factual and deontic states are not. Deontic states, as the readings of terms like 'ought' and 'must' in Sect. 4.5.2.1 suggest, are instrumental states.[13] In general, when we say that an action is permitted, prohibited, or obligatory, we are saying that the action may or must not or ought to be performed in order to achieve some end. More specifically, when we say that an action is morally permitted, prohibited, or obligatory, we are saying that the action may or must not or ought to be performed in order to achieve a moral end. In both the general and specifically moral cases, we are asserting instrumental facts, instrumental states of affairs. Abandoning a child is morally impermissible because it is incompatible with treating the child beneficently (to name just one moral end). Wearing either my red sweater or my blue sweater is morally permissible in normal circumstances because neither one would prevent the attainment of moral ends. The contrast between ontic and deontic states is not factual versus nonfactual, therefore; it is phenomenal versus instrumental. A morally phenomenal state like the injustice of a certain policy contrasts with a morally instrumental state like the impermissibility of injustice.

To ignore the differences between ethical theory and scientific theory would be foolhardy, but they do have this much in common: just as quantum mechanics, say, posits phenomenal states like the position of a photon and instrumental states directing the two-slit experiment, ethical theories relate their own phenomenal and instrumental states. What I propose, then, is to take ethical theory at its word and recognize the mixture of states, phenomenal and instrumental alike, that are subjects of moral discourse. I will refer to the states posited by Kantian, Benthamite, and Frankenian theory (in conjunction with statements of initial conditions) as Kantian, Benthamite, and Frankenian states.

The number of these states could conceivably reach a potential infinity. But human inquirers can meaningfully consider only a finite number of states or classes of states.[14] Fortunately, the fact that we are working with rival theories serves to focus our attention. Since we cannot process all of our theories' implications about states, we single out some for serious consideration. The implications are chosen with an eye to contrast, where one theory implies one state and another theory another. The contrasting states could be phenomenal, like the justice and injustice of a certain action, or instrumental, such as obligations to perform mutually exclusive actions. In either case, the chosen implications function like test implications in science, and their truth and falsehood are analogous to the results of crucial experiments.[15] The truth and falsehood of the chosen implications become the inductive basis of any generalization we may hazard about the theory as a whole. These chosen implications are necessarily finite.

[13] Teleological states are instanced in the final paragraph of Sect. 5.5.

[14] "The reason why we cannot survey an infinite totality is not the deficiency of human capabilities: it is that it is *senseless* to imagine an infinite task completed. An infinite task is by definition one that cannot in principle be completed" (Dummett 2006, pp. 70–71).

[15] This process of confirmation is described more fully in Sect. 5.4.

The plausibilities to be estimated are the plausibilities that certain states hold given the evidence. But what evidence could show that moral states hold? We might attempt to respond to this question with answers of varying lengths. A short answer would be that evidence is whatever the agent thinks it is. That is, comparative decision theory is a branch of individual decision theory, and individual decision theory accepts whatever evidential and plausibilistic inputs agents provide. At the other extreme, a long answer to the question would take the form of another book. The reader may be relieved to know that such a book will not materialize here.

However, a medium-length answer to the question might take shape along the following lines. Many agents take the evidence for morally relevant states to be a mixed bag that includes at least five sorts of evidence:

1. logical consistency;
2. the adequacy of any explanations a theory might make (why a lie, say, might be morally required);
3. the truth of any predictions ventured (that certain actions are not conducive to happiness, for example);
4. coherence with background knowledge (psychological, historical, sociological, etc.);
5. coherence with what Alan Donagan called "common morality" (the large part of traditional morality that does not depend on theistic assumptions) (1977, pp. 27–28; cf. Gert 2004).

Note that to take common morality as one strand of evidence is not to assume it uncritically. Like any other element in our belief structure, beliefs reflecting common morality may have to be revised or rejected (Rawls 1971, p. 49). But just as elementary mathematical beliefs guide construction at the foundations of mathematics,[16] elementary moral beliefs like the impermissibility of torturing children guide the construction of moral theory.

To assess the plausibility of Bentham's act-utilitarian states, we will review several well-known difficulties with the position. The first is what J. S. Mill called "the only real difficulty in the utilitarian theory of morals": the charge that utility does not imply justice ([1861] 1969, Chap. v, p. 259). Critics of utilitarianism usually urge this point through counterexamples. To establish that utility does not imply retributive justice, they instance the utility-maximizing execution of an innocent person to prevent rioters who demand a culprit from exacting greater loss of life (Foot 1967, p. 8).[17] To show that utility does not imply distributive justice, they moot a utility-maximizing proposal to kill a healthy patient so that her organs could be used in multiple life-saving operations (Foot 1967, p. 9). A second problem is that act-utilitarianism would encourage me to make a promise I have no intention of keeping because the promise would produce more utility than not making it. Integrity, as Bernard Williams pointed out, is not a utilitarian virtue (Williams and Smart

[16] Cf. the quotation from *Principia Mathematica* in Sect. 1.2.

[17] Kai Nielsen argues that there are utilitarian and consequentialist grounds for not framing and executing the innocent in such cases (1972).

1973, pp. 108–118). Still another problem is that the hedonistic calculus appears to condone pleasures that are sadistic, racist, and sexist in the same way and to the same degree as pleasures that, by nonutilitarian standards, are innocent. This list of ills is all too brief, but it could easily be extended (Harwood 2003, pp. 179–192). Even in this attenuated form, however, I suggest that it is sufficient for an initial appraisal of the plausibilities of many act-utilitarian instrumental states.

A possible objection to such an appraisal is that one person's modus tollens is another's modus ponens. The arguments sketched in the preceding paragraph have the underlying form of modus tollens, but a convinced act-utilitarian could counter by recasting them as modus ponens. For example: if act-utilitarianism is correct, then the utility-maximizing killings instanced above are just; act-utilitarianism is correct; therefore these killings are just. But because such counters must bear a crushing burden of proof, I will ask the reader to assume momentarily, for the sake of illustration at least, that such killings are as unjust as they appear. Later on, in Sect. 6.2, I argue that there is a way to submit such act-utilitarian counters and their common-morality contradictories to a critical tribunal.

Turn now to Kant's strong deontologism, which posits an instrumental state that prohibits moral consideration of consequences. We ran into this issue already in discussing the life-saving lie and the scratched finger in Sect. 4.5.4. If Kant is right about the role of consequences in ethics, the life-saving consequences of both these actions are morally irrelevant. We do not have to be strict consequentialists to find this impracticable. Even Kant was unable to practice what he preached, as Mill points out. When the categorical imperative is applied to "the most outrageously immoral rules of conduct," all Kant managed to show "is that the *consequences* of their universal adoption would be such as no one would choose to incur" ([1861] 1969, Chap. i, p. 207). For example, someone who has formed the maxim to never help the needy could not universalize it, as Kant observed, for "many cases could occur in which one would need the love and sympathy of others and in which, by such a law of nature arisen from his own will, he would rob himself of all hope of the assistance he wishes for himself" (1785, Ak 4:423). This is consequentialist thinking. Any attempt to be more consistently anti-consequentialist than Kant would be inconsistent with all other forms of practical reason, for no other area of practical reason treats consequences as forbidden fruit. These considerations strongly suggest that practical reason that ignores consequences is just impractical. We could put the point by paraphrasing Kant himself: ethics without consequences are empty; consequences without ethics are blind.

Frankena's mixed deontologism has two fundamental principles: a principle of beneficence that enjoins doing good and preventing harm, and a principle of justice that prescribes equal treatment. Pluralistic theories like Frankena's have a marked disadvantage with respect to monistic theories like Bentham's and Kant's: the logical possibility of deontic inconsistency among the principles. In Frankena's case, the principle of beneficence might require one action and the principle of justice another. If conflict between justice and beneficence arises, which should come first? Frankena acknowledges the problem and admits that he has no *general* solution to offer. But he does offer some remarks that are pertinent to managing the conflict on a case-by-case basis:

> It is tempting to say that the principle of justice always takes precedence over that of beneficence: do justice though the heavens fall. But is a small injustice never to be preferred to a great evil? Perhaps we should lean over backwards to avoid committing injustice, but are we never justified in treating people unequally? One might contend that the principle of equal treatment always has priority at least over the fourth or positive part of the principle of beneficence, but is it never right to treat people unequally when a considerable good is at stake? The answer to these questions, I regret to say, does not seem to me to be clearly negative…. (1973, pp. 52–53)

Frankena's position here amounts to positing two sets of instrumental states. (1) Normally, justice should trump beneficence. (2) Exceptionally, beneficence *might* trump justice if a small injustice would avoid a great evil or obtain a great good. These remarks leave a good deal of residual indeterminism, of course. But they do provide some guidance in resolving cases of conflict between the principles.

Let us now attempt to assess the relative plausibilities of states relevant to our three theories. Bentham's act-utilitarianism, as we have seen, is plagued with grave difficulties: injustice, insincere promises, and illicit pleasures among them. By contrast, Kantian theory does not have such strongly counterintuitive implications. I suggest, then, that the overall plausibility of the Kantian states is greater on the evidence than the overall plausibility of the Benthamite states. Yet Kantian ethics suffers from the impracticality of its ban on consequences. Frankenian ethics does not, though it does have an unresolved problem of conflict between its principles. Neither problem is trivial, but the Kantian difficulty affects every moral decision without exception, whereas Frankena's arises only in exceptional cases. More importantly, the two problems are of radically different kinds. The Kantian problem affects the plausibility of the instrumental state that bans consideration of consequences, but the Frankenian problem diminishes the theory's effectiveness as a guide to action. Hence the Frankenian problem does not really affect the plausibilities of states; rather, it reduces the theory's epistemic utility. I suggest, then, that the overall plausibility of the Frankenian states is greater on the evidence than the overall plausibility of the Kantian states. Drawing on the relation of supraplausibility defined in Sect. 3.3.2.6, we can represent the relative plausibilities π of a finite number n of Benthamite act-utilitarian states ($s_{b1} \ldots s_{bn}$), Kantian strong deontological states ($s_{k1} \ldots s_{kn}$), and Frankenian mixed deontological states ($s_{f1} \ldots s_{fn}$) on the total evidence e as follows:

$$\pi(s_{f1} \ldots s_{fn} \mid e) > \pi(s_{k1} \ldots s_{kn} \mid e) > \pi(s_{b1} \ldots s_{bn} \mid e).$$

This ordering might have been anticipated by reflecting that Frankena appears to have aimed his version of mixed deontologism at the principal weaknesses of the other two. The principle of beneficence focuses on consequences—more fundamentally, Frankena argued, than the principle of utility—and thus remedies Kant's consequential emptiness (1973, pp. 45–46). The principle of justice corrects the utilitarian weakness acknowledged by Mill.

As already noted, these comparative plausibilities reflect degrees of belief—specifically, the writer's degrees of belief. Admittedly, however, I am recommending them to others. But if others are unpersuaded, if their judgments of comparative plausibility differ from mine, the argument that follows can still play a useful role.

It can illustrate the extent to which comparative plausibilities can be subsumed within a decision-theoretic framework. How competing estimates of comparative plausibility could be employed will become evident as we proceed.

4.6.2 Utility

The second stage of our inquiry concerns the utilities of outcomes that would result from choosing the theories in our field. Theory choice may yield outcomes of very different sorts, and many of these outcomes could be of interest to the decision maker alone or to the decision maker and an idiosyncratic clique. But in a community whose members share a cognitive aim, information conducive to this aim is valued by the entire community. In such a community, therefore, any theory perceived to be more informative than others about phenomena in the community's cognitive domain should receive, to that extent, a higher intersubjective utility.

The distinction between partly cognitive and purely cognitive decisions drawn in Sect. 3.3.2.3 is relevant to the argument that follows. A partly cognitive decision is based upon consideration of information as one relevant outcome among others. A purely cognitive decision rests on information as the only relevant outcome. The choice among the theories in our field will be treated here as purely cognitive, that is, as a decision based on informational outcomes alone. Hence the only utilities to be considered are utilities of information.

Information, we assumed in Sect. 3.3.2.3, is reduction of uncertainty about states of the world. An ethical theory can reduce uncertainty about morally phenomenal states such as those instanced at the outset of Sect. 4.6.1. It can also reduce uncertainty about the best way to act. The reduction of uncertainty about how to act is what Luciano Floridi has called "instructional information" (2005, § 3.1). Note that instructional information is not information in some new sense of the term. We have already seen in Sect. 4.5.2.1 how sentences bearing terms like 'must' and 'ought' can be naturally construed as instrumental sentences that bear truth values. We saw that such construals can conform to Audi's practical inference schemata. Sentences that purport to provide action-guiding information can be interpreted along the same factual lines. The instrumental premise 'If you turn left at the first stoplight, you will find the gas station' and the conclusion 'Turn left at the first stoplight' may instantiate a sufficient reason schema, for example. Or the instrumental premise 'You will get your mortgage only if you sign here' and the conclusion 'Sign here' may instantiate a necessary condition schema. Action-guiding information is, at bottom, factual information about how to realize possible states of the world.

Looked at instrumentally, then, a moral theory should provide action-guiding information; it should inform by reducing uncertainty about how to act. To determine how successfully the theories in our field do this, I propose consideration of a specific moral dilemma. Take Sophie's choice in William Styron's eponymous novel, for instance. Sophie, a young Polish woman in Nazi-occupied Warsaw, has been caught stealing a ham for her tubercular mother. The Nazis deport her and her two young children, Jan and Eva, to Auschwitz. Upon their arrival, a drunken SS doctor accosts Sophie. The pertinent part of their dialogue is as follows:

> [T]he doctor said, "You may keep one of your children."
> "*Bitte*?" said Sophie.
> "You may keep one of your children," he repeated. "The other one will have to go. Which one will you keep?"
> "You mean, I have to choose?"
> "You're a Polack, not a Yid. That gives you a privilege—a choice."
> Her thought processes dwindled, ceased. Then she felt her legs crumple. "I can't choose! I can't choose!" She began to scream. Oh, how she recalled her own screams! Tormented angels never screeched so loudly over hell's pandemonium. "*Ich kann nicht wählen!*" she screamed.
> The doctor was aware of unwanted attention. "Shut up!" he ordered. "Hurry now and choose. Choose, goddamnit, or I'll send them both over there. Quick!"
> She could not believe any of this. She could not believe that she was now kneeling on the hurtful, abrading concrete, drawing her children toward her so smotheringly tight that she felt that their flesh might be engrafted to hers even through layers of clothes. Her disbelief was total, deranged. It was disbelief reflected in the eyes of the gaunt, waxy-skinned young Rottenführer, the doctor's aide, to whom she inexplicably found herself looking upward in supplication. He appeared stunned, and he returned her gaze with a wide-eyed expression, as if to say: I can't understand this either. (1979, Chap. 15, p. 589)

What are poor Sophie's options here? She can try to save Jan by sacrificing Eva. She can attempt to save Eva by sacrificing Jan. She can refuse to sacrifice either child. Or, out of despair or hatred or sheer human perversity, she could offer to sacrifice both. Thus she has four fundamental options.[18]

Now just as someone who has lost her way might ask a passerby for directions, we can imagine Sophie asking the theories in our field. If she were to do her asking decision-theoretically, she would need to conceptualize the situation as follows. The theoretical options under consideration are choosing Kantian theory, choosing Benthamite theory, and choosing Frankenian theory. The states are the first-order instrumental states relevant to Sophie's choice: she ought to attempt to save Jan; she ought to attempt to save Eva; she ought to attempt to save both; and she ought to attempt to save neither. Different combinations of acts and states result in different information outcomes. Suppose that the states posited by Bentham's act-utilitarianism are more truthlike than their rivals. If Sophie were to choose Benthamite theory, then, the outcome would be some degree of information about instrumental states of the world. But had she chosen Kantian or Frankenian theory instead, the result would be some degree of misinformation about these states.

Since information is reduction of uncertainty, the success of these theories in providing action-guiding information is measured by the extent to which they eliminate options. Information can be quantified in a natural way by relating the number of eliminated options to the total number of options. The relation, expressed for information i, eliminated options e, and total options t, is traditionally taken to be $i = e/t$ (Bar-Hillel and Carnap 1953, pp. 147–157; Levi 1967, p. 74). For our comparative purposes, however, it is sufficient to know that a theory that eliminates two of Sophie's four options, say, is more informative than a theory that eliminates just one.

[18] This is the most natural parsing of the situation, but other sets of options are conceivable. The issue is discussed in Sect. 6.2.

Bentham's act-utilitarianism would surely reject the options of acting to save both children and acting to lose both.[19] The argument would be that these options would probably result in the death of both children, whereas the other two options would probably lead to the death of one child each. Given the case as described, however, there seems to be no act-utilitarian reason for sacrificing one child rather than the other. Both these acts would be seen as permissible but neither as obligatory (cf. Zimmerman 1996, p. 209, pp. 220–221). Granted, it is not hard to imagine further circumstances that would incline an act-utilitarian in one direction rather than another. Nor is it hard to counter this move by imagining other circumstances that would make the decision a toss-up. But circumstance padding in either of these forms is beside the point. Any such circumstances would have to be added to the case, and that is to change the subject. If we stick to the case as described, the act-utilitarian recommendation would be to sacrifice one child or the other.

Let us turn, then, to Kantian ethics. As Kantians, our criterion of duty would be the categorical imperative. Suppose we apply its second formulation, the formulation of the end in itself, to Sophie's choice. This formulation would evidently forbid Sophie to sacrifice either child for the other, for that would be to treat one child as a mere means in order to save the life of the other. Nor, of course, could Sophie offer to sacrifice both children, for that would be to treat both as mere means. As a result, the Kantian solution would be to attempt to save both children.

Finally, consider Frankena's mixed deontologism. Offering to sacrifice both children would clearly violate Frankena's principle of beneficence, which enjoins us to prevent harm. And, given the probable consequences, attempting to save both children would violate beneficence as well. The most beneficent acts, it would appear, are the two in which one child is sacrificed for the other. But these acts are also egregiously unjust. Hence there is a conflict between the Frankenian principles of beneficence and justice. As we have seen, Frankena recommends that if the two principles conflict, justice normally comes first. But he leaves the door open to exceptions in favor of beneficence if a small injustice would prevent a great evil or obtain a great good. So the question is whether the most beneficent acts the situation permits qualify as exceptions by these lights. A moment's reflection indicates that they do not; to sacrifice one child for another is a grievous injustice. Hence Frankena's ethics, like Kant's, would prohibit sacrifice; it would insist on the attempt to save both children.

The results of our inquiry about information are then as follows. Bentham's act-utilitarianism narrows Sophie's options down to two. By contrast, Kant's strong deontologism and Frankena's mixed deontologism determine unique results. Consequently, it seems appropriate to say that all three theories are informative, but they are not all informative to the same degree. For Sophie's choice, Kant's and Frankena's theories are more informative than Bentham's, and the two deontological theories are equally informative. Since we are treating choice among the

[19] However, Daniel Hausman has suggested that utilitarianism may have no direct implications for conduct at all (1991, pp. 273–278). For an argument that all forms of consequentialism provide moral standards but no decisive reasons for following them, see Hurley (2006, pp. 680–706).

three theories as a purely cognitive decision, utility is preference for information alone. Drawing on the relations of supradesirability and equidesirability sketched in Sect. 3.3.2.6, we can therefore represent the utilities υ of our three information outcomes *i* as follows:

$$\upsilon(i_f) > \upsilon(i_b); \upsilon(i_k) > \upsilon(i_b); \upsilon(i_f) = \upsilon(i_k).$$

Just as we arrived at estimates of the comparative plausibilities of these theories above, we now have estimates of their comparative utilities for Sophie's choice.

4.6.3 Plausibilistic Expectation

The third stage of our inquiry is to bring the results of the first two stages to bear on the plausibilistic expectations of the theories in our field. I will proceed by breaking the comparison down into two sequential decisions. The first is the choice between Frankena or Kant, on the one hand, and Bentham, on the other. Here the alternatives are Frankena *or* Kant versus Bentham because Frankena and Kant occupy the same comparative positions with respect to Bentham. Given the foregoing considerations, that is, Frankenian and Kantian states are both more plausible than Benthamite states, and Frankenian and Kantian outcomes are both more informative than Benthamite outcomes. From this comparative perspective, then, which one we consider does not matter. After disposing of this initial choice, we will take up that between Frankena and Kant. Both choices will be treated as decisions under risk. For the duration of the exposition, the comparative results of Sects. 4.6.1 and 4.6.2 will be assumed without further ado. To represent these results, let *U* and *u* express higher and lower utility, and let *P* and *p* stand respectively for higher and lower plausibility based on the evidence.

Decision-theoretic comparison of Frankenian or Kantian theory with Benthamite theory calls for the use of *PE* (Sect. 3.3.2.7, Eq. 3.4). We can prepare the way by noting that Frankenian and Kantian theory promise utility *U* with plausibility *P* and disutility $-u$ with plausibility *p*. Bentham's theory, on the other hand, offers utility *u* with plausibility *p* and disutility $-U$ with plausibility *P*. By *PE*, then, the plausibilistic expectation for Frankena or Kant is:

$$PE_{f,k} = UP \oplus -up$$

whereas the plausibilistic expectation for Bentham is:

$$PE_b = up \oplus -UP.$$

Since $U > u$ and $P > p$, the plausibilistic expectation for Frankena or Kant is positive while that for Bentham is negative. We have an instance of case 5 from Table 3.4

(Sect. 3.5.1). By these lights, Bentham's act-utilitarianism is inferior to the other two theories.

The second round of comparison is the runoff between Frankena and Kant. Since both theories proffer the same advice about Sophie's choice, they carry the same utility U and disutility $-U$. However, Frankena's theory yields U with plausibility P and $-U$ with plausibility p; for Kant, the plausibilities are reversed. According to *PE*, then, the plausibilistic expectation for Frankena is:

$$PE_f = UP \oplus -Up = U(P \oplus -p),$$

and that for Kant is:

$$PE_k = Up \oplus -UP = U\left(p \oplus -P\right).$$

Consequently, the ratio formed by the two expectations is determined entirely by their comparative plausibilities. This value is positive for Frankenian theory and negative for Kantian theory. This is an instance of case 6 from Table 3.4 (Sect. 3.5.1). On these grounds, then, Frankenian theory is superior to Kantian theory.

Combining the results of both rounds of comparison would permit the plausibilistic expectations of the three theories to be represented as follows:

$$PE\left(f\right) > PE\left(k\right) > PE\left(b\right).$$

On decision-theoretic grounds, Frankenian theory is a better guide than its rivals for Sophie's choice. If this is so, then Sophie's choice is conceptually clear, though undoubtedly charged with enormous residues of guilt and regret. Sophie should refuse to sacrifice either child in an effort to save them both.

I invite the reader to consider this last claim in the context of the familiar distinction between obligation dilemmas and prohibition dilemmas. Obligation dilemmas impose choice among morally obligatory actions; a standard example is Sartre's student, agonizing between aiding his mother and joining the French resistance ([1946] 1956, pp. 295–297; cf. Lebus 1990, p. 116). Prohibition dilemmas, on the other hand, require choice among morally prohibited actions. Patricia Greenspan treats Sophie's choice as a prohibition dilemma, indeed a dilemma of "exhaustive prohibition," because all of Sophie's options are morally wrong (1983, p. 118).

Greenspan's analysis coincides largely but not entirely with the foregoing discussion. Both take the three options that involve sacrificing one or both children to be immoral. But Greenspan's claim that the refusal to sacrifice either child is immoral contradicts my claim that this refusal is moral. The reason for the discrepancy, I think, is that morally relevant variants of this fundamental option have not been unpacked. Greenspan understands the refusal option as "letting die," "simply standing by," and "doing nothing"; she correctly observes that this would violate Sophie's duty to protect her young children (1983, pp. 118–119). But refusal can be active as well as passive. Sophie and her children were not alone with the doctor. The doctor's aide was present, "stunned" with "disbelief," and the doctor was

"aware of unwanted attention" due to Sophie's screams. Twice in succession, the doctor urged her to hurry. Why did he want her to act quickly? Whose attention did he not want to attract? That attention and the implicit divergence in the attitudes of doctor and aide represent the chance that Sophie could have resisted by continuing to scream, continuing to solicit the attention of anyone less evil than the doctor. The struggle to bring this chance to life is not immoral. It is, on the contrary, morally required, enjoined by Sophie's condition as a parent of young children. And it is seconded not only by Frankenian theory but also by Kantian ethics, Frankena's strongest rival in the case.

My claim about the decision-theoretic superiority of Frankenian ethics should not be digested without a pair of caveats. The first is that this conclusion is plainly relative to the theories in our field. It is logically possible that some theory not included in the field—a more sophisticated form of consequentialism, for example—could best Frankena's theory on decision-theoretic grounds. The other caveat is that the estimates of the theories' plausibilistic expectations are supported by arguments of unequal scope. The argument about the comparative plausibilities of their states is entirely general, whereas the argument about the comparative utilities of their outcomes relies on a single case. Strictly speaking, then, the conclusion about the theories' plausibilistic expectations is limited to the case upon which the comparative utilities are based. It is logically possible that a theory that is more (or less) informative about Sophie's choice might turn out to be less (or more) informative about some other case, and that this could alter the theories' comparative expectations for the other case. I will therefore return to the relative advantages of our three approaches to morality in Sect. 5.5.

Anyone pondering a decision-theoretic approach to Sophie's choice might very well question the tactics employed above. The questioning might run as follows. Why take the long way around? Why make the arduous detour through ethical theory? Why not bring decision theory directly to bear on Sophie's choice? My response is that it is possible and it is direct. But there are two great advantages to dealing first with theory, then with Sophie's choice.

The first can be seen against the backdrop of a problem already noticed in connection with act-utilitarianism: the tolerance of the hedonistic calculus for pleasures that are sadistic, racist, and sexist. Decision theory is comparable in this respect: it has no resources whatever for screening out immoral preferences. Suppose that Sophie, terrified for her own safety and enraged with her children for their behavior on the thirty-hour train trip, prefers that they be led away to their deaths. Decision theory would accept this preference on a par with the exalted preferences of a saint. One advantage of passing through ethical theory before turning to Sophie's choice, then, is that the theory acts as a kind of filter for identifying immoral preferences.

The other advantage emerges from the fact that moral deliberation is not a series of one-night stands. We do not decide and decide anew without looking for continuity in our moral experience. But concentrating exclusively on the moral dilemma at hand neither draws on our experience in the past nor extends it toward the future. By contrast, moral theory connects the dots. It relates the problem of the moment to other cases extending back and forward in time. Hence a reliable theory, if one could be found, would help us not just this time but the next.

4.7 Conclusion

Instrumental discourse purports to tell the truth about how to achieve ends. This is no less true of moral than nonmoral discourse. Instrumental moral discourse has been the focus of this chapter, which has argued that the veridicality of such discourse can be assessed in relatively straightforward ways: for individual sentences, by careful observation and inductive inference (Sect. 4.4); for practical inferences, by the standard of inductive cogency (Sect. 4.5.2); and for theory, by comparative decision theory (Sect. 4.6). So whether our focus is an individual sentence, an inference, or a theory, instrumental moral discourse attempts to indicate how to obtain certain ends. Instrumental moral discourse is therefore cognitive. The task of separating what is genuinely cognitive from what is cognitive in intent, but not result, can be daunting indeed. But this is true of instrumental discourse of all kinds.

The following chapter is as closely linked to this one as ends are to means. It undertakes an examination of teleological moral discourse. The comparative evaluation of moral theories undertaken here on instrumental grounds (Sect. 4.6) is continued there on teleological grounds (Sect. 5.5). The arguments are meant to be complementary.

References

Allan, D. J. 1955. The practical syllogism. In *Autour d'Aristote,* ed. Auguste Mansion, 325–340. Louvain: Publications Universitaires de Louvain.

Angner, Erik. 2013. Is it possible to measure happiness? *European Journal for Philosophy of Science* 3:221–240.

Aquinas, Thomas. 1265–1273. *Summa theologica*. English edition: Aquinas, Thomas. 1947. *Summa theologica* (trans: Fathers of the English Dominican Province). New York: Benziger Brothers.

Aristotle. 1984. *Movement of animals*. In *The complete works of Aristotle,* ed. Jonathan Barnes, Vol. I, 1087–1096. Princeton: Princeton University Press.

Aristotle. 1984. *Nicomachean ethics*. In *The complete works of Aristotle,* ed. Jonathan Barnes, Vol. II, 1729–1867. Princeton: Princeton University Press.

Aristotle. 1984. *On the soul*. In *The complete works of Aristotle,* ed. Jonathan Barnes, Vol. I, 641–692. Princeton: Princeton University Press.

Aristotle. 1984. *Politics*. In *The complete works of Aristotle,* ed. Jonathan Barnes, Vol. II, 1986–2129. Princeton: Princeton University Press.

Audi, Robert. 1989. *Practical reasoning*. London: Routledge.

Audi, Robert. 2004. Reasons, practical reason, and practical reasoning. *Ratio* 17:110–149.

Bar-Hillel, Yehoshua, and Rudolf Carnap. 1953. Semantic information. *The British Journal for the Philosophy of Science* 4:147–157.

Bentham, Jeremy. 1789. *An introduction to the principles of morals and legislation*. Rpt. London: University of London, Athlone Press, 1970.

Bok, Sissela. 1979. *Lying: Moral choice in public and private life*. New York: Vintage Books.

Broome, John. 2001. Normative practical reasoning. *Supplement to the Proceedings of The Aristotelian Society* 75:175–193.

Dewey, John. 1922. *Human nature and conduct: An introduction to social psychology.* In The middle works, 1899–1924, eds. Jo Ann Boydston and Patricia Baysinger, vol. 14. Carbondale: Southern Illinois University Press, 1983.

Dewey, John. 1938. *Logic: The theory of inquiry.* In The later works, 1925–1953, ed. Jo Ann Boydston, vol. 12. Carbondale: Southern Illinois University Press, 1986.
Donagan, Alan. 1977. *The theory of morality*. Chicago: The University of Chicago Press.
Dummett, Michael. 2006. *Thought and reality*. Oxford: Clarendon Press.
Feldman, Fred. 1978. *Introductory ethics*. Englewood Cliffs: Prentice-Hall.
Floridi, Luciano. 2005. Semantic conceptions of information. In *Stanford encyclopedia of philosophy*. http://plato.stanford.edu/entries/information-semantic. Accessed 30 April 2014.
Foot, Philippa. 1967. The problem of abortion and the doctrine of the double effect. *The Oxford Review* 5:5–15. Rpt. in *Virtues and vices and other essays in moral philosophy*, 19–32. Oxford: Basil Blackwell. (Berkeley: University of California Press, 1978).
Frankena, William K. 1973. *Ethics*. 2nd ed. Englewood Cliffs: Prentice-Hall.
Gert, Bernard. 2004. *Common morality: Deciding what to do*. New York: Oxford University Press.
Greenspan, Patricia. 1983. Moral dilemmas and guilt. *Philosophical Studies* 43:117–125.
Harman, Gilbert. 1978. Reasons. In *Practical reasoning,* ed. Joseph Raz, 110–117. Oxford: Oxford University Press.
Harwood, Sterling. 2003. Eleven objections to utilitarianism. In *Moral philosophy: A reader,* ed. Louis Pojman, 3rd ed., 179–192. Indianapolis: Hackett.
Hausman, Daniel M. 1991. Is utilitarianism useless? *Theory and Decision* 30:273–278.
Hempel, Carl G. 1965. *Aspects of scientific explanation*. New York: The Free Press. (London: Collier-Macmillan).
Hobbes, Thomas. 1651. *Leviathan*. Rpt. Indianapolis: Hackett, 1994.
Hume, David. 1739. *A treatise of human nature*. Rpt. Oxford: Clarendon Press, 1973.
Hurley, Peter E. 2006. Does consequentialism make too many demands, or none at all? *Ethics* 116:680–706.
Irwin, T. H. 2003. A sort of political science. In *The classics of Western philosophy: A reader's guide,* eds. Jorge J. E. Gracia, Gregory M. Reichberg, and Bernard N. Schumacher, 56–69. Oxford: Blackwell.
Kant, Immanuel. 1785. *Grundlegung zur metaphysik der sitten*. English edition: Kant, Immanuel. 1996. *Groundwork of the metaphysics of morals* (trans. and ed: Gregor, M. J.). In *Practical philosophy*, 41–108. Cambridge: Cambridge University Press.
Kant, Immanuel. 1788. *Kritik der praktischen vernunft*. English edition: Kant, Immanuel. 1996. *Critique of practical reason* (trans. and ed: Gregor, M. J.). In *Practical philosophy*, 137–271. Cambridge: Cambridge University Press.
Kant, Immanuel. 1797. Über ein vermeintes recht aus menschenliebe zu lügen. English edition: Kant, Immanuel. 1996. On a supposed right to lie from philanthropy (trans. and ed: Gregor, M. J.). In *Practical philosophy,* 609–615. Cambridge: Cambridge University Press.
Kenny, Anthony. 1979. *Aristotle's theory of the will*. New Haven: Yale University Press.
Kohn, Melvin. 1974. Occupational structure and alienation. *American Journal of Sociology* 82:111–130.
Layard, Richard. 2005. *Happiness: Lessons from a new science*. London: Penguin.
Lebus, Bruce. 1990. Moral dilemmas: Why they are hard to solve. *Philosophical Investigations* 13:110–125.
Levi, Isaac. 1967. *Gambling with truth: An essay on induction and the aims of science*. New York: Alfred A. Knopf.
MacIntyre, Alasdair. 1981. *After virtue*. Notre Dame: University of Notre Dame Press.
Mill, John Stuart. 1861. *Utilitarianism*. Rpt. in Collected works of John Stuart Mill, vol. 10, 203–259. Toronto: University of Toronto Press and London: Routledge & Kegan Paul, 1969.
Nettle, Daniel. 2005. *Happiness: The science behind your smile*. Oxford: Oxford University Press.
Nielsen, Kai. 1972. Against moral conservatism. *Ethics* 82:113–124.
Prichard, H. A. 1912. Does moral philosophy rest upon a mistake? Rpt. in *Moral obligation: Essays and lectures*, 1–17. Oxford: Clarendon Press, 1949.
Rawls, John. 1971. *A theory of justice*. Cambridge: The Belknap Press of Harvard University Press.

Richardson, Henry S. 1994. *Practical reasoning about final ends*. Cambridge: Cambridge University Press.
Sartre, Jean-Paul. 1946. L'existentialisme est un humanisme. English edition: Sartre, Jean-Paul. 1956. Existentialism is a humanism. In *Existentialism from Dostoevsky to Sartre,* ed. Walter Kaufmann, 287–311. Cleveland: World.
Schacht, Richard. 1994. *The future of alienation*. Urbana: University of Illinois Press.
Seeman, Melvin. 1959. On the meaning of alienation. *American Sociological Review* 24:783–791.
Sinnott-Armstrong, Walter. 2006. *Moral skepticisms*. Oxford: Oxford University Press.
Spielthenner, Georg. 2007. A logic of practical reasoning. *Acta Analytica* 22:139–153.
Styron, William. 1979. *Sophie's choice*. New York: Bantam.
Von Wright, Georg Henrik. 1963. *The varieties of goodness*. London: Routledge & Kegan Paul.
Von Wright, Georg Henrik. 1983. *Practical reason: Philosophical papers*, vol. 1. Oxford: Basil Blackwell.
Welch, John R. 1991. Reconstructing Aristotle: The practical syllogism. *Philosophia* 21:69–88.
Welch, John R. 1994. Science and ethics: Toward a theory of ethical value. *Journal for General Philosophy of Science* 25:279–292.
Williams, Bernard, and J. J. C. Smart. 1973. *Utilitarianism: For and against*. Cambridge: Cambridge University Press.
Wittgenstein, Ludwig. 1978. *Remarks on the foundations of mathematics*. Revised ed. Cambridge: The MIT Press.
Zimmerman, Michael J. 1996. *The concept of moral obligation*. Cambridge: Cambridge University Press.

Chapter 5
Securing Our Moral Ends

Abstract Chapter 5 concentrates on the teleological stratum of moral discourse. The ends of Kantian, Benthamite, and Frankenian moral theories are the good will, the greatest happiness of the greatest number, and beneficence with justice respectively. Since each of these ends can generate moral advice that differs from that of the others, the relative merits of these ends is a fundamental moral issue. Is there any rational way to choose among them? This chapter argues for an affirmative answer. When ends are expressed in teleological descriptions such as 'The highest good is the greatest happiness of the greatest number', the descriptions can be viewed as hypotheses and thereby confirmed or disconfirmed through hypothetico-deductive reasoning analogous to that employed in the sciences. But when ends are expressed in teleological directives such as 'Act with a good will', choice among directives can be reasonably guided by the comparative decision theory of Chap. 3. Finally, the chapter addresses the higher-order end of morality itself: Can the aim of acting morally be rationally justified? The chapter urges that coherence requires the adoption of 'I ought to be moral' and that coherence in turn can be justified.

5.1 Moral Teleology

Fontenelle, a scientist and literary dandy from the time of Louis XIV, lived to be almost a hundred. Asked what he felt as he approached his final birthday in full health, Fontenelle replied, *Rien, rien du tout ... seulement une certaine difficulté d'être* (quoted in Ortega y Gasset [1936] 1961, p. 201). A certain difficulty of being marks every human life. This difficulty is not that of finding food or shelter or a mate, though these difficulties can be great and, in extreme cases, insuperable. The fundamental difficulty is that of spending time.[1] Time is our primal resource and, like other resources, it is scarce. All men are mortal.

Life can be looked at as a "resource-allocation problem" (Morton 1991, p. 110). Decisions on how to allocate our temporal resources should address the constraints on human action imposed by morality. Some ways of managing time are moral; others are not. Drawing the line between the moral and the immoral requires all three

[1] Marx considered the most fundamental form of economics to be the economics of time ([1939] 1977, p. 362).

J. R. Welch, *Moral Strata,* Theory and Decision Library A:, 49,
DOI 10.1007/978-3-319-08013-0_5,

of our discursive strata: phenomenal, instrumental, and teleological. The teleological stratum has a certain pride of place, however, for it acts as compass for the rest. Instrumental moral discourse is evidently dependent on moral teleology, for means make sense only relative to ends. "Ends are more important than means" (Dyson 1988, p. 215). Phenomenal moral discourse is just as dependent on moral teleology, though the dependence may be less evident. If our ends are consequentialist, we have no need for morally phenomenal descriptions of intentions; but if our ends are Kantian, there is no moral point in describing consequences. Ends determine what is phenomenally relevant.

As Sect. 4.1 observed, the Aristotelian thesis that every action is aimed at a good enables two kinds of critical reflection: instrumental and teleological. Chapter 4 was devoted to instrumentality; the present chapter explores teleology. Just as Chap. 4 began with some general remarks on instrumentality before focusing on moral instrumentality, this chapter will proceed from teleology in general to moral teleology. The two immediately following sections discuss ends in general. Section 5.2 distinguishes different sorts of ends, and Sect. 5.3 reviews different strategies for justifying them. The next five sections then buckle down to moral teleology. Section 5.4 investigates the possibility of justifying moral ends identified by teleological descriptions; Sect. 5.5, those proposed by teleological directives. Section 5.6 then outlines an amplified version of Frankena's mixed deontology, already the subject of decision-theoretic evaluation in Sect. 4.6. Section 5.7 addresses the higher-order end of morality itself: Can the aim of acting morally be rationally justified? The attempt to answer this question leads to a discussion of coherence that occupies Sect. 5.8.

5.2 Distinguishing Ends

If every action is aimed at a good, there must be a staggering number of goods. Experience confirms this, just as it confirms their variety. But despite the number and variety of goods or ends, they can be readily classified in several ways. They can be distinguished by their content, by their relation to other goods or ends, and by their linguistic expression. The following paragraphs briefly treat each in turn.

Ends can be differentiated according to content. Some ends are things; you may act to obtain a jacket or a pet, for instance. Other ends are actions, such as a colleague's attending a meeting. Still other ends are states of affairs. Some states of affairs are meant to be experienced by the agent, such as feeling cool on a hot day, while others are intended to be experienced by others, like a child's financial security. Finally, some ends are compounds of these fundamental types. An athlete might aim to win a bet on a game (a thing), garner votes for most valuable player (a state of affairs), experience the thrill of victory (agent's experience), and gratify fans (others' experience) all at the same time.

How ends are related to other ends is a further basis for distinction. Some ends are instrumental, while others are intrinsic. The end of getting paid may be instrumental

to traveling to Hong Kong, for example, and traveling to Hong Kong may be instrumental to riding the Star Ferry. But the end of riding the Star Ferry may be intrinsic, for it may be valued in itself, not for the sake of something else. Aristotle, Hume, and Kant are each committed to what Robert Audi calls "behavioral foundationalism": the view that "all intentional action is linked by a purposive chain to at least one thing the agent wants for its own sake" (1989, p. 36, 47; cf. 2001, p. 205).

Finally, ends can be classified according to their linguistic expression. Just as instrumental descriptions differ from instrumental directives, as we saw in Sect. 4.2, teleological descriptions are distinct from teleological directives. 'There is no higher end than friendship' is a teleological description. 'Know yourself' is a teleological directive.

5.3 Justifying Ends

Some writers believe that ends (or goals) are neither rational nor irrational; instead, rationality and irrationality are properties of means. Herbert Simon, for example, expounds this instrumental view of reason:

> We see that reason is wholly instrumental. It cannot tell us where to go; at best it can tell us how to get there. It is a gun for hire that can be employed in the service of whatever goals we have, good or bad. (1983, pp. 7–8)

Illustrating his point with reference to Hitler's *Mein Kampf*, Simon adds:

> And so it is not its reasoning for which we must fault *Mein Kampf*, but its alleged facts and its outrageous values….
> And so we learned, by bitter experience and against our first quick judgments, that we could not dismiss Hitler as a madman, for there was method in his madness. His prose met standards of reason neither higher nor lower than we are accustomed to encountering in writing designed to persuade. Reason was not, could not have been our principal shield against Nazism. Our principal shield was contrary factual beliefs and values. (1983, pp. 10–11)

Other writers defend a slightly less anemic conception of reason. Maurice Allais maintains that rationality can impose the minimal requirement of logical consistency on ends:

> It cannot be too strongly emphasized *that there are no criteria for the rationality of ends as such other than the condition of consistency*. Ends are completely arbitrary…. This area is like that of tastes: they are what they are, and differ from one person to the next. ([1953] 1979, p. 70)

Still other writers hold a more robust conception of reason. John Rawls, for instance, affirms that ends may be rationally rejected if their descriptions are meaningless, contradict established truths, or arise from overgeneralization or accidental associations (1971, pp. 419–420). In addition, he mentions temporal principles like the principle of postponement, which recommends not committing to an end "until we have a clear view of the relevant facts," and the principle of continuity, which stipulates "Not only must effects [of planned actions] between [temporal] periods

be taken into account, but substantial swings up and down [from one period to another] are presumably to be avoided" (1971, pp. 420–421).

A comparably robust conception of reason is defended by Larry Laudan, who points out two ways of rationally criticizing goals (1984, pp. 50–62). Though he proposes these tactics for evaluating scientific goals, they are straightforwardly applicable to other goals as well. One maneuver is to reject any goal that is utopian. This could be done in several ways: by showing that, given our understanding of logic or the laws of nature, a goal is demonstrably utopian; by showing that, because a goal cannot be described precisely or unambiguously, it is semantically utopian; or by showing that, because there is no criterion for determining if a goal has been achieved or not, it is epistemically utopian. The second recourse for criticizing goals is, in effect, to harp on the proverb 'Practice what you preach', pointing out any inconsistencies that might exist between explicit goals that are acknowledged but not acted upon and implicit goals that are acted upon but not acknowledged.

Consistent with Laudan's critique of utopian ends, Karin Edvardsson and Sven Ove Hansson assume that goals are set in order to be attained—or at least approximated (2005). From this point of view, "A rational goal is a goal that performs its achievement-inducing function well" (2005, p. 347). In order for a goal to induce achievement, it must both guide and motivate action. Edvardsson and Hansson claim that goals must meet certain structural criteria if they are to guide and motivate. These criteria are derived from epistemic, ability-related, and volitional considerations. Epistemic criteria include precision, which permits the agent to know what the goal is, and evaluability, which allows the agent to gauge the effectiveness of her actions in achieving it. An ability-related criterion is attainability—or approachability, at least, in case some goals can only be approximated. Finally, a volitional criterion is motivational power: the power to motivate the appropriate kind of action. Rational goals, therefore, should be precise, evaluable, approachable, and motivating.

Edvardsson and Hansson observe that these four criteria are not mutually independent; in order for a goal to be evaluable, for example, it must be precise (2005, p. 350). They also remark that logical consistency is implied by their criteria, for precision and attainability (the strong form of approachability) presuppose it (2005, p. 350).

Robert Audi distinguishes instrumental and intrinsic desires (2001, Chaps. 3 and 4). Though he is directly concerned with desires rather than ends, his remarks are easily extended to ends because ends are special sorts of desires: desires that are meant to be acted on. His views on instrumental desires, intrinsic desires, conflicting intrinsic desires, and justification of intrinsic desires are briefly summarized here.

The instrumental case is relatively simple. Section 5.2 mentioned the instrumental end of getting paid in order to ride the Star Ferry in Hong Kong. The two ends are connected by an instrumental belief. The Audian configuration of these three elements has been mapped out already in the material on practical inference in Sect. 4.5.2:

I intend to ride the Star Ferry in Hong Kong.
Unless I get paid, I cannot ride the Star Ferry in Hong Kong.
Therefore I must get paid.

Whether the end of getting paid is justified depends on whether the end of riding the Star Ferry is.

That brings us to the intrinsic case. Audi claims that an intrinsic desire can be mistaken if based on false beliefs or presuppositions about that which is desired (2001, p. 88). The point applies to ends as well. If the Star Ferry is no longer in operation, for instance, my aim of riding it would be mistaken. The mistake would be objective as well as subjective if I know about the end of operations; objective but not subjective if I do not.

One intrinsic desire can rationally defeat another, according to Audi, if the agent realizes that satisfying one desire would thwart a second, stronger one (2001, p. 71, 79). Consider the following remarks by Derek Jarman, a painter, poet, and filmmaker who died of AIDS in 1994:

> [E]veryone should have the type of sex they want to have, and if they decide they don't want to have safe sex, it's their funeral, quite literally. You can't start being moralistic about it. A friend of mine had sex with someone whom he knew was HIV-positive and didn't use any protection. Now they're both HIV-positive. I always thought this was mad, for anyone to deliberately have the disease. On the other hand, you can't really say it was wrong, if this was what they wanted to do. (Shulman 1993, p. 60)

Jarman seems to be denying that moral criticism of informed, unprotected sex such as that engaged in by his friend is legitimate. For anyone sympathetic to the Kantian view that we have moral duties to ourselves as well as to others, this is a questionable claim. But my point concerns the rationality, not the morality, of such behavior. If Jarman's friend has an overriding commitment to certain ends—political, athletic, literary, or whatever—the end of practicing unsafe sex with someone known to be HIV-positive would be irrational if the friend realizes that the sexual end jeopardizes the overriding end. Regardless of whether the agent has an overriding end, the end of unsafe sex with someone known to be HIV-positive is shortsighted in normal circumstances because it permits lesser but immediate ends to outweigh greater but mediate ones.[2] Such actions would violate what Sidgwick called the principle of "impartial concern for all parts of our conscious life," which implies "that a smaller present good is not to be preferred to a greater future good" (1907, III.xiii.3, p. 381). Consequently, we can say that such sexual practices are irrational in at least some situations.

Audi takes unjustified desires to be highly significant, for if a desire "can be ill-grounded, one would expect that it can be well-grounded when it is free of the relevant vitiating basis" (2001, p. 88). Intrinsic desires, in Audi's view, can be defeasibly grounded in experience, "above all [in] those experiential qualities intrinsic to pleasure and pain and to the happy exercise of our capacities, including conscious

[2] Cf. psychologists Gilbert and Wilson on 'miswanting': to want things we would not like if we manage to get them (Irvine 2006, p. 104).

states of rewarding contemplation, whether aesthetic, intellectual, religious, or of any other kind" (2001, pp. 100–101). An intrinsic desire to listen to Schubert, say, is prima facie rational (or not) if relevantly similar experiences have been worthwhile (or not) from the perspective of the agent. The point carries over directly to ends; an intrinsic end to have dinner with a friend can be prima facie rational or not depending on the agent's experiences.

The balance of this chapter sides with those like Rawls, Laudan, Edvardsson and Hansson, and Audi whose conceptions of reason are sufficiently robust to support the claim that ends can be rationally justified. Unlike the just-mentioned authors, however, I will attempt to support this claim by drawing on some ideas from confirmation theory and decision theory. We will initially treat ends identified by teleological descriptions, then ends recommended by teleological directives.

5.4 Teleological Moral Descriptions

From this point on we will narrow our focus to moral teleology. I propose to start with single sentences in the form of descriptions. Two observations by Dewey will serve as points of departure: "all moral judgment is experimental and subject to revision by its issue" ([1922] 1983, p. 194); and moral principles "require verification by the event," existing "as hypotheses with which to experiment" ([1922] 1983, pp. 164–165). In the spirit of these remarks, I propose to consider teleological descriptions as hypotheses subject to confirmation or disconfirmation. To say that a hypothesis is confirmed or disconfirmed is not to say that it is proved or disproved. It is to say that the evidence raises or lowers the hypothesis' credibility to some—not necessarily great—degree.

'Confirmation', in its usual sense, is increase of probability due to new evidence. But this sense of 'confirmation' can be generalized. Section 3.3.2.4 pointed out that plausibilities, unlike probabilities, need not be numeric; hence plausibility can be used in many cases where probability cannot. For example, John Woods and Dov Gabbay have studied the use of plausibility in legal contexts. Referring to judgments of the so-called "balance of probabilities," Woods remarks that

> probability is nothing that any probability theorist to date has ever turned his mind to. This is chiefly because—or so we [Woods and Gabbay] think—that in these contexts it is not a probability measure that is in play, but rather a plausibility measure, never mind the name that lawyers give it. If this is right, there is much about legal probability that demands working out in a logic of plausibility. (Pigozzi 2009, p. 4)

Just as many kinds of legal judgment adhere to a logic of plausibility, so do many kinds of moral judgment. In what follows, then, confirmation will be understood as increase of plausibility due to new evidence. This expanded sense of 'confirmation' includes the increase-of-probability sense as a special case.

To flesh out the notion that teleological descriptions can be confirmed, recall the standard notion of scientific method as hypothetico-deductive. Let us review some simple patterns of confirmation and disconfirmation within this

hypothetico-deductive framework. Modus tollens provides a simple model of disconfirmation. If a hypothesis *h* has a logical consequence *c* and *c* turns out to be untrue, then *h* is strongly disconfirmed:

If *h*, then *c*.
$-c$.
Therefore $-h$.

Though we are accustomed to say that '*h*' is falsified, the weaker claim that '*h*' is strongly disconfirmed is better advised. For there is some possibility, however slight, that modus tollens could turn out to be invalid. After all, even the validity of modus ponens is under fire (McGee 1985). So if modus tollens should turn out to be invalid, the premises of an argument of that form could be true and the conclusion false, in which case '*h*' would be true. Nevertheless, the evidence of the premises makes the plausibility that '*h*' is false extremely high.

Alternatively, if a hypothesis has a logical consequence and the consequence turns out to be true, then the hypothesis is confirmed. The inference takes the following form:

If *h*, then *c*.
c.
Therefore *h*.

This form is invalid, of course; it incurs the fallacy of affirming the consequent. But the truth of '*c*' increases the plausibility that '*h*' is true. In the probabilistic special case, the truth of '*c*' increases the probability that '*h*' is true unless the prior probability of '*h*' is 0 or that of '*c*' is 1 (Jeffrey 1992, pp. 57–58). George Polya calls this form the "fundamental inductive pattern" (1954, vol. II, p. 4; cf. Franklin 2001, pp. 329–330).

Note that the crucial difference between affirming the consequent and modus tollens is the second premise: '*c*' for affirming the consequent, '$-c$' for modus tollens. Thus whether '*h*' is confirmed or disconfirmed hinges on the choice between '*c*' and '$-c$'. Where our aim is cognitive, this choice must be based on the plausibilities of the two sentences relative to the evidence. Hence '*h*' is confirmed only if $\pi(c \mid e) > \pi(-c \mid e)$ and disconfirmed only if $\pi(-c \mid e) > \pi(c \mid e)$.

In summoning these elementary patterns to the context of morality, my point is not that moral teleology is or can be made scientific. The point is the structural parallel between reasoning in science and reasoning in morals. Philosophies of science that do not fall prey to scientism are often open to this insight. Theo Kuipers, for example, remarks "A test for a hypothesis may be experimental or natural. That is, a test may be an experiment, an active intervention in nature or culture, but it may also concern the passive registration of what is or was the case, or what happens or has happened" (2000, p. 19). Good philosophy, he notes, conforms to his Principle of Testability, which calls for theories with implications that "can be tested for their truth-value by way of observation" (2000, p. 123, 131). Philosophical observation may not be of empirical events, of course, but of cases that are handled satisfactorily or not and adequacy conditions that are satisfied or not.

What I suggest, then, is that teleological hypotheses are subject to natural tests. As an initial example, take the teleological description that the highest good is the greatest misery of the greatest number. If the moral skeptic were right about moral teleology, we would have no choice but to admit this description as a rival to the best of Kant and Mill. But consider an argument based upon it:

> The highest good is the greatest misery of the greatest number.
> Anything that contributes to the highest good is good.
> Famine contributes to the greatest misery of the greatest number.
> Therefore famine is good.

Contrast this conclusion with its contradictory: it is not the case that famine is good. On which statement would we choose to bet? We would bet on the conclusion's contradictory, I submit, for two sorts of reasons. The first is the biological evidence that famine is not good for individual organisms: all forms of life generally attempt to avoid it or—if this is not possible—to relieve it. The second is the sociopolitical evidence that famine is not good for human collectivities: human social planners almost invariably attempt to prevent famine, not induce it. Hence the plausibility of 'It is not the case that famine is good' on this evidence is far superior to the plausibility of 'Famine is good' on the same evidence. 'Famine is good' is therefore disconfirmed. By modus tollens, then, the conjunction of premises that implies 'Famine is good' is also disconfirmed. Thus the evidence suggests that at least one of these premises is false. Since the initial teleological description is by far the likeliest culprit, the teleological description is disconfirmed as well.

Bentham's rival teleological claim is that the highest good is the greatest happiness of the greatest number. We might use it to produce a rival of the miserable argument of the preceding paragraph:

> The highest good is the greatest happiness of the greatest number.
> Anything that contributes to the highest good is good.
> Adequate food contributes to the greatest happiness of the greatest number.
> Therefore adequate food is good.

That something is instrumentally good—and not just perceived to be instrumentally good—is borne out by the results of obtaining it. The results of adequate food are demonstrated across the plant and animal kingdoms, and they are positive in the extreme. The fact that an ascetic may consciously seek to be poorly fed, at least for a time, is no exception to the rule, for the ascetic consciously sacrifices one good in order to achieve another. Hence the biological evidence in favor of 'Adequate food is good' and against its contradictory is overwhelming. This disparity sets off a chain reaction of three confirmations. 'Adequate food is good' is confirmed, first of all. Affirming the consequent supplies a second confirmation: the conjunction of premises that implies 'Adequate food is good'. Finally, since there is some evidence that all three premises are true, there is some evidence that the first premise in particular is true. Hence Bentham's teleological description is confirmed, though the degree of confirmation may not be high.

Notice that this final confirmation does not play out like the disconfirmation of 'The highest good is the greatest misery of the greatest number' two paragraphs above. To single out this teleological description as the likely cause of a false conjunction, we had to rely on its meaning; the disconfirmation was semantic. But the confirmation of the greatest happiness premise is not a semantic matter; it follows syntactically from the prior confirmation of conjoined premises.

Many variants of the arguments sketched in the foregoing paragraphs are possible, of course. We might confirm Bentham's view by building on the idea that pleasure contributes to happiness:

> If the highest good is the greatest happiness of the greatest number, then pleasure is good.
> Pleasure is good.
> Therefore the highest good is the greatest happiness of the greatest number.

And we might disconfirm the miserable rival to Bentham's view by arguing:

> If the highest good is the greatest misery of the greatest number, then pain is good.
> Pain is not good.
> Therefore the highest good is not the greatest misery of the greatest number.

Although the sample confirmations and disconfirmations in this section are admittedly not very exciting, I have devoted some space to them because they seem to be seldom noticed. Successive sections will raise the degree of teleological difficulty.

5.5 Teleological Moral Directives

Whether an end finds linguistic expression in a description or a directive is often a matter of stylistic convenience. A Kantian might say "The highest good is a good will" or "Always act with a good will." But Sect. 5.4 has focused on teleological moral descriptions, so the present section will treat teleological moral directives.

I will approach the topic by returning to the problem of theory choice, for there is more than one legitimate way to evaluate theories. A moral theory might be favored as a guide in a specific situation, which was the instrumental approach to evaluating moral theories as they apply to Sophie's choice (Sects. 4.6.1–4.6.3). But a moral theory might also be valued for the suitability of the end or ends it recommends. This teleological point of view informs the present section. I propose that we consider the teleological directives characteristic of the moral theories of Sect. 4.6: the Kantian directive to act with a good will, the utilitarian directive to act for the greatest happiness of the greatest number, and the Frankenian directive to act justly and beneficently. Are there grounds for adopting one of these directives rather than the others?

Each of the just-cited directives claims that we ought to act in a determinate way. But the 'ought' of practical inference is instrumental, as I argued in Sect. 4.5.2.1.

That is, if we conclude that we ought to do something, we ought to do it to achieve some end. Might a teleological 'ought' then be instrumental as well? The answer is affirmative. We have noted repeatedly that an end may be a means to some further end. But what might be the further end or ends at which the Kantian, Benthamite, and Frankenian directives are aimed? This part, at least, is easy: each of these directives is aimed at the same overarching end: to act morally. Kant, Bentham, and Frankena think that we should act in the ways they prescribe in order to be moral. But now comes the hard part. How could we ever decide which directive provides better advice about how to act morally?

The task is admittedly complicated by the vagueness of 'acting morally'. Even morality's staunchest defenders would have to admit that the end of acting morally compares unfavorably in clarity with the end of reaching Pike's Peak by sundown. But 'reaching Pike's Peak by sundown' is not completely clear, and 'acting morally' is not completely vague. Roy Sorensen has argued that all vague predicates are cognitive since vagueness requires borderline cases, borderline cases imply clear cases, and clear cases have truth values (1990). And we do have reason to think that adhering to the directive 'Maim as many people as possible', for example, does not conduce to the end of acting morally. So let us see whether the end of morality is sufficiently clear to evaluate our three directives for their effectiveness in furthering moral action.

Section 4.6 deployed a comparative version of decision theory to choose a theory on instrumental grounds, and I suggest that we try a similar line to select a teleological directive. The idea is to choose a teleological directive that maximizes plausibilistic expectation. Note that the attempt to maximize plausibilistic expectation does not automatically tip the scales in favor of consequentialist approaches like Bentham's. Charles Larmore makes an analogous point about maximizing expected utility:

> In fact, within the moral domain itself it [maximizing expected utility] does not, strictly speaking, privilege "consequentialist" over "deontological" ways of reasoning, despite the common perception of an elective affinity between maximization and consequentialism (which holds that one is to act so as to bring about the most good overall). For the deontologist maximizes too when he conforms as best he can to the moral principle he holds supreme, which is that one is to respect certain rights, whatever the consequences…. (2008, p. 102)

To choose among our teleological directives on decision-theoretic grounds, we must be able to specify the relevant acts, states, and outcomes. Adopting a moral directive belongs to the category of mental acts; like other acts, mental acts are localized in space and time. In this context, three possible mental acts are under consideration: adopt the Kantian directive (d_k), adopt the Benthamite directive (d_b), or adopt the Frankenian directive (d_f) as a guide to acting morally. States can be understood as propositions, some of which describe abstract features of the world (Sect. 3.3.1.2). Instances of such features are the teleological states posited by our three theories: the highest good is the good will (s_k), the highest good is the greatest happiness of the greatest number (s_b), and the highest goods are justice and beneficence (s_f),

Table 5.1 Choosing a teleological directive

	s_k	s_b	s_f
d_k	$+i$	$-i$	$-i$
d_b	$-i$	$+i$	$-i$
d_f	$-i$	$-i$	$+i$

respectively. Since these states are not jointly exhaustive, no one of them need be true. But one can be closer to the truth than the others. Suppose, for example, that the Kantian state is more truthlike than its rivals and that we choose the Kantian directive. The result would be good advice—some degree of action-guiding information ($+i$). On the other hand, if the Kantian state is more truthlike and we choose the Benthamite or Frankenian directive, the result would be some degree of misinformation about how to act ($-i$). We can summarize the decision-theoretic situation by representing the acts of choosing a teleological directive, the relevant teleological states, and the resulting information outcomes as in Table 5.1.

5.5.1 Plausibility

To estimate the plausibilistic expectation associated with our three directives, we must consider the state plausibilities and outcome utilities in each case. Fortunately, much of the spadework for the relevant state plausibilities has already been done. Section 5.4 indicated how teleological descriptions like ‘The highest good is the greatest happiness of the greatest number’ can be confirmed or disconfirmed. Section 4.6.1 cited some standard scenarios for disconfirming Bentham’s act-utilitarianism in which utility maximization would require executing an innocent person, killing a healthy patient, making an insincere promise, etc. (The objection that these scenarios depend on an uncritical acceptance of common morality is addressed in Sect. 6.2.) For the Kantian state, Sect. 4.5.4 made the case that moral action is no less valuable than good will, and Sect. 4.6.1 argued that consequences must be part of moral deliberation. These arguments, I submit, disconfirm the Kantian description ‘The highest good is the good will’. By contrast, the consequential focus of Frankena’s beneficence corrects Kant’s consequential emptiness, and Frankena’s emphasis on justice remedies the act-utilitarian tendency to let utility steamroll justice. Additional confirmation of a more practical sort can be gleaned from the Belmont Report, an established standard in medical ethics (The National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research 1979). Commissioned by the United States Congress to identify the basic ethical principles governing the use of human research subjects, the Report lists beneficence, justice, and respect for persons. Though the Report’s inclusion of respect for persons could be taken to confirm at least some elements of Kantian ethics, the inclusion of the consequentialist principle of beneficence is plainly anti-Kantian. To some extent, then, the Frankenian description ‘The highest goods are justice and beneficence’ is confirmed. The result is that the comparative state plausibilities for

entire theories observed in Sect. 4.6.1 are mirrored here by comparative state plausibilities for teleological parts:

$$\pi(s_f|e) > \pi(s_k|e) > \pi(s_b|e).$$

5.5.2 Utility

What, then, about the outcome utilities? In keeping with the normative decision theory deployed in this work, I have argued that information should be considered as an outcome of cognitive choice and that utility should be proportional to information. Much as in Sect. 4.6.2, let us consider the choice among the directives in this section as a purely cognitive decision, that is, a decision based wholly on the directives' effectiveness in providing action-guiding information. Note, however, that the focus is different from that of Sect. 4.6.2. There we compared the action-guiding information provided by our three theories in a particular case: Sophie's choice. Here we are dealing with the action-guiding information provided by one part of those theories—their teleological directives—in the general case.

There are at least two ways to proceed. One is to key on the fact that each of our directives can falter in the attempt to provide action-guiding information. Whether utilitarianism has any implications for action at all has been doubted (Hausman 1991). But even if this is too extreme, we can say at least that Bentham's act-utilitarianism guides less effectively than Kantian and Frankenian theory in Sophie's choice, as we have seen in Sect. 4.6.2. More generally, what the greatest happiness of the greatest number requires is often most unclear. Even if the hedonistic calculus is pressed into service to implement the utilitarian directive, it cannot be relied upon whenever accurate numbers for intensity, duration, certainty, remoteness, fecundity, purity, and extent are unavailable—and this is by far the usual case.

Yet the categorical imperative in the formulation of universal law fails to guide when different people at the same time—or the same person at different times—can will to universalize different maxims. There are analogous problems with the formulation of the end in itself. Kant appealed to the formulation of the end in itself to defend retributive punishment, but the requirement to treat others as ends can be interpreted to permit only rehabilitative punishment (cf. Scheid 1983, pp. 272–277). This is a difference that makes a difference, for variant views of the end in itself can result in very different punishments for the same crime.

Finally, as Frankena himself points out, his theory can fail to guide in cases of conflict between justice and beneficence (1973, pp. 52–53). Though Frankena usually places justice before beneficence, the priorities might—or might not—be reversed if a small injustice would avoid a great evil or obtain a great good, as indicated in Sect. 4.6.1.

Thus all three directives are vague (cf. Flanagan 1986, pp. 57–58). To be able to compare their degrees of vagueness would be useful; unfortunately, I know of no effective way to do so. But there is another way to proceed. In dealing with Sophie's choice, we noted that the quantity of information i could be straightforwardly

expressed by relating the number of eliminated options e to the total number of options t: $i = e/t$. Though our interests in this section are ultimately comparative, I propose to work with this quantitative measure as long as we can muster enough data to do so.

Consider first the total number of options t. The life of a human being can be represented as a world line with the acts she performs plotted along it. Some of these acts can be mental acts of adopting a teleological moral directive. These special acts can be viewed as occupying a finite number m of positions along the world line. In addition, there is a finite number n of teleological moral directives—'Always act with a good will', for example—that constitute the agent's moral horizon. I will assume for simplicity that n is constant throughout the agent's lifetime. Though this assumption is probably false for many agents, greater realism would not affect the overall shape of the argument. At each of the m points where the agent adopts a teleological moral directive, then, her world line can branch in n directions; the agent might adopt any one of the n directives under consideration. The total number of options for teleological moral directives in the agent's lifetime is therefore $mn = t$.[3]

Turn now to the number of options e eliminated by each of our directives. This number is evidently quite large. At each of the m decision points, all but one of the n options will be eliminated: $n-1$ of them. During the agent's lifetime, therefore, the number of options each directive will eliminate is $m(n-1) = e$.

Consequently, the information provided by our directives would be expressed by Eq. 5.1.

$$i = \frac{m(n-1)}{mn} = \frac{n-1}{n}. \tag{5.1}$$

Because this ratio is the same for all three directives, all three provide the same amount of information. Since we have assumed that our choice among directives is a purely cognitive decision and that utility is proportional to information, the utilities of these information outcomes should be equal as well:

$$u(i_f) = u(i_k) = u(i_b).$$

5.5.3 Plausibilistic Expectation

To calculate the plausibilistic expectation of our three choices, we can apply *PE* (Sect. 3.3.2.7, Eq. 3.4). Because the comparative decision theory employed in this work focuses on binary choice, a field with options o_1, o_2, and o_3 requires two comparisons: o_1 with o_2, say, and the winner with o_3. One way to proceed, then, is to begin by comparing the plausibilistic expectations of adopting the Kantian and Benthamite directives. Since the plausibility of the Kantian state is greater than that of the Benthamite state while the utilities of the information outcomes are equal,

[3] The issue of alternative sets of options is treated in Sect. 6.2.

the plausibilistic expectation of the Kantian directive is greater than that of the Benthamite directive. This is an instance of case 6 from Table 3.4 (Sect. 3.5.1): greater plausibility for the Kantian directive, equal utility, therefore the Kantian directive. The second step is therefore the comparison of the Kantian and Frankenian options. The plausibility of the Frankenian state is greater than that of the Kantian state, but the utilities of their information outcomes are equal. Hence the plausibilistic expectation of the Frankenian directive is greater than that of the Kantian directive—another instance of case 6 from Table 3.4: greater plausibility for the Frankenian directive, equal utility, hence the Frankenian directive. Combining the results of both comparisons gives us the following comparative ranking:

$$PE(f) > PE(k) > PE(b).$$

This ranking should not be confused with the parallel ranking of Sect. 4.6.3. There we ranked the plausibilistic expectation of three theories; here we have ranked the plausibilistic expectation of three directives from those theories. Nevertheless, the two results are complementary. The superiority of Frankenian theory asserted in Sect. 4.6.3 was based on instrumental grounds: its efficacy as a moral guide in the unique circumstances of Sophie's choice. Hence it is entirely possible that another theory might prove superior in another set of circumstances. But the argument for the superiority of Frankenian teleology strengthens the presumption in favor of Frankenian theory considerably, for the argument we have just rehearsed is entirely general. Since the definitive differences among moral theories are at the teleological level, any moral theory with teleological advantages has much in its favor indeed.

The results of this section are limited to the three directives under consideration. It is possible, therefore, that some version of moral teleology not considered here is decision-theoretically superior to Frankenian teleology. In addition, it is quite possible that the arguments for the comparative plausibilities and utilities on which this section's expectation ranking is based are mistaken or variously inadequate. These are admittedly deep waters, and we should proceed with a certain amount of moderate Humean skepticism. But if others' judgments of comparative plausibility and utility should differ from mine, the comparative decision theory employed in this section could still be applied with a different set of inputs.

Though we are not directly concerned with theory choice in this section, our discussion has uncovered a subterranean connection between two of the four approaches to theory choice described in Sect. 3.2: probabilism and decision theory. Both of the foregoing comparisons of teleological directives are based on unequal state plausibilities and equal outcome utilities. In these situations, as we have seen, the decisive factor is plausibility. The structure of these situations is frequently instantiated in purely cognitive decision making. In these cases, when the information offered by rival theories is equal, or roughly so, the decision is made on plausibilistic grounds. A venerable example is cited by Moses Maimonides in his discussion of whether or not the world is eternal. Maimonides says he has employed the method described long before him by Alexander of Aphrodisias:

> For Alexander has explained that in every case in which no demonstration is possible, the two contrary opinions with regard to the matter in question should be posited as hypotheses, and it should be seen what doubts attach to each of them: the one to which fewer doubts attach should be believed. Alexander says that things are thus [or: that he conducts things thus] with respect to all the opinions regarding the divine that Aristotle sets forth and regarding which no demonstration is possible…. We have acted in this way when it was to our own mind established as true that, regarding the question whether the heavens are generated or eternal, neither of the two contrary opinions could be demonstrated. ([1190] 1963, II.22, p. 320)

In treating the issue of the world's eternity as a purely cognitive matter, Maimonides recommends the solution that carries less doubt. This, of course, is the more plausible solution. Since probability is a special case of plausibility, probabilism turns out not to be an independent approach to theory choice. It is a special case of decision-theoretic choice.

5.6 Amplifying Mixed Deontologism

Though I have been campaigning in favor of Frankenian teleology, bringing the directives of justice and beneficence to bear on concrete cases does present a problem: Frankena's specifications of what his principles require leave much to be desired. For example, he breaks the principle of beneficence down into four subsidiary principles that can be summarized as follows (1973, p. 47):

1. Do not cause evil.
2. Prevent evil.
3. Remove evil.
4. Do good.

Though Frankena prefaces his discussion of utility and beneficence by focusing on "nonmoral good and evil" (1973, p. 35), he says little else about what these nonmoral qualities are.

I would like to propose a partial remedy in the form of a hybridized version of mixed deontologism. While the Frankenian principles of justice and beneficence constitute the basic framework, the principle of beneficence is hybridized by joining its secondary principles on evil ('Do not cause evil', 'Prevent evil', and 'Remove evil') to Bernard Gert's concept of nonmoral evil. In Gert's treatment, death, pain, disability, loss of freedom, and loss of pleasure are all nonmoral evils (1998, p. 90). The three secondary principles of beneficence, which we recall are prima facie, then ramify as follows:

Do not cause death.
Do not cause pain.
Do not cause disability.
Do not cause loss of freedom.
Do not cause loss of pleasure.

Prevent death.
Prevent pain.
Prevent disability.
Prevent loss of freedom.
Prevent loss of pleasure.

Remove pain.
Remove disability.
Remove loss of freedom.
Remove loss of pleasure.

The result, I suggest, is an ethical theory that is far more concrete and explicit than Frankenian ethics in its original form. But the comparative decision-theoretic ranking of Kantian, Benthamite, and Frankenian teleological directives carried out above does not depend on this modified version of mixed deontologism.

5.7 Morality as an End

The moral directive 'Don't harm others' enjoins non-maleficence as an end and is therefore teleological. But since non-maleficence can be a means to the higher end of acting morally, 'Don't harm others' is morally instrumental as well. This is another instance of a pattern we noticed in Sects. 4.3 and 5.2: some ends are simultaneously means to higher ends. Hence some moral sentences can be seen as borderline between moral teleology and moral instrumentality.

What are the prospects for justifying directives like these in one form or another? Instrumental descriptions embedded in enthymematic practical inferences can justify instrumental directives, as we saw in Sects. 4.4–4.5.1. Hence we might attempt a similar approach here. We might construct a practical inference around the idea that the part of non-maleficence that is morally required (not supererogatory) is a necessary condition for morality. That is, a person who violates the requirements of non-maleficence is immoral to the extent of her failure. A simple example of such an inference is:

I want to be moral.
Unless I am non-maleficent, I will not be moral.
Therefore I must be non-maleficent.

This is an instrumental justification of 'Don't harm others'. But the justification depends on an established desire to be moral, and since the premise that expresses this desire is unlikely to be true of the sort of person who would challenge 'Don't harm others', the justification is unlikely to persuade the unpersuaded. On the other hand, the sort of person the first premise is likely to be true of is unlikely to need a justification of 'Don't harm others'.

For these reasons I want to explore an alternate route. This route requires replacing the higher-order description in the first premise of the previous inference with a higher-order directive:

I ought to be moral.
Unless I am non-maleficent, I will not be moral.
Therefore I must be non-maleficent.

This inference makes whether I should be non-maleficent depend on whether I should be moral. But that runs us smack into the monster with many heads: Why be moral?

We notice immediately that the question 'Why be moral?' has collective and individual senses. Its collective sense is explicit in 'Why should human beings be moral?'; its individual sense, in 'Why should I be moral?'. Here we are concerned with the individual sense, which challenges the first premise of the previous paragraph's justification of non-maleficence.

We also notice that the question appears to rule out a moral reason for being moral. It is asked from the outside looking in; that is, it contemplates morality from an external point of view, questioning it as a whole. Thus the question asks for a nonmoral reason for being moral.

So understood, the question 'Why should I be moral?' has been thought illegitimate. H. A. Prichard argued famously that it rests on a mistake: "the mistake of supposing the possibility of proving what can only be apprehended directly by an act of moral thinking" ([1912] 1949, p. 16). Charles Larmore distinguishes the everyday sense of 'Why should I be moral?', which asks "whether we must really do some particular morally required action," from the traditional philosophical sense, which wonders "whether to heed moral requirements at all—as though only in this way could a truly adequate answer to the initial worry be found" (2008, p. 88). Like Prichard, Larmore then attacks the question in its traditional sense:

> Fueling the traditional approach, I believe, is a failure to recognize that morality forms a realm of irreducible value. We cannot explain why it should matter to us by appealing to interests and motivations that are presumably more basic. In that, I agree with the moral nihilist, but I am no nihilist myself, since my position is that moral reasons possess an intrinsic authority. Morality speaks for itself. This truth has too often been missed or obscured by philosophical theory, and not least by the ethics of autonomy, which draws its inspiration from Kant and aims to ground the moral point of view in a conception of human freedom as self-legislation. All the more appropriate, it seems to me, to reverse the terms and to speak instead of *the autonomy of morality*. There is no way to reason ourselves into an appreciation of moral value from some standpoint outside it. Morality only makes sense in its own terms. (2008, p. 88)

Both Prichard and Larmore claim that there is no external standpoint from which we can reason our way into morality. This may very well be so, but such claims are only as good as the search for such a standpoint that presumably precedes the claims. I, for one, would like to extend the search a little further. The remainder of this section is an attempt to do so.

One way of answering the question 'Why should I be moral?' would be to present a practical inference connecting morality with some nonmoral good *x*. The inference might take the following form:

I want *x*.
Unless I am moral, I will not obtain *x*.
Therefore I ought to be moral.

An answer of this type would conform to von Wright's interpretation of 'ought', summarized already in Sect. 4.5.2.1. The game would then reduce to finding a suitable value for '*x*'.

Let us review some standard options. Section 1.4 cited Sidgwick's systematization of ultimate reasons for action: happiness, whether individually or collectively conceived, and self-perfection. This would give us three possible values for '*x*': egoistic happiness, utilitarian happiness, and perfectionist self-fulfillment.

The egoistic approach is the oldest by far. "Egoism in this sense was assumed in the whole ethical controversy of ancient Greece; that is, it was assumed on all sides that a rational individual would make the pursuit of his own good his supreme aim," as Sidgwick remarks (1907, I.vii.1, pp. 91–92). Plato, for example, contends that justice (in the wide Greek sense that is roughly synonymous with morality) is necessary for happiness (*Republic* 576b–580c). The egoistic answer has a grave problem, however: morality does not seem to be universally necessary for individual happiness—not on this earth, at any rate. Kant thought it morally necessary to assume the existence of God and immortality so that imbalances of moral merit and happiness could be redressed. Similarly, Hindus understand karma to extend across lifetimes.[4]

The classical egoistic justification of morality may attempt to prove too much. This, at least, would appear to be the assessment of David Gauthier, who is ethical egoism's best-known contemporary exponent. Gauthier's *Morals by Agreement* (1986) acquiesces in a certain shrinkage of egoism. The point of morality shrinks, in effect, for its point is not happiness, according to Gauthier, but a part of happiness: the benefits of cooperation. Morality shrinks as well. Locke's remark, "an Hobbist… will not easily admit a great many plain duties of morality," applies to Gauthier's extrapolation of Hobbes' egoism (1986, p. 17, 179, 268). Impartiality and equality have no fundamental place in Gauthier's theory (1986, p. 17, 270), and some forms of coercion and exploitation are permitted (1986, p. 17, 232). Since contractarian morals are motivated by mutual advantage, only contributing members of society can ratify the moral contract and reap its benefits. "Animals, the unborn, the congenitally handicapped and defective, fall beyond the pale of a morality tied to mutuality" (1986, p. 268). So do "people overseas" under certain conditions (Braybrooke 1987, p. 756), strangers in general, and the weak (Larmore 2008, pp. 98–99). Not surprisingly, whether morality is rational depends heavily on circumstances. Gauthier approves of Hume's advice that a person who falls into "the society of ruffians" should "consult the dictates of self-preservation alone"

[4] I am indebted to Jawara Sanford for this last observation.

(1986, p. 181). At best, therefore, *Morals by Agreement* shows that part of happiness requires part of morality part of the time.

To wonder whether some further justification of morality could be found is understandable, therefore. Suppose we consider the other two standard approaches. Initial premises tantamount to 'I want collective happiness' and 'I want self-perfection' would be false of many people, and those they would be false of would be those most in need of a justification of 'I ought to be moral'. Considerations like these led Philippa Foot to argue that morality is a system of hypothetical imperatives (1972). Hypothetical moral imperatives suitable to the present context would have antecedents of the form 'If I want collective happiness…' or 'If I want self-perfection…'. From this perspective, then, reason would permit morality but not require it. But reason would also permit Hume to prefer the destruction of the world to the scratching of his finger, just as he claimed ([1739] 1973, 2.3.3.6, p. 416). These hypothetical justifications suffice for those with the requisite desires. But the fact is that they remain precariously hypothetical.[5]

That is the incentive for trying out yet another approach. This new approach is different in kind from the egoistic, utilitarian, and perfectionist alternatives just sketched. Instead of trying to show that being moral is necessary to obtain a non-moral good, I will argue that accepting 'I ought to be moral' is necessary for a non-moral good. The argument takes a linguistic turn, driven by this volume's focus on the phenomenal, instrumental, and teleological strata of moral discourse.

The argument builds on two familiar ideas: every action is aimed at a good (Sect. 4.1), and accepting 'I ought to be moral' is a mental act (cf. Sect. 5.5). If we put these ideas together, we can infer that accepting 'I ought to be moral' is aimed at a good. But what good? The answer may depend on individual psychology: one person could accept 'I ought to be moral' for one reason and another for a different reason. But an answer that does not depend entirely on individual psychology would be possible if there turns out to be a good that could motivate acceptance for any rational person.

I want to explore this possibility by recalling Laudan's hierarchical and reticulated models of scientific rationality as outlined in Sect. 1.4. The hierarchical model requires justification to flow invariably downward: teleological discourse justifies instrumental discourse, and instrumental discourse justifies phenomenal discourse. But the reticulated model permits bidirectional justification, upward from phenomena to instrumentality to teleology as well as in the opposite direction. The reticulated model is not limited to scientific rationality, as we noticed in Sect. 1.4. This volume has adapted its levels of scientific discourse to moral discourse, and the following paragraph draws on the reticulated model to justify moral teleology.

[5] These assessments of attempts to justify morality in terms of egoistic happiness, utilitarian happiness, and perfectionist self-fulfillment intersect with work by J. L. Mackie. Though Mackie takes morality to be necessary for human well-being in general, he asserts that the rational calculation of long-term self-interest requires only that some people be moral—not all (1977, pp. 107–124, 189–192).

The guiding idea is that moral teleology can be justified by appealing to moral phenomena and instrumentality.

Wide reflective equilibrium as outlined in Sect. 1.4 is traditional in its emphasis on coherence but novel in stressing coherence among morally phenomenal, instrumental, and teleological discourse together with background theories. What I propose for consideration is a justification of accepting 'I ought to be moral' in terms of wide reflective equilibrium. Put as simply as possible, the argument would run along the following lines:

> I want wide reflective equilibrium.
> Unless I accept 'I ought to be moral', I will not attain wide reflective equilibrium.
> Therefore I must accept 'I ought to be moral'.

When interpreted along the von Wrightian lines of Sect. 4.5.2.1, the argument's conclusion amounts to the claim that one of my ends will be unrealized unless I accept 'I ought to be moral'. The argument takes no stand on the relative strength of the end of wide reflective equilibrium and, by implication, the urgency of accepting 'I ought to be moral'. Whether morality is in some sense overriding is a vexed question that will be taken up below in Sect. 6.5.3.

Some support for the argument's second premise can be gleaned from the following considerations. Since wide reflective equilibrium is here defined in terms of coherence, the premise asserts that accepting 'I ought to be moral' is necessary for a specific sort of coherence. But what is coherence? Any defensible account of coherence would have to avoid maximal and minimal extremes (BonJour 1998, § 3). A maximal notion is mutual entailment, which would unrealistically require that each member of a set of propositions entail and be entailed by all other members of the set. A minimal notion is logical consistency, which would permit a set of propositions that are completely unrelated to each other to be considered coherent. An intermediate concept of coherence for belief systems has been proposed by Walter Sinnott-Armstrong (2006). Drawing on Geoffrey Sayre-McCord's work on coherence for moral theory (1985, p. 171), Sinnott-Armstrong claims that a belief system is coherent "to the extent that its beliefs are jointly consistent, connected, and comprehensive" (2006, p. 222). The connectedness condition, which requires that beliefs in a system provide each other with mutual support, is particularly illuminating in this context.

Consider a set composed of phenomenal discourse to the effect that certain actions and persons are just, courageous, and honest; instrumental discourse that recommends just, courageous, and honest actions as means to moral ends; and the teleological directive 'I ought to be moral'. This discursive set, which I will call the C-set, contrasts with the D-set, which is phenomenally and instrumentally identical to the C-set but replaces the teleological directive 'I ought to be moral' with its deontic contradictory: 'It is not the case that I ought to be moral'. The phenomenal strata of both sets presuppose the interest of justice, courage, and honesty, and the instrumental strata urge their importance. But 'It is not the case that I ought to be moral' in the D-set fails to connect with the lower discursive strata. Imagine

children who are given clothes for a wedding, instructed to wear them properly at the wedding, and then told they need not go to the wedding. Like the phenomenal and instrumental strata of the D-set, they would be all dressed up with no place to go. By contrast, the phenomenal and instrumental strata of the C-set are smoothly connected by 'I ought to be moral'. Hence accepting 'I ought to be moral' is necessary for the connectedness of discursive sets like the C-set. Since connectedness is necessary for coherence, accepting 'I ought to be moral' is necessary for the coherence of such sets. Consequently, it is necessary for the wide reflective equilibrium of such sets, since wide reflective equilibrium is a specific sort of coherence. The point is immediately generalizable to more complex sets of moral discourse. Accepting 'I ought to be moral' is necessary for the wide reflective equilibrium of moral discourse in general.

To further motivate the appeal to coherence in justifying 'I ought to be moral', consider the problem of justification from a wider perspective: the Platonic trinity of the good, the true, and the beautiful (*Republic* 508e–509a). Moral, scientific, and esthetic discourses that impinge on these ends can be analyzed into phenomenal, instrumental, and teleological strata. Imagine that our scientific discourse includes phenomenal descriptions of geological faults, instrumental directives regulating the use of seismographs, and a teleological directive to accurately predict earthquakes. Similarly, esthetic discourse might contain phenomenal descriptions of the dramatic unities of classical plays, instrumental directives to observe the unities of time, place, and action, and a teleological directive to produce esthetically compelling drama. Scientific teleology and esthetic teleology face the same kind of justificatory challenge as moral teleology. We might just as well ask 'Why seek truth?' and 'Why seek beauty?' as 'Why seek goodness?'. And we might just as well provide the same kind of answer: phenomenal and instrumental discourse in these fields would be disoriented without the good, the true, and the beautiful as ends.

Uncomplicated as the foregoing justification of accepting 'I ought to be moral' in terms of wide reflective equilibrium appears to be, it is open to at least three immediate objections. One is that coherence among morally phenomenal, instrumental, and teleological discourse is not a desideratum at all. Moral skeptics of a radical bent might want to pulverize all three strata and so consider their mutual coherence to be beside the point.

The proper response to that, I think, is to point out that substantial portions of morally phenomenal and instrumental discourse rest on solid ground. Let us review the phenomenal case. Sections 2.2–2.3 developed the theses that thick core classifications are carried out by analogy with prototypes and that these analogies can be based entirely on factual properties, as moral subjectivists like J. L. Mackie concede (1977, pp. 16–17, 25–27, 41). Section 2.4.1 advanced a classically-inspired standard of inductive cogency that can be used to save the moral phenomena, that is, to show that established classifications are inductively cogent analogies whereas their contradictories are not. The case for the instrumental stratum is at least as strong. Whether an action type or token conduces to a given end is really just a matter of fact: the purported means leads to the end, or it does not. Granted, knowing the instrumental facts can be difficult or even impossible in special circumstances;

technological limitations can prevent us from knowing what they are, for example. But even in the worst epistemological case, the facts are still the facts, and they can in principle be known. The knowing requires keying on causal relations asserted by individual sentences (Sect. 4.4); on inductive cogency for practical inferences (Sect. 4.5.2); and on decision theory for theories (Sect. 4.6).

To throw out moral discourse as a whole would be to jettison huge chunks of discourse, some of which have firmly reasonable grounds. It would be worse than this, in fact, for the damage could not be contained. If moral classification based on analogies among factual properties cannot be justified, then what grounds could be found for nonmoral classification based on analogies among factual properties? If it makes no sense to say that moral means conduce to their ends, what sense would it make to say that nonmoral means conduce to their ends? And if moral ends are rationally groundless, wouldn't other ends be rationally groundless as well? This last question would prompt ready assent from those who hold a purely subjective theory of value, but the first two questions would not. Imposing a no-morality policy on cognitive discourse would simply eviscerate the language.

A second objection to the justification of accepting 'I ought to be moral' is that the coherence card can be played in contexts that are deeply immoral. Think of the many discursive contexts throughout history characterized by phenomenal discourse identifying certain individuals and groups as ethnically unclean; instrumental discourse about various means of ethnic cleansing; and a teleological directive to achieve ethnic cleanliness. In such situations, an appeal to wide reflective equilibrium might be mounted as follows:

> I want wide reflective equilibrium.
> Unless I endorse ethnic cleanliness, I will not attain wide reflective equilibrium.
> Therefore I must endorse ethnic cleanliness.

'Endorse ethnic cleanliness' evidently unifies the phenomenal and instrumental strata of this discourse far more effectively than 'Do not endorse ethnic cleanliness'. Hence, we might conclude, 'Endorse ethnic cleanliness' is justified.

The mistake in this maneuver is the narrowness of the equilibrium it intends. However coherent the discourse on ethnic cleansing may be, it clashes head-on with moral discourse on several fronts. Most obviously, it clashes with the moral perspectives of those singled out for ethnic cleansing. It also clashes with what Donagan called "common morality": the part of traditional morality that does not rely on theistic commitments (1977, pp. 27–28; cf. Gert 2004). And, without exception, it clashes with established moral theory. The result is narrow equilibrium within the mini-discourse on ethnic cleansing but wide disequilibrium between it and moral discourse. This is why, as a practical matter, those who attempt to extend the meme of ethnic cleansing throughout society typically resort to subterfuge. They employ a disingenuous vocabulary of final solutions and hygienic cleansings. They disqualify their intended victims as subhuman to deny them moral standing. They portray their victims as mortal threats in order to activate the mechanisms of self-defense. Whatever serves, in short, to smudge the lines of conflict between ethnic purity and morality.

A third objection to the prior justification of accepting 'I ought to be moral' is to deny the first premise: 'I want wide reflective equilibrium'. Just as the utilitarian and perfectionist defenses of morality can be weakened by the moral skeptic's replying 'I don't want collective happiness' or 'I don't want self-perfection', the moral skeptic can attack the equilibrium justification by claiming 'I don't want wide reflective equilibrium'. The problem would then be to counter this counter. Why wide reflective equilibrium? Or, in more general terms, why coherence?

One response to this challenge is to claim that coherence is valuable because it is truth-conducive: mutually coherent propositions are more likely to be true than mutually incoherent ones. The following section will explore this response in detail. In the meantime, however, might the moral skeptic contest this response by challenging the goal of truth? Of course. Just as the skeptic can deny the importance of collective happiness, self-perfection, and coherence, the skeptic can deny the importance of truth.

But what was the point of the skeptic's challenge to moral teleology? Wasn't it that morally teleological discourse has no rational basis? And doesn't this imply the claim that morally teleological discourse has no claim to truth? Then this claim presupposes the importance of truth. If the skeptic replies that the presupposition is mine, not hers, I need only observe that the claim that morally teleological discourse has no rational basis is a truth claim. Hence the skeptic has presupposed the importance of truth. The skeptic would therefore be guilty of performative contradiction (cf. Searle 1969, p. 30, 137); that is, the skeptic would have both presupposed the importance of truth and denied the importance of truth. And that, I think, is to jump the Ship of Reason.

5.8 Why Coherence?

Why strive for coherence? As just noted, a standard answer is that coherence is truth-conducive: mutually coherent propositions are more likely to be true than mutually incoherent ones. But this answer cannot mean that coherence is the only epistemic consideration, for other conditions such as prior probability and witness reliability bear epistemic weight as well (Bovens and Hartmann 2003, pp. 10–11). That coherence is truth-conducive *ceteris paribus* is a more defensible claim. I propose that we consider this claim in three contexts: probabilistic coherence (Sect. 5.8.1); plausibilistic coherence (Sect. 5.8.2); and contemporary epistemology (Sect. 5.8.3).[6]

5.8.1 Probabilistic Coherence

C. I. Lewis illustrated the probabilistic truth-conduciveness claim with an example of several "relatively unreliable witnesses who independently tell the same

[6] Sections 5.8.1–5.8.2 draw on Welch (2014).

circumstantial story," arguing that "congruence of the reports establishes a high probability of what they agree upon" (1946, p. 346). Wally, Wendy, and Willy may each tell a dubious story, but the evidence that their stories independently cohere raises the probability that their common part is true.

Unfortunately, recent investigation casts serious doubts on this claim. Tomoji Shogenji offers a counterexample of more coherence but less probability:

> [C]onsider an epistemically ultraconservative agent who only holds a few extremely unspecific beliefs—say, some rocks are heavier than others; some animals sleep sometimes; and someone is humming some tune somewhere. It is very likely that her beliefs are all true even though they do not hang together. Meanwhile a huge collection of highly specific beliefs—such as the entire body of medical science—almost certainly contains errors even though they tightly hang together. (1999, p. 342)

Even worse for the prospects of Lewis' claim are the impossibility results obtained independently by Luc Bovens and Stephan Hartmann (2003, pp. 19–22) and Erik J. Olsson (2005a, b, pp. 134–204, 211–215). Olsson maintains that the constraints that have been imposed on the concept of coherence are jointly incompatible, making coherence comparable to a square circle (2005a, pp. 409–410). As a result, whatever probabilistic measure of coherence is chosen, counterexamples to truth-conduciveness can be found.

Nevertheless, there are multiple strategies for defending the truth-conduciveness claim (Douven and Meijs 2007a). Bovens and Hartmann initially define the relation 'is no less coherent than' as an ordering in order to derive their impossibility result, but they manage to sidestep this result by weakening 'is no less coherent than' to a partial ordering (2003, Chap. 2).[7] Franz Dietrich and Luca Moretti show that coherence is confirmation-conducive; that is, the coherence of a set of statements, beliefs, hypotheses, etc. transmits confirmation to its members. This makes coherence truth-conducive provided there is evidence that confirms a member of a coherent set (2005; Moretti 2007). Jonah Schupbach establishes that *ceteris paribus* conditions other than Bovens and Hartmann's are both plausible and avoid their impossibility result (2008). Staffan Angere reminds us that the perfect can be the enemy of the good:

> What is common about both these theorems [Bovens and Hartmann's, Olsson's] is that they work by finding counterexamples, or rather by showing that, whatever measure C of coherence we pick, counterexamples to their truth-conduciveness can be found. But this does not mean that all measures are equally bad—just that no measure is perfect. For everyday reasoning, it can be held that we can have good use for merely an indication of truth, or even an indication of higher probability, since even good probabilistic data can be hard to come by. (2007, p. 322)

Angere proceeds to demonstrate that several standard coherence measures are *partially* truth-conducive: for belief sets with up to 15 members, greater coherence correlates with greater probability about 75 % of the time (2007, pp. 328–331; 2008,

[7] Terminology varies. Bovens and Hartmann actually use the term 'quasi-ordering' to describe a relation that is reflexive, transitive, but not necessarily complete (2003, p. 25).

pp. 13–19). This result provides weighty evidence, it seems to me, that coherence is a useful but defeasible heuristic for finding the truth.

Useful it may be, but coherence is a notoriously vague notion. One way to clarify it is to define probabilistic measures of coherence by appealing to probabilistic measures of confirmation. Some coherence measures are identical to probabilistic measures of confirmation (Shogenji 1999, p. 339), while others are extensions of such measures (Douven and Meijs 2007b, pp. 410–412). Since there are at least nine probabilistic measures of confirmation (Atkinson et al. 2009, pp. 4–5), options for defining probabilistic coherence abound.

To introduce one such option, we need the notion of a relevance measure of confirmation. Relevance measures can be defined for probability p, hypothesis h, evidence e, and confirmation c as follows:

if $p(h|e) > p(e)$, then $c(\langle h, e\rangle) > 0$;
if $p(h|e) = p(e)$, then $c(\langle h, e\rangle) = 0$;
if $p(h|e) < p(e)$, then $c(\langle h, e\rangle) < 0$.

One relevance measure of confirmation is the difference measure (Carnap 1962, p. 361), expressed by Eq. 5.2.

$$c(h,e) = p(h \mid e) - p(h). \tag{5.2}$$

Though there are other relevance measures, the difference measure does enjoy a certain prominence in the literature. Richard Jeffrey called it "the basic concept of probabilistic methodology" (1992, p. 72), and Igor Douven and Wouter Meijs favor it over six other measures of confirmation (2007b, p. 417).

Douven and Meijs also note five ways of understanding the coherence of a finite set of propositions (Douven and Meijs 2007b, pp. 407–408). Of these five ways, they claim that any-any coherence best captures the notion of a set of propositions hanging together well (2007b, p. 409, 411). Any–any coherence is a function of the dependence of any non-null subset on any other non-overlapping, non-null subset.

To define any–any coherence, the difference measure is first generalized for ordered pairs of sets. Where S is a finite set of propositions $\{R_1, \ldots, R_n\}$ with non-empty, non-overlapping subsets S' and S*, the degree to which S* confirms S' is given by the modified difference measure in Eq. 5.3.

$$c(\langle S', S^*\rangle) = p(\wedge S' \mid \wedge S^*) - p(\wedge S'). \tag{5.3}$$

Then any–any coherence can be defined as a simple average of the confirmation each subset affords the other. For example, if $S = \{R_1, R_2\}$, $S' = \{R_1\}$, and $S^* = \{R_2\}$, the coherence κ of S would be determined by Eq. 5.4.

$$\kappa(S) = \frac{\left[c(\langle S', S^*\rangle) + c(\langle S^*, S'\rangle)\right]}{2}. \tag{5.4}$$

The result is a manageable notion of probabilistic coherence. Reliance on the difference measure is inessential, however; any preferred measure could be used instead.

5.8.2 Plausibilistic Coherence

When coherence can be understood in probabilistic terms, I think it should be. The problem is that often it cannot be so understood. Many situations are so cognitively impoverished that we are unable to gauge the prior and posterior probabilities required for probabilistic measures of confirmation. In such situations, it is nonetheless frequently possible to estimate non-probabilistic prior and posterior plausibilities. I propose, then, that the foregoing approach to probabilistic coherence be extended to plausibilistic coherence. If we generalize our approach to coherence by expanding it from probability to plausibility, we can still rely on the stronger notion of probabilistic plausibility when possible but fall back on the weaker notion of non-probabilistic plausibility when necessary.

An unrefined claim about probabilistic truth-conduciveness is 'the more coherent, the more probable'. An equally unrefined claim about plausibilistic truth conduciveness is 'the more coherent, the more plausible'. This claim is plainly overbold, however, because it ignores the epistemic role of other considerations. Weakening it accordingly results in the claim that greater coherence is correlated with greater plausibility *ceteris paribus*. Yet even this is too strong. Like the probabilistic truth-conduciveness claim, it runs aground on the impossibility results for the probabilistic special case. Consequently, it needs to be weakened further: *ceteris paribus*, greater coherence is correlated with greater plausibility most of the time. 'Most of the time' is subject to various interpretations: stronger, as in 'well over half the time', or weaker, as in 'more than half' (Angere 2007, pp. 325–326, 2008, pp. 14–15). Hence a truth-conduciveness claim employing this expression has stronger and weaker forms. The plausibilistic truth-conduciveness claim that follows is meant to be generic, interpretable according to one's preferred sense of 'most of the time':

> *Partial truth-conduciveness ceteris paribus*: for a coherence measure κ, a plausibility measure π, and sets of propositions S_1 and S_2, κ is truth-conducive if, and only if, the following condition is satisfied: *ceteris paribus*, if $\kappa(S_1) > \kappa(S_2)$, then $\pi(S_1) > \pi(S_2)$ most of the time.

Choice among stronger and weaker forms of this claim need not concern us here since the subsequent discussion focuses on the antecedent of the truth-conduciveness condition.

How could we tell whether the antecedent's inequality '$\kappa(S_1) > \kappa(S_2)$' is satisfied? Since probabilistic coherence can be defined with reference to probabilistic confirmation, an analogous tactic might work for plausibilistic coherence. Though 'confirmation' typically means 'increase of probability due to new evidence', Sect. 5.4 noted that the term can be straightforwardly generalized to mean 'increase of plausibility due to new evidence'. Accordingly, the probabilistic difference measure could first be generalized for the plausibility measure π as in Eq. 5.5.

$$c(h,e) = \pi(h \mid e) - \pi(h). \tag{5.5}$$

When the plausibility values are numeric, as in the probabilistic special case, the minus sign would have its usual arithmetical sense. But when the plausibilities are non-numeric, the minus sign would be interpreted as a bifunctor that assigns a set of non-numeric plausibilities **P** to a set of non-numeric differences **D**. For instance, where **P** = {high, medium, low} and **D** = {great, small, null}, the right side of the difference measure would generate mappings such as these: {high, low} → {great}; {medium, low} → {small}; {low, low} → {null}.

Then, similar to Douven and Meijs' probabilistic generalization of the difference measure for ordered pairs of nonempty, non-overlapping subsets *S*' and *S** (Eq. 5.3), the plausibilistic generalization of the difference measure could be stated by Eq. 5.6.

$$c(\langle S', S^* \rangle) = \pi(\wedge S' \mid \wedge S^*) - \pi(\wedge S'). \tag{5.6}$$

Finally, we could treat any–any coherence as an average of the confirmation afforded to *S*' by *S** and vice versa. As noted for the probabilistic case, other measures of confirmation could be employed *mutatis mutandis*.

The advantage of a plausibilistic approach to coherence is its viability in situations that are too amorphously characterized for probabilistic coherence. That is, plausibilistic coherence is viable provided the situation is rich enough to determine prior and posterior plausibilities such that the generalized difference measure can be applied. Naturally, the plausibilities to which the difference measure can be applied are quite varied. But even when we cannot determine numeric differences among them, non-numeric differences such as 'great' and 'small' can turn out to be highly useful.

Now consider the set $S_1 = \{R_1, R_2\}$, where R_1 and R_2 are neither logically equivalent nor logically contradictory.[8] Let the results of applying the difference measure to the subsets $S' = \{R_1\}$ and $S^* = \{R_2\}$ be one of five values:

D (strong confirmation)
d (weak confirmation)
0 (no confirmation)
$-d$ (weak disconfirmation)
$-D$ (strong disconfirmation).

The comparative relations among these values are evidently $D > d > 0 > -d > -D$. These values are admittedly coarse, but they could be used to represent many real-life situations where an agent is able to estimate only that the difference between a posterior and prior plausibility is large and positive, small and positive, nonexistent, small and negative, or large and negative. In this sense, the values are quite realistic.

Under these assumptions, a coherence measure κ can be defined such that its values are averages of confirmation values. Given that averaging with respect to

[8] This restriction is inessential. Logically equivalent and contradictory propositions could be included by stipulating that the conditional plausibility of equivalents is the maximal value $\top$ and that of contradictories is the minimal value $\bot$.

Table 5.2 Numerators of averaged confirmation values

Positive	Zero	Negative
$D+D$	$D+(-D)$	$-d+0$
$D+d$	$0+0$	$-D+d$
$D+0$	$d+(-\mathrm{d})$	$-d+(-\mathrm{d})$
$d+d$		$-D+0$
$D+(-d)$		$-D+(-\mathrm{d})$
$d+0$		$-D+(-D)$

S' and *S** will require some of our five values to be paired with others, the number of possible binary combinations with repetition of these values is given by Eq. 5.7.

$$\binom{5+2-1}{2} = \frac{6!}{2!(6-2)!} = 15. \tag{5.7}$$

Each of these combinations is an average such as $(D+d)/2$. But we can simplify slightly since our focus is the inequality '$\kappa(S_1)>\kappa(S_2)$' of the partial truth-conduciveness claim defined above and the κ-values are averages whose denominators are identical. Because the numerators alone determine the truth value of the inequality, only they are listed here. Table 5.2 organizes them into three groups: positive, zero, and negative. Taken together, these groupings reflect the symmetry of the initial values obtained from the generalized difference measure.

Happily, some of these numerators are eliminable. The reason can be illustrated with the probabilistic special case. Bayes' theorem in its simplest form is expressed by Eq. 5.8.

$$p(h \mid e) = p(h) \times \frac{p(e \mid h)}{p(e)}. \tag{5.8}$$

In this form, Bayes' theorem immediately yields Eq. 5.9.

$$\frac{p(h \mid e)}{p(h)} = \frac{p(e \mid h)}{p(e)}. \tag{5.9}$$

That is, the relation between posterior and prior probabilities is the same for the hypothesis as it is for the evidence. Hence *h* is confirmed (disconfirmed, neither confirmed nor disconfirmed) by *e* if, and only if, *e* is confirmed (disconfirmed, neither confirmed nor disconfirmed) by *h*. This eliminates mixed cases of the following sorts:

a. the evidence confirms the hypothesis but the hypothesis disconfirms the evidence;
b. the evidence disconfirms the hypothesis while the hypothesis confirms the evidence;
c. the evidence confirms the hypothesis but the hypothesis neither confirms nor disconfirms the evidence;
d. the evidence neither confirms nor disconfirms the evidence while the hypothesis confirms the evidence;

Table 5.3 Relevant numerators of averaged confirmation values

Positive	Zero	Negative
$2D$	0	$-2d$
$D+d$		$-D-d$
$2d$		$-2D$

e. the evidence disconfirms the hypothesis but the hypothesis neither confirms nor disconfirms the evidence;
f. the evidence neither confirms nor disconfirms the evidence while the hypothesis disconfirms the evidence.

The underlying logic extends to the plausibilistic general case. Since hypothesis and evidence are positively correlated, the relation between posterior and prior plausibilities for the hypothesis is mirrored by that between posterior and prior plausibilities for the evidence. As a result, mixed cases characterized by confirmation and disconfirmation, or confirmation and neutral confirmation, or disconfirmation and neutral confirmation will not occur. They can therefore be eliminated from the values in Table 5.2. After obvious simplifications, we are left with the array of numerators in Table 5.3. The comparative relations among these numerators are as follows: $2D > (D+d) > 2d > 0 > -2d > (-D-d) > -2D$.

To get a more concrete feel for plausibilistic coherence, let us consider three simple examples. In the first, we contrast two sets of propositions whose coherence is intuitively quite different: $S_1 = \{B, F\}$, where

B: Bobo is a bird;
F: Bobo flies;

and $S_2 = \{B, T\}$, where

B: Bobo is a bird;
T: Baldo is a toad.

Suppose that the plausibilities of these propositions are expressed in terms of a set of plausibilities $\mathbf{P} = \{H, I, L\}$, where H = high, I = intermediate, and L = low. In addition, differences among these plausibilities can be represented through the set of differences $\mathbf{D} = \{D, d, 0, -d, -D\}$ introduced four paragraphs above.

Let an agent equipped with a plausibility measure π apply Eq. 5.6, the generalized difference measure, as follows:

For S_1:

$$c(\langle\{B\}, \{F\}\rangle) = \pi(B|F) - \pi(B) = I - L = d.$$
$$c(\langle\{F\}, \{B\}\rangle) = \pi(F|B) - \pi(F) = H - L = D.$$

For S_2:

$$c(\langle\{B\}, \{T\}\rangle) = \pi(B|T) - \pi(B) = L - L = 0.$$
$$c(\langle\{T\}, \{B\}\rangle) = \pi(T|B) - \pi(T) = L - L = 0.$$

In these circumstances, then, $\kappa(S_1) = (d+D)/2$ and $\kappa(S_2) = (0+0)/2$. Hence $\kappa(S_1) > \kappa(S_2)$.

The second example is also structured by two sets of propositions, but it employs triples instead of pairs. Let $S_1 = \{B, G, P\}$, where

B: Burning fossil fuels boosts the greenhouse effect;
G: Global warming is underway;
P: The polar ice caps are melting.

By contrast, $S_2 = \{\overline{B}, \overline{G}, \overline{P}\}$, where

$\overline{B}$: Burning fossil fuels does not boost the greenhouse effect;
$\overline{G}$: Global warming is not underway;
$\overline{P}$: The polar ice caps are not melting.

Very few people will find S_1 and S_2 equally plausible, yet they appear to be equally coherent. This intuitive equicoherence can be confirmed under plausible assumptions by calculating plausibilistic coherence with the sets of plausibilities **P** and differences **D** used in the first example.

Suppose that an agent applies a plausibility measure π and Eq. 5.6, the generalized difference measure, as follows.

For S_1:

$c(\langle\{B\}, \{G\}\rangle) = \pi(B|G) - \pi(B) = H - I = d.$
$c(\langle\{B\}, \{P\}\rangle) = \pi(B|P) - \pi(B) = H - I = d.$
$c(\langle\{B\}, \{G, P\}\rangle) = \pi(B|G \wedge P) - \pi(B) = H - I = d.$
$c(\langle\{G\}, \{B\}\rangle) = \pi(G|B) - \pi(G) = H - I = d.$
$c(\langle\{G\}, \{P\}\rangle) = \pi(G|P) - \pi(G) = H - I = d.$
$c(\langle\{G\}, \{B, P\}\rangle) = \pi(G|B \wedge P) - \pi(G) = H - I = d.$
$c(\langle\{P\}, \{B\}\rangle) = \pi(P|B) - \pi(P) = H - I = d.$
$c(\langle\{P\}, \{G\}\rangle) = \pi(P|G) - \pi(P) = H - I = d.$
$c(\langle\{P\}, \{B, G\}\rangle) = \pi(P|B \wedge G) - \pi(P) = H - I = d.$
$c(\langle\{B, G\}, \{P\}\rangle) = \pi(B \wedge G|P) - \pi(B \wedge G) = H - I = d.$
$c(\langle\{B, P\}, \{G\}\rangle) = \pi(B \wedge P|G) - \pi(B \wedge P) = H - I = d.$
$c(\langle\{G, P\}, \{B\}\rangle) = \pi(G \wedge P|B) - \pi(G \wedge P) = H - I = d.$

For S_2:

$c(\langle\{\overline{B}\}, \{\overline{G}\}\rangle) = \pi(\overline{B}|\overline{G}) - \pi(\overline{B}) = I - L = d.$
$c(\langle\{\overline{B}\}, \{\overline{P}\}\rangle) = \pi(\overline{B}|\overline{P}) - \pi(\overline{B}) = I - L = d.$
$c(\langle\{\overline{B}\}, \{\overline{G}, \overline{P}\}\rangle) = \pi(\overline{B}|\overline{G} \wedge \overline{P}) - \pi(\overline{B}) = I - L = d.$
$c(\langle\{\overline{G}\}, \{\overline{B}\}\rangle) = \pi(\overline{G}|\overline{B}) - \pi(\overline{G}) = I - L = d.$
$c(\langle\{\overline{G}\}, \{\overline{P}\}\rangle) = \pi(\overline{G}|\overline{P}) - \pi(\overline{G}) = I - L = d.$
$c(\langle\{\overline{G}\}, \{\overline{B}, \overline{P}\}\rangle) = \pi(\overline{G}|\overline{B} \wedge \overline{P}) - \pi(\overline{G}) = I - L = d.$
$c(\langle\{\overline{P}\}, \{\overline{B}\}\rangle) = \pi(\overline{P}|\overline{B}) - \pi(\overline{P}) = I - L = d.$
$c(\langle\{\overline{P}\}, \{\overline{G}\}\rangle) = \pi(\overline{P}|\overline{G}) - \pi(\overline{P}) = I - L = d.$
$c(\langle\{\overline{P}\}, \{\overline{B}, \overline{G}\}\rangle) = \pi(\overline{P}|\overline{B} \wedge \overline{G}) - \pi(\overline{P}) = I - L = d.$
$c(\langle\{\overline{B}, \overline{G}\}, \{\overline{P}\}\rangle) = \pi(\overline{B} \wedge \overline{G}|\overline{P}) - \pi(\overline{B} \wedge \overline{G}) = I - L = d.$
$c(\langle\{\overline{B}, \overline{P}\}, \{\overline{G}\}\rangle) = \pi(\overline{B} \wedge \overline{P}|\overline{G}) - \pi(\overline{B} \wedge \overline{P}) = I - L = d.$
$c(\langle\{\overline{G}, \overline{P}\}, \{\overline{B}\}\rangle) = \pi(\overline{G} \wedge \overline{P}|\overline{B}) - \pi(\overline{G} \wedge \overline{P}) = I - L = d.$

Consequently, under these assumptions, $\kappa(S_1)=d$ and $\kappa(S_2)=d$. Hence $\kappa(S_1)=\kappa(S_2)$.

Even though these two sets may be equally coherent, they are not equally plausible. What then of coherence as a criterion of truth? If coherence is really truth-conducive, why doesn't it lead to truth here? We recall that the truth-conduciveness claim for plausibilistic coherence contains a *ceteris paribus* clause: other things being equal, greater coherence is correlated with greater plausibility most of the time. In this example, the *ceteris paribus* condition is violated: other things are not equal. Other relevant things include prior probability and witness reliability, as noted above, in addition to prior plausibility, which we can now add to the list of epistemic considerations. The prior plausibilities of *B* ('Burning fossil fuels boosts the greenhouse effect'), *G* ('Global warming is underway'), and *P* ('The polar ice caps are melting') are unequal to those of $\overline{B}$, $\overline{G}$, and $\overline{P}-I$ versus *L* respectively, according to the hypothetical agent of our example. Hence the two sets of propositions are not equally truth-conducive despite being equally coherent.

The third example involves sets of propositions whose coherence contrasts less starkly than those of the first example. Like the second example, it involves two sets of triples. $S_1=\{M, R, W\}$, where

M: Dennis the Menace is near the window;
R: A rock is near the window;
W: The window is broken.

By contrast, $S_2=\{M, F, W\}$, where

M: Dennis the Menace is near the window;
F: A flower is near the window;
W: The window is broken.

We assume the same sets of plausibilities **P** and differences **D** as in the previous examples.

Let an agent apply a plausibility measure π and Eq. 5.6, the generalized difference measure, along the following lines.

For S_1:

$c(\langle\{M\}, \{R\}\rangle)=\pi(M|R)-\pi(M)=L-L=0.$
$c(\langle\{M\}, \{W\}\rangle)=\pi(M|W)-\pi(M)=I-L=d.$
$c(\langle\{M\}, \{R, W\}\rangle)=\pi(M|R\wedge W)-\pi(M)=H-L=D.$
$c(\langle\{R\}, \{M\}\rangle)=\pi(R|M)-\pi(R)=L-L=0.$
$c(\langle\{R\}, \{W\}\rangle)=\pi(R|W)-\pi(R)=L-L=0.$
$c(\langle\{R\}, \{M, W\}\rangle)=\pi(R|M\wedge W)-\pi(R)=I-L=d.$
$c(\langle\{W\}, \{M\}\rangle)=\pi(W|M)-\pi(W)=I-L=d.$
$c(\langle\{W\}, \{R\}\rangle)=\pi(W|R)-\pi(W)=L-L=0.$
$c(\langle\{W\}, \{M, R\}\rangle)=\pi(W|M\wedge R)-\pi(W)=H-L=D.$
$c(\langle\{M, R\}, \{W\}\rangle)=\pi(M\wedge R|W)-\pi(M\wedge R)=I-L=d.$
$c(\langle\{M, W\}, \{R\}\rangle)=\pi(M\wedge W|R)-\pi(M\wedge W)=I-L=d.$
$c(\langle\{R, W\}, \{M\}\rangle)=\pi(R\wedge W|M)-\pi(R\wedge W)=I-L=d.$

For S_2:

$c(\langle\{M\}, \{F\}\rangle) = \pi(M|F) - \pi(M) = L - L = 0.$
$c(\langle\{M\}, \{W\}\rangle) = \pi(M|W) - \pi(M) = I - L = d.$
$c(\langle\{M\}, \{F, W\}\rangle) = \pi(M|F \wedge W) - \pi(M) = I - L = d.$
$c(\langle\{F\}, \{M\}\rangle) = \pi(F|M) - \pi(F) = L - L = 0.$
$c(\langle\{F\}, \{W\}\rangle) = \pi(F|W) - \pi(F) = L - L = 0.$
$c(\langle\{F\}, \{M, W\}\rangle) = \pi(F|M \wedge W) - \pi(F) = L - L = 0.$
$c(\langle\{W\}, \{M\}\rangle) = \pi(W|M) - \pi(W) = I - L = d.$
$c(\langle\{W\}, \{F\}\rangle) = \pi(W|F) - \pi(W) = L - L = 0.$
$c(\langle\{W\}, \{M, F\}\rangle) = \pi(W|M \wedge F) - \pi(W) = I - L = d.$
$c(\langle\{M, F\}, \{W\}\rangle) = \pi(M \wedge F|W) - \pi(M \wedge F) = I - L = d.$
$c(\langle\{M, W\}, \{F\}\rangle) = \pi(M \wedge W|F) - \pi(M \wedge W) = L - L = 0.$
$c(\langle\{F, W\}, \{M\}\rangle) = \pi(F \wedge W|M) - \pi(F \wedge W) = I - L = d.$

Hence $\kappa(S_1) = (d + \frac{1}{3}D)/2$ and $\kappa(S_2) = d/2$. Consequently, $\kappa(S_1) > \kappa(S_2)$.

Limited though this third example is, it is highly suggestive. It suggests both what can and cannot be done within the current framework. What can be done is to calculate the coherence of larger sets of propositions. The calculations are inevitably more tedious, of course, but they can be carried out provided the sets are finite. What cannot be done, unfortunately, is to compare the results of these calculations in every case. Though the results are often comparable, as we have seen in all three of our examples, there are also results as simple as $\kappa(S_1) = \frac{1}{2}D$ and $\kappa(S_2) = d$ that are incomparable because we do not know by how much D is greater than d.

Granted, we might devise a scoring system whereby D is worth two points, say, and d is worth one—in which case the two sets would be judged equally coherent. But this would be tantamount to substituting a cardinal scale of difference for an ordinal one, and this would violate the spirit of the exercise. We have been concerned to see whether coherence can be measured in information-poor situations where only non-numeric plausibilities and differences are at hand.

A more consonant approach would be to replace the set of differences **D** with the more limited set $\mathbf{D^*} = \{d, 0, -d\}$. This would amount to ignoring the differences between D and d, on the one hand, and $-D$ and $-d$, on the other. If we were to adopt this expedient, many comparative relations would be preserved. In the Dennis the Menace example above, we would have $\kappa(S_1) = \frac{2}{3}d$ and $\kappa(S_2) = \frac{1}{2}d$, which preserves the relation $\kappa(S_1) > \kappa(S_2)$. In addition, some incomparable differences would be converted to comparable ones. Whereas $\kappa(S_1) = \frac{1}{2}D$ and $\kappa(S_2) = d$ are incomparable in terms of the five-valued set of differences **D**, they would be comparable in terms of the three-valued set **D***, for $\kappa(S_1) = d$ and $\kappa(S_2) = d$.

Constricting the set of differences can be a costly maneuver, however. If $\kappa(S_1) = D$ and $\kappa(S_2) = d$ in terms of the five-valued set of differences, then $\kappa(S_1) = d$ and $\kappa(S_2) = d$ in terms of the three-valued set. Here the greater coherence of S_1 according to the more nuanced set of differences would be transformed to equal coherence by the less nuanced set. That would be a mistake, I think. While a policy that would convert some incomparables to comparables might seem attractive at first glance, a policy that would change some inequalities to equalities would not. The problem

with such a policy is that it ignores some of the relevant information at hand (the inequality of *D* and *d*, for instance). It therefore violates the requirement of total evidence that Carnap enunciated for inductive logic: "the total evidence available must be taken as a basis for determining the degree of confirmation" (1962, p. 211). The requirement applies well beyond the ambit of inductive logic. In fact, as Hempel pointed out, "One might well say that it [the requirement of total evidence] is simply a partial explication of conditions governing rational belief and rational choice" (1965, p. 66).

In short, a second glance at a policy that would convert incomparables to comparables might not be as favorable as the first. The plain fact is that cognitive quantities—probabilities, plausibilities, plausibilistic differences, and utilities, to name a few—are sometimes incomparable. Even if we hold that such quantities are ultimately comparable *sub specie aeternitatis*, we who operate *sub specie temporis* are sometimes unable to discern the ultimate relations of comparability. Hence to adopt a policy that converts incomparables to comparables would be to falsify some cognitive relations. Rather than rush pell-mell to judgment, the Socratic virtue of admitting ignorance is occasionally *de rigueur*. This does not preclude us from affirming our knowledge when we have it, as in the three examples above. But apportioning our belief to the evidence requires that belief be equivocal when the evidence is. This need not be a loss. "The essence of critical thinking," observed Dewey, "is suspended judgment" (1910, p. 74).

5.8.3 Coherence in Contemporary Epistemology

If coherence can be defined along the lines of the previous section, the question 'Why coherence?' has a promising answer: *ceteris paribus*, plausibilistic coherence is truth-conducive most of the time. How much support could such a defense of coherence command? It should be congenial to several camps of contemporary epistemologists. Coherentists would evidently feel at home with it because greater coherence raises plausibility and—once plausibility is sufficiently raised—justifies belief (e.g., Lehrer 2000). Evidentialists who take experiential states involved in rational thought as evidence (e.g., Conee and Feldman 2010, p. 124) could concur that coherence is evidence and that high coherence (perhaps in conjunction with other forms of evidence) can justify belief. Modest foundationalists should be able to live with these views as well, though the situation is more complex. Like C. I. Lewis, some foundationalists grant that either coherence alone or coherence in conjunction with other properties can justify belief. "Considerations of coherence might sometimes, by themselves, suffice to justify beliefs. And perhaps all of your perceptual beliefs are justified in part by such considerations of coherence" (Pryor 2000, p. 535). But other foundationalists claim that even though incoherence can defeat justification, coherence is not a source of justification but a sign (e.g., Audi 2001, p. 24, 28, 71). If coherence is merely a sign of justification, greater coherence signals—but does not itself produce—greater plausibility and possible justification.

The relation between coherence and justification would be mere correlation, rather like an aching joint that signals but does not cause imminent rain.

The claim that incoherence can defeat justification but coherence cannot create it strikes me as dubious. Coherence is a matter of degree. High incoherence is just low coherence, and high coherence is low incoherence. So if a high degree of incoherence (or low coherence) can defeat justification, why could not a high degree of coherence (or low incoherence) be a source of it? Let me suggest an answer in foundationalist terms. According to Audi, well-grounded beliefs are grounded by experience, broadly conceived: perception, introspection, memory, and reason. But reason, I maintain, has coherence as a goal. Hence two belief systems that are epistemically comparable except that one is highly coherent while the other is not are not equally grounded by reason; the highly coherent system is better grounded because reason demands coherence. I suggest, then, that foundationalism can legitimately incorporate coherence as a source, though not the only source, of justification. But even if coherence is merely a sign of justification, it remains epistemically relevant. From any epistemic point of view, coherence is a desideratum (Alston 1993, pp. 529–531).

5.9 Conclusion

Morally teleological discourse is the third of the strata highlighted in Sect. 1.4. The present chapter has argued that this genre of discourse, whether in the form of descriptions or directives, is subject to rational evaluation. Teleological descriptions can be viewed as hypotheses and thereby confirmed or disconfirmed through hypothetico-deductive reasoning. Teleological directives can be chosen reasonably with the aid of comparative decision theory. Hence the teleological stratum, like the instrumental stratum of Chap. 4, is cognitive. In addition, this chapter has endorsed the ends of Frankena's mixed deontologism, already defended on instrumental grounds in Sect. 4.6, and proposed modifications that plug some of the gaps in the original theory. Finally, the chapter has argued that the higher-order end of acting morally can be justified by appealing to coherence with the rest of our moral discourse, much of which can legitimately claim to be factual. Coherence is understood to be wide reflective equilibrium as outlined in Sect. 1.4.

References

Allais, Maurice. 1953. Fondements d'une théorie positive des choix comportant un risque et critique des postulats et axiomes de l'école américaine. *Econometrie* 40:257–332. English edition: Allais, Maurice. 1979. The foundations of a positive theory of choice involving risk and a criticism of the postulates and axioms of the American school. In *Expected utility hypotheses and the Allais paradox: Contemporary discussions of decisions under uncertainty with Allais' rejoinder,* ed. Maurice Allais and Ole Hagen, 27–145. Dordrecht: D. Reidel.

Alston, William P. 1993. Epistemic desiderata. *Philosophy and Phenomenological Research* 53:527–551.
Angere, Staffan. 2007. The defeasible nature of coherentist justification. *Synthese* 157:321–335.
Angere, Staffan. 2008. Coherence as a heuristic. *Mind* 117:1–26.
Atkinson, David, Jeanne Peijnenburg, and Theo Kuipers. 2009. How to confirm the conjunction of disconfirmed hypotheses. *Philosophy of Science* 87:1–21.
Audi, Robert. 1989. *Practical reasoning*. London: Routledge.
Audi, Robert. 2001. *The architecture of reason: The structure and substance of rationality*. Oxford: Oxford University Press.
BonJour, Laurence. 1998. Knowledge and justification, coherence theory of. In *Routledge encyclopedia of philosophy,* ed. E. Craig. London: Routledge. http://www.rep.routledge.com/article/P009. Accessed 30 April 2014.
Bovens, Luc, and Stephan Hartmann. 2003. *Bayesian epistemology*. Oxford: Oxford University Press.
Braybrooke, David. 1987. Social contract theory's fanciest flight. *Ethics* 97:750–764.
Carnap, Rudolf. 1962. *Logical foundations of probability*. 2nd ed. Chicago: University of Chicago Press.
Conee, Earl, and Richard Feldman. 2010. In *A companion to epistemology,* ed. Jonathan Dancy, Ernest Sosa, and Matthias Steup, 2nd ed., 123–130. Chichester: Wiley-Blackwell.
Dewey, John. 1910. *How we think*. Lexington: D. C. Heath.
Dewey, John. 1922. *Human nature and conduct: An introduction to social psychology*. In The middle works, 1899–1924, eds. Jo Ann Boydston and Patricia Baysinger, vol. 14. Carbondale: Southern Illinois University Press, 1983.
Dietrich, Franz, and Luca Moretti. 2005. On coherent sets and the transmission of confirmation. *Philosophy of Science* 72:403–424.
Donagan, Alan. 1977. *The theory of morality*. Chicago: The University of Chicago Press.
Douven, Igor, and Wouter Meijs. 2007a. On the alleged impossibility of coherence. *Synthese* 157:347–360.
Douven, Igor, and Wouter Meijs. 2007b. Measuring coherence. *Synthese* 156:405–425.
Dyson, Freeman. 1988. *Infinite in all directions*. London: Penguin.
Edvardsson, Karin, and Sven Ove Hansson. 2005. When is a goal rational? *Social Choice and Welfare* 24:343–361.
Flanagan, Owen. 1986. Admirable immorality and admirable imperfection. *The Journal of Philosophy* 83:41–61.
Foot, Philippa. 1972. Morality as a system of hypothetical imperatives. *Philosophical Review* 81:305–316. In *Virtues and vices and other essays in moral philosophy,* 157–173. Oxford: Basil Blackwell and Berkeley: University of California Press, 1978.
Frankena, William K. 1973. *Ethics*. 2nd ed. Englewood Cliffs: Prentice-Hall.
Franklin, James. 2001. *The science of conjecture: Evidence and probability before Pascal*. Baltimore: The Johns Hopkins University Press.
Gauthier, David. 1986. *Morals by agreement*. Oxford: Clarendon Press.
Gert, Bernard. 1998. *Morality: Its nature and justification*. New York: Oxford University Press.
Gert, Bernard. 2004. *Common morality: Deciding what to do*. New York: Oxford University Press.
Hausman, Daniel M. 1991. Is utilitarianism useless? *Theory and Decision* 30:273–278.
Hempel, Carl G. 1965. *Aspects of scientific explanation*. New York: The Free Press and London: Collier-Macmillan.
Hume, David. 1739. *A treatise of human nature*. Oxford: Clarendon Press, 1973.
Irvine, William B. 2006. *On desire: Why we want what we want*. Oxford: Oxford University Press.
Jeffrey, Richard C. 1992. *Probability and the art of judgment*. Cambridge: Cambridge University Press.
Kuipers, Theo A. F. 2000. *From instrumentalism to constructive realism*. Dordrecht: Kluwer Academic Publishers.
Larmore, Charles. 2008. *The autonomy of morality*. Cambridge: Cambridge University Press.

Laudan, Larry. 1984. *Science and values: The aims of science and their role in scientific debate*. Berkeley: University of California Press.
Lehrer, Keith. 2000. *Theory of knowledge*. 2nd ed. Boulder: Westview.
Lewis, C. I. 1946. *An analysis of knowledge and valuation*. La Salle: Open Court.
Mackie, J. L. 1977. *Ethics: Inventing right and wrong*. Harmondsworth: Penguin.
Maimonides, Moses. 1190. *Dalālat al-ḥā'irīn*. English edition: Maimonides, Moses. 1963. *The guide of the perplexed*. Chicago: University of Chicago Press.
Marx, Karl. 1939. *Grundrisse der kritik der politischen ökonomie (Rohentwurf) 1857–1858*. Moskau: Verlag für Fremdsprachige Literatur. English excerpts: Marx, Karl. 1977. Grundrisse. In *Karl Marx: Selected Writings,* ed. David McLellan, 345–387. Oxford: Oxford University Press.
McGee, Vann. 1985. A counterexample to modus ponens. *The Journal of Philosophy* 82:462–471.
Moretti, Luca. 2007. Ways in which coherence is confirmation conducive. *Synthese* 157:309–319.
Morton, Adam. 1991. *Disasters and dilemmas*. Oxford: Basil Blackwell.
National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. 1979. *The Belmont report*. Office of Human Subjects Research. http://www.hhs.gov/ohrp/humansubjects/guidance/belmont.html. Accessed 30 April 2014.
Olsson, Erik J. 2005a. The impossibility of coherence. *Erkenntnis* 63:387–412.
Olsson, Erik J. 2005b. *Against coherence: Truth, probability, and justification*. Oxford: Oxford University Press.
Ortega y Gasset, José. 1936. History as a system. In *Philosophy and history: Essays presented to Ernst Cassirer,* eds. Raymond Klibansky and H. J. Paton, 283–322. Oxford: The Clarendon Press. In *History as a system and other essays toward a philosophy of history,* 165–233. New York: Norton, 1961.
Pigozzi, Gabriella. 2009. Interview with John Woods. *The Reasoner* 3 (3): 1–4.
Plato. *Republic*. In *Complete works*, ed. John M. Cooper, 971–1223. Indianapolis: Hackett, 1997.
Polya, George. 1954. *Mathematics and plausible reasoning*. Princeton: Princeton University Press.
Prichard, H. A. 1912. Does moral philosophy rest upon a mistake? In *Moral obligation: Essays and lectures,* 1–17. Oxford: Clarendon Press, 1949.
Pryor, James. 2000. The skeptic and the dogmatist. *Nous* 34:517–549.
Rawls, John. 1971. *A theory of justice*. Cambridge: The Belknap Press of Harvard University Press.
Sayre-McCord, Geoffrey. 1985. Coherence and models for moral theorizing. *Pacific Philosophical Quarterly* 66:170–190.
Scheid, Don E. 1983. Kant's retributivism. *Ethics* 93:262–282.
Schupbach, Jonah N. 2008. On the alleged impossibility of Bayesian coherentism. *Philosophical Studies* 141:323–331.
Searle, John R. 1969. *Speech acts: An essay in the philosophy of language*. Cambridge: Cambridge University Press.
Shogenji, Tomoji. 1999. Is coherence truth conducive? *Analysis* 59:338–345.
Shulman, Ken. 1993. AIDS without tears. *Newsweek,* Atlantic edition, August 23:60.
Sidgwick, Henry. 1907. *The methods of ethics*. 7th ed. London: Macmillan.
Simon, Herbert A. 1983. *Reason in human affairs*. Oxford: Basil Blackwell.
Sinnott-Armstrong, Walter. 2006. *Moral skepticisms*. Oxford: Oxford University Press.
Sorensen, Roy A. 1990. Vagueness implies cognitivism. *American Philosophical Quarterly* 27:1–14.
Welch, John R. 2014. Plausibilistic coherence. *Synthese* 191:2239–2253. doi:10.1007/s11229-013-0395-9.

Chapter 6
Remedies for Reflective Disequilibrium

Abstract Chapter 6 proposes remedies for a common affliction: reflective disequilibrium. This affliction can result from inconsistencies within moral strata or between moral and nonmoral discourse. The chapter claims that reflective disequilibrium within the phenomenal stratum can be reduced by appeal to the standard of inductive cogency. Reflective disequilibrium within the instrumental stratum, which is illustrated by the classic case of *United States v. Holmes*, can be intra-theoretic or inter-theoretic. Intra-theoretic instrumental disequilibrium can sometimes be resolved by judicious use of moral theory, while the inter-theoretic variety typically requires teleological ascent. Like instrumental disequilibrium, teleological disequilibrium can be intra-theoretic or inter-theoretic. Inter-theoretic cases can be managed with the resources of comparative decision theory. While intra-theoretic cases can be more recalcitrant, they may nonetheless become tractable over time through increased understanding of consequences of alternative moral ends. Finally, reflective disequilibrium can also arise through conflict between moral and nonmoral discourse. Citing the conflict between Gauguin's commitments to his family and his art, the chapter maintains that extra-moral disequilibrium can sometimes be ameliorated by adhering to an overridingness thesis stated in terms of supererogation and moral obligation.

6.1 Reflective Disequilibrium

The backdrop to the explorations of morally phenomenal, instrumental, and teleological discourse in this work has been the quest for reflective equilibrium among these strata and background theories. Not infrequently, however, the quotidian reality of our moral discourse is reflective *dis*equilibrium. Some elements of our moral discourse turn out to be inconsistent with others. The question I would like to address in this chapter is how we might find relief from the disturbances of this condition. Once we realize we are in such a state, how might we exchange reflective disequilibrium for reflective equilibrium?

A glance at some recent reflections on reflective equilibrium will serve as our point of departure. Nelson Goodman's views are pivotal, as we noticed in Sect. 1.2. In discussing the justification of general rules and particular inferences in logic, Goodman remarks, "The process of justification is the delicate one of making

J. R. Welch, *Moral Strata*, Theory and Decision Library A:, 49,
DOI 10.1007/978-3-319-08013-0_6,

mutual adjustments between rules and accepted inferences; and in the agreement achieved lies the only justification needed for either" (1979, p. 64). John Rawls cites this passage in developing his notion of reflective equilibrium as a "process of mutual adjustment of principles and considered judgments" that "is not peculiar to moral philosophy" (1971, p. 20 n. 7).

Other philosophers influenced by Goodman and Rawls have followed suit. Hilary Putnam emphasizes the "dialectic" between general principles and particular judgments in ethics and indeed throughout philosophy:

> In ethics we start with judgments that individual acts are right or wrong, and we gradually formulate maxims (*not* exceptionless generalizations) based on those judgments, often accompanied by reasons or illustrative examples, as for instance 'Be kind to the stranger among you, because you know what it was like to be a stranger in Egypt'. These maxims in turn affect and alter our judgments about individual cases, so that new maxims supplementing or modifying the earlier ones may appear. After thousands of years of this dialectic between maxims and judgments about individual cases, a philosopher may come along and propose a moral conception, which may alter both maxims and singular judgments and so on.
>
> The very same procedure may be found in all of philosophy (which is almost coextensive with theory of rationality). (1983, pp. 201–202)

Similarly emphasizing the dialectic of particular judgments and general principles, Catherine Elgin takes reflective equilibrium to be the standard of rational acceptability for any system of thought:

> We proceed dialectically. We mold specific judgments to accepted generalizations, and generalizations to specific judgments. We weigh considerations of value against antecedent judgments of fact. We synchronize ends and means, reconcile principle and practice. A process of delicate adjustments occurs, its goal being a system in reflective equilibrium. (1996, p. 106)

Henry Richardson defends a conception of extended reflective equilibrium that builds on Rawls' notion by admitting the data of emotion and perception (1986, 1994). Like Rawls, however, he admits "bidirectional" justification that proceeds "from the specific to the general or in the other direction" (1994, pp. 176–177). Martha Nussbaum recommends an ideal of perceptive equilibrium akin to Richardson's extended reflective equilibrium. In perceptive equilibrium, "concrete perceptions 'hang beautifully together,' both with one another and with the agent's general principles" (1990, pp. 182–183). Frances Kamm subscribes to a version of reflective equilibrium that begins with one's own case-based judgments, seeks a principle that would account for them, and then considers whether the principle can be justified or not (2007, p. 5). If not, adjustments must be made to the case-based judgments, the principle, or both.

As a proponent of reflective equilibrium, I am generally sympathetic to these points of view. But I confess to a certain dissatisfaction with them all. This dissatisfaction has a double root. Nobody seems to have much to say about how to make the adjustments required for reflective equilibrium, first of all. "When theory conflicts with what is taken to be fact, we sometimes give up the theory and sometimes give up the 'fact'," as Putnam observes (1992, p. 137). Yes, but how? Surely this is not a stochastic process. Some philosophers seem to have a standing preference for fact.

Socrates rejects Cephalus' abstract definition of justice in Book I of the *Republic* by relying on a concrete instance, as Martha Nussbaum reminds us (1997, p. 38). Yet philosophers like David Gauthier favor theory:

> We shall find no simple fit, or lack of fit, between our theory and the supposedly "plain duties" of conventional morality. Trusting theory rather than intuition, we should advocate the view of social relationships sketched in this chapter without regard to the intellectual fashions of the moment. If the reader is tempted to object to some part of this view, on the ground that his moral intuitions are violated, then he should ask what weight such an objection can have, if morality is to fit within the domain of rational choice. (1986, p. 269)

More recently, the same divergence can be noted in the work of Frances Kamm, who trusts our intuitions about moral facts, and Peter Singer, who places his faith in theory (Voorhoeve 2009, pp. 20, 50–51). But why should theory take precedence over fact? Or why should fact prevail over theory? How should we decide what to reject and what to retain?

The second reason for dissatisfaction is that the nature of reflective disequilibrium is often loosely described. To state that there is a conflict between principles and specific judgments or between theory and fact is not sufficiently explicit. Suppose that a principle embedded in a moral theory has the form 'All F are G' and a specific moral judgment takes the form '$Fa \wedge -Ga$'. Roughly speaking, the principle and the judgment are inconsistent, but the root of the inconsistency is a contradiction between statements of the same type. 'All F are G' (the principle) and 'Fa' (implied by the judgment) jointly imply 'Ga', but the judgment '$Fa \wedge -Ga$' implies '$-Ga$'. Hence we have 'Ga' and '$-Ga$'. Stating the inconsistency as explicitly as possible requires reference to the same discursive type. Hence I will refer to disequilibria that result from conflicts between phenomenal and phenomenal, instrumental and instrumental, or teleological and teleological discourse, but not to conflicts between phenomenal and teleological discourse, for instance. Rivals, in order to be rivals, must vie for the same place.

This chapter examines reflective disequilibrium in its different forms and explores various possibilities of relief. Sections 6.2–6.4 address phenomenal, instrumental, and teleological disequilibrium in that order. These forms of disequilibria are all internal to moral discourse, but disequilibrium can also arise through conflict between moral and nonmoral discourse. Section 6.5 develops the theme of extramoral disequilibrium by drawing on the concepts of supererogation and moral obligation.

6.2 Phenomenal Disequilibrium

To explore the idea of conflict within the phenomenal stratum, I want to peruse an initial example of a nonmoral sort.[1] The anomalous accelerations of the Pioneer, Galileo, and Ulysses space probes were noted in passing in Sect. 3.2. John Anderson

[1] This example was treated in preliminary fashion in Welch (2013, pp. 325–326).

of the Jet Propulsion Laboratory and his team of investigators have been unable to explain why the probes are slowing down slightly more than predicted (1998). They have considered the possibility of error in the observed accelerations, the descriptions of the probes' initial conditions, the general theory of relativity, and background assumptions like the principle of equivalence (the Newtonian assumption that gravity affects all objects equally). To the best of their knowledge, none of these is mistaken—yet the anomaly persists (Turyshev and Toth 2010).

The Anderson team is in a state of reflective disequilibrium. Something has gone wrong, but what? The investigating physicists are decidedly not neutral on the question of where they expect the error to be found. They note "it is interesting to speculate on the possibility that the origin of the anomalous signal is new physics. This is true even though the probability is that some 'standard physics' or some as-yet-unknown systematic [initial conditions of the system] will be found to explain this 'acceleration'" (Anderson et al. 1998, p. 2860). In the same vein, they remark "we feel that some systematic or combination of systematics (such as heat or gas leaks) will most likely explain the anomaly" (Anderson et al. 1999, p. 1891). Similarly, "Until more is known, we must admit that the most likely cause of this effect is an unknown systematic. (We ourselves are divided as to whether 'gas leaks' or 'heat' is this 'most likely cause'.)" (Anderson et al. 2002, p. 44). In other words, these physicists anticipate that the error will turn out to lie in the description of the probes' initial conditions.

To interpret their language ("the probability is," "most likely explain," "most likely cause"), we recall that the term 'probability' is ambiguous. In addition to the strict mathematical sense of the probability calculus, the term has an ordinary language sense roughly synonymous with 'plausibility' (cf. Sect. 5.4). Anderson and his colleagues appear to be relying on this ordinary language sense, and the following discussion proceeds on this assumption. But the argument could easily be refocused from the general case of plausibility to the special case of probability.

I submit that Anderson and his colleagues are implicitly appealing to a comparative decision-theoretic model of cognitive choice. In fact, they appear to be focusing on a special case described by this model. Consider a generic choice between acts whose outcome utilities are equal, as in cases 3, 6, and 9 from Table 3.4 (Sect. 3.5.1). In such a case, the decisive considerations in choosing how to act are plausibilistic. More specifically, since information is the only relevant outcome in purely cognitive decisions, where the information afforded by the options under consideration is equal, the utilities of information outcomes are equal as well. Relative plausibility therefore tips the scales.

That the equal information condition is satisfied in the case of the space probes follows directly from Sect. 3.3.2.3's gloss of information as reduction of uncertainty about attributable states of the world. To see this, let us consider four possible answers to the question 'What went wrong?' in the case of the space probes: observations, initial conditions, the general theory, and background assumptions. Each of these answers is actually a complex disjunction, but since the question is being posed in general terms (in terms of the initial conditions as a whole, for example, rather than a specific initial condition), finer-grained analysis is beside the

point. These four answers are not the only possibilities, however. More than one of these alternatives could go wrong, so their possible combinations must be considered as well. Assuming that something really has gone wrong, there are 4 unary + 6 binary + 4 ternary + 1 quaternary = 15 possible answers to the question 'What went wrong?'. Provided quantity of information *i* is equal to the ratio of eliminated options to total options, *i* is 14/15 for each of the fifteen possible answers; that is, each excludes fourteen of the fifteen possibilities. Thus they each provide the same amount of information.

If the information afforded by these answers is equal and the decision is purely cognitive, the utilities of these information outcomes should be equal as well, as we have noted. Hence comparative decision theory would require that successive binary comparisons among the answers be based on their relative plausibilities. That is exactly what the Anderson team was doing. In their judgment, the plausibility that the problem lies in the description of the probes' initial conditions is greater than that of any other possible answer. Where 't_1' represents the answer about initial conditions and 't_2' any of the other fourteen answers, the Anderson team would be implicitly relying on instances of case 6 from Table 3.4 (Sect. 3.5.1): greater plausibility for t_1, equal utility, hence t_1.

The claim that the fifteen possible answers are equally informative needs to be carefully understood. In particular, it should be understood against the backdrop of pragmatic views of information (e.g., Levi 1984). According to these views, information depends on context. Take the statement that there is a grocery store on the corner, for example. This statement may be informative if I do not know this but uninformative if I do. If quantity of information *i* is equal to the ratio of eliminated options to total options and *n* is the number of possible locations of the nearest grocery store, *i* for 'There is a grocery store on the corner' is $n-1/n$ if I do not know this but $0/n$ if I do.

With this pragmatic backdrop in mind, we note that the contexts of theory use and theory mention are importantly different. Take theory use, for example, where fourteen of fifteen diagnostic options are implausible but the implausible fourteenth, not the plausible fifteenth, turns out to hold. In one sense, I learn more from this discovery than from learning that the fifteenth option holds because I can now begin to use the fourteenth to obtain much new information. Before this can happen, however, I must mention the theory in the context established by the question 'Which of these fifteen options holds?'. In this prior context of theory mention, I have not learned more from mentioning the implausible fourteenth option rather than the plausible fifteenth, for each of the fifteen options would answer the question as well as the rest. Hence each is as informative as the rest. This is not to deny that, in the subsequent step of using the theory to explain or predict, the fifteen options are not equally informative. The equal-information claim about the theory and its alternatives applies only to their mention.

Of course, the information space corresponding to the question 'What went wrong?' can be partitioned differently. In the Pioneer case, for example, the ternary and quaternary possibilities might be fused as a five-clause disjunction, which would leave ten possible answers instead of fifteen. But the fact that there are other

possible partitions is not an obstacle to applying individual decision theory, which here accepts the inputs of the actual decision makers as givens. To ignore the Anderson team's partition and impose my own would be to abandon a real decision for a fictional one. More importantly, even if alternative partitions are admitted and they do yield different numbers of possible answers, the equal-information claim would still hold. Suppose that one partition includes fifteen mutually exclusive answers while another includes ten. Although each possibility in the fifteen-answer partition has an information value of 14/15 and each in the ten-answer partition has an information value of 9/10, each is as informative as any other within the partition. The inter-partition variation of information values like 14/15 and 9/10 is reminiscent of what Quine used to call "don't-cares" (1960, p. 182, 259). That is, the relevant relations are intra-partition, not inter-partition, for the question 'What went wrong?' is answered relative to a partition.

If comparative decision theory is the key to correcting reflective disequilibrium in the case of the space probes, could it also be used to remedy reflective disequilibrium in ethics? Imagine a half-hearted utilitarian who realizes that utilitarian theory has phenomenal implications that conflict with some considered phenomenal judgments. Say that our utilitarian is inclined to the view that executing an innocent person to save greater loss of life is just, since it maximizes utility and utility implies justice. But she is also concerned that the same action could be unjust because it violates the right to life. She is therefore in a state of phenomenal disequilibrium.

Since something has obviously gone wrong, the question is what. The team investigating the space probes considered the possibility of four types of error, and our utilitarian might do much the same thing. Perhaps the "observation," the phenomenal description of the action as unjust, is false. Maybe the description of initial conditions, the concrete circumstances of agent, place, time, etc. that configure the action and its context, is inadequate. Alternatively, the theoretical principles of utilitarianism could be at fault. Or background assumptions such as the logic that leads from utilitarian principles to conclusions might be defective. Just as in the case of the space probes, these alternatives are not the only possibilities; any combination of them could be the source of the problem. As before, then, under the assumption that something has gone wrong, there are fifteen generic possibilities. Each of these possibilities is an equally informative answer to the question 'What went wrong?', for each excludes the other fourteen. The fact that other partitions with different numbers of possible answers could be formulated does not belie the equal-information claim. As we have just noted, the question 'What went wrong?' is answered relative to a partition, and a pragmatic view of information ensures that each mutually exclusive answer within a partition is equally informative. Now if the information outcomes of a purely cognitive decision are equal, the utilities associated with each option should be equal. That places the weight of the decision solely on the relative plausibilities of the options—just like the case of the space probes.

How might the key plausibilities be determined? This is a crucial question since decision theory, if conceived along the lines of Patrick Maher's unqualified Bayesianism, provides no check on the plausibility and utility inputs that generate expectation outputs (1993, p. 29). They can be subjective and even arbitrary. But the

normative version of comparative decision theory deployed in these pages is not as permissive. Like Maher's qualified Bayesianism (1993, p. 29), it permits further rational constraints on utility and plausibility inputs. We have already required that cognitive decisions admit information as a possible outcome and that epistemic utilities be proportional to information. Additional constraints can be reasonably imposed on plausibilities. The trick is to draw on the standard of inductive cogency outlined in Sect. 2.4.1.

As an illustration, consider a simplified version of our half-hearted utilitarian's predicament in which the possibility of error in describing initial conditions and background assumptions is regarded as negligible. Our utilitarian believes that utilitarian theory is false or the description of the action as unjust is false but that theory and description are not both false. She may reason, on the one hand, as follows: if the theory is right, then the action is just; the theory is right; hence the action is just. This would license the inference that what went wrong was the description of the action as unjust. But one person's modus ponens is another's modus tollens. Hence she may also reason that if the theory is right, the action is just; the action is not just; hence the theory is wrong. This chain of reasoning places the blame on the theory, not the description. So what's a poor reasoner to do?

In assessing the plausibility of act-utilitarianism, Sects. 4.6.1 and 5.5.1 both appealed to standard objections that act-utilitarianism licenses injustice, insincere promises, racist and sexist pleasures, and other forms of immorality. Both sections postponed consideration of the act-utilitarian counter that these objections assume the verdicts of common morality uncritically. It is time to address this counter. If an act-utilitarian insists that a common-morality classification of an action as unjust is mistaken, that the action is in fact just, we have a case of phenomenal disequilibrium.

The path to phenomenal equilibrium, I suggest, passes through the classificatory procedures of Chap. 2. The disequilibrium's fulcrum is the description of the action. One chain of reasoning (act-utilitarianism's modus ponens) classifies it as just; the other (common morality's modus tollens) classifies it as unjust. According to the analogy thesis of Sects. 2.2–2.3, positive core classification using thick terms like 'just' is by analogy, and negative core classification with terms like 'unjust' is by disanalogy. If this is so, any hope for breaking out of the impasse lies in our ability to distinguish good and bad arguments by analogy. Since arguments by analogy appeal to an inductive standard of argumentation, the standard for inductive cogency proposed in Sect. 2.4.1 can be applied to the case. Imagine that done and that the results favor describing the action as just. Our utilitarian would then have reason to think that the other suspect, the contradictory description of the action, is probably the culprit. But if the inductive standard favors the analogy leading to the conclusion that the action is unjust, utilitarian theory is probably at fault. I am under no illusions about the practical difficulties of carrying out these procedures. Appeal to the standard of inductive cogency is, as I have urged in Sect. 2.6, an in-principle solution to the problem of vagueness, and the obstacles to putting it into practice can be formidable. But I know of no reason to think them insuperable.

6.3 Instrumental Disequilibrium

Reflective disequilibrium appears in many guises. We have seen how it can emerge within the phenomenal stratum of moral discourse. We will now take up disequilibrium within the instrumental stratum. Instrumental disequilibrium can be intra-theoretic, when one theory renders conflicting instrumental judgments, or it can be inter-theoretic, when the instrumental conflict is generated by more than one theory.

To illustrate both types of instrumental disequilibrium, I will draw on a famous case in American law: *United States v. Holmes*.[2] Holmes, a seaman, was charged with manslaughter in the maritime death of Francis Askin. The case was tried in the U. S. Circuit Court of the Eastern District of Pennsylvania in 1842. Here are some excerpts from the court record:

> The American ship William Brown left Liverpool on the 13th of March, 1841, bound for Philadelphia, in the United States. She had on board (besides a heavy cargo) 17 of a crew, and 65 passengers, Scotch and Irish emigrants. About 10 o'clock on the night of the 19th of April, when distant 250 miles southeast of Cape Race, Newfoundland, the vessel struck an iceberg, and began to fill so rapidly that it was evident she must soon go down. The long-boat and jolly-boat were cleared away and lowered. The captain, the second mate, 7 of the crew, and 1 passenger got into the jolly-boat. The first mate, 8 seamen, of whom the prisoner [Holmes] was one (these 9 being the entire remainder of the crew), and 32 passengers, in all 41 persons, got indiscriminately into the long-boat. The remainder of the passengers, 31 persons, were obliged to remain on board the ship. In an hour and a half from the time when the ship struck, she went down, carrying with her every person who had not escaped to one or the other of the small boats. Thirty-one passengers thus perished. On the following morning (Tuesday) the captain, being about to part company with the long-boat, gave its crew several directions, and, among other counsel, advised them to obey all the orders of the mate, as they would obey his, the captain's. This the crew promised that they would do. The long-boat was believed to be in general good condition; but she had not been in the water since leaving Liverpool, now 35 days; and as soon as she was launched, began to leak. She continued to leak the whole time; but the passengers had buckets, and tins, and, by bailing, were able to reduce the water, so as to make her hold her own. The plug was about an inch and a half in diameter. It came out more than once, and finally, got lost; but its place was supplied by different expedients.
>
> It appeared by the depositions of the captain, and of the second mate, (the latter of whom had followed the sea 21 years; the former being, likewise, well-experienced), that on Tuesday morning when the two boats parted company, the long-boat and all on board were in great jeopardy…. And… before the boats parted company, the mate, in the long-boat, told the captain, in the jolly-boat, that the long-boat was unmanageable, and, that unless the captain would take some of the long-boat's passengers, it would be necessary to cast lots and throw some overboard. "I know what you mean," or, as stated by one witness, "I know what you'll have to do," said the captain. "Don't speak of that now. Let it be the last resort." There was little or no wind at this time, but pieces of ice were floating about.
>
> Notwithstanding all this, the long-boat, loaded as she is above described to have been [including "provisions for 6 or 7 days"], did survive throughout the night of Monday, the day of Tuesday, and until 10 o'clock of Tuesday night,—a full twenty-four hours after the ship struck the iceberg. The crew rowed, turn about, at intervals, and the passengers bailed. On Tuesday morning, after the long-boat and jolly-boat parted, it began to rain, and continued to rain throughout the day and night of Tuesday. At night the wind began to freshen, the

[2] Hugo Bedau brought this case to my attention long ago.

> sea grew heavier, and once, or oftener, the waves splashed over the boat's bow so as to wet, all over, the passengers who were seated there. Pieces of ice were still floating around, and, during the day, icebergs had been seen. About 10 o'clock of Tuesday night, the prisoner and the rest of the crew began to throw over some of the passengers, and did not cease until they had thrown over 14 male passengers. These, with the exception of two married men and a small boy, constituted all the male passengers aboard. Not one of the crew was cast over.…
> None of [the witnesses from the long-boat] spoke in a manner entirely explicit and satisfactory in regard to the most important point, viz. the degree and imminence of the jeopardy at 10 o'clock on Tuesday night, when the throwing over began. As has been stated, few words were spoken. It appeared, only, that, about 10 o'clock of Tuesday night, it being then dark, the rain falling rather heavily, the sea somewhat freshening, and the boat having considerable water in it, the mate, who had been bailing for some time, gave it up, exclaiming: "This work won't do. Help me, God. Men, go to work." Some of the passengers cried out, about the same time: "The boat is sinking. The plug's out. God have mercy on our poor souls." Holmes and the crew did not proceed upon this order; and after a little while, the mate exclaimed again: "Men, you must go to work, or we shall all perish." They then went to work; and, as has been already stated, threw out, before they ended, 14 male passengers, and also 2 women [Francis Askin's sisters, who may or may not have gone over voluntarily]. The mate directed the crew "not to part man and wife, and not to throw over any women." There was no other principle of selection. There was no evidence of combination among the crew. No lots were cast, nor had the passengers, at any time, been either informed or consulted as to what was now done. Holmes was one of the persons who assisted in throwing the passengers over.…
> On Wednesday morning the weather cleared, and early in the morning the long-boat was picked up by the ship "Crescent." All the persons who had not been thrown overboard were thus saved. (*United States v. Holmes* 1842)

In order to advance our discussion of instrumental disequilibrium, I propose to treat this case from the standpoint of Frankenian ethical theory. As noted in Sect. 4.6.1, Frankena recognizes two basic principles: beneficence and justice. Had the first mate on the *William Brown*'s long-boat been a Frankenian, he might have reasoned along the following lines. Although throwing people overboard would cause evil, it would likely prevent the greater evil of all of us being lost; hence beneficence requires that some people be thrown overboard. But to throw people overboard would sacrifice some solely for the benefit of others; since justice requires equal treatment, we should not throw anyone overboard. The first mate would then be in a state of instrumental disequilibrium. The principle of beneficence implies the instrumental directive to throw some people overboard; the principle of justice implies the instrumental directive to refrain. Unlike the phenomenal disequilibrium of the half-hearted utilitarian, which arose from a conflict between a theoretical implication and an extra-theoretic judgment, the first mate's instrumental disequilibrium is generated by theoretical implications of a single theory. This is the intra-theoretic case.

Fortunately, Frankena does not just recommend the goals of beneficence and justice. He proposes a tentative ranking whereby justice usually comes first. Nevertheless, he advises us to consider the possibility of putting beneficence before justice in case a small injustice would avoid a great evil or obtain a great good. Are these conditions satisfied in the case of *United States v. Holmes*? In fact, they are not. To throw some overboard so that others may survive is a clear violation of equality; in fact, it is a paradigm of great injustice. According to Frankenian principles, then, justice would have to come first.

Even so, this does not necessarily mean that no one could be thrown overboard. Attentive readers of the case will have noticed that what was finally done did not correspond to the captain's and first mate's initial conversation on the matter: "it would be necessary to cast lots and throw some overboard." At the crucial moment, no lottery was used to select those to be thrown overboard: "No lots were cast, nor had the passengers, at any time, been either informed or consulted as to what was now done." Had lots been cast and the passengers been informed or consulted, might not some have been thrown overboard in accordance with the principle of justice? Since fair lotteries require equal treatment, justice as equality would arguably have been served. Informed participants in the lottery would have been treated with respect, dealt with as human beings instead of as ballast.

The instrumental disequilibrium we have been considering is internal to Frankenian theory. Happily, as it turns out, the theory is up to the task of resolving it. But instrumental disequilibrium can also be inter-theoretic; it can arise when two or more theories are applied to the same case. An act-utilitarian reasoning about *United States v. Holmes* would derive an instrumental directive to throw some people overboard, but a Kantian relying on the categorical imperative in the formulation of the end in itself would deny this directive. The obvious maneuver for dealing with disequilibrium of this sort is reflected by the older, hierarchical model of justification described in Sect. 1.4: teleological ascent. That is, to deal with inconsistency in the instrumental stratum, we move up to teleology. The disequilibrium persists within this higher level, of course, but that leads us directly to the following section.

6.4 Teleological Disequilibrium

Like instrumental disequilibrium, teleological disequilibrium can be intra-theoretic or inter-theoretic. Since the conclusion of the preceding section feeds directly into the inter-theoretic case, I will begin with it before proceeding to the intra-theoretic case.

Suppose that, in an effort to resolve the impasse between Kantian and act-utilitarian instrumental directives about *United States v. Holmes*, we resort to teleological ascent and address the conflict there. We would confront the teleological directive to act for the greatest happiness of the greatest number with the teleological directive to act with a good will. How to deal with conflicts of this sort was the subject of Sect. 5.5, where I argued for two theses that are apposite here. One was the general claim that comparative decision theory is capable of guiding us to reasonable decisions about teleological directives. The second was the specific claim that the plausibilistic expectation of the Kantian teleological directive is greater than that of its act-utilitarian counterpart. If these theses are acceptable, we have an illustration of how to resolve inter-theoretic teleological disequilibrium.

That would leave us with the intra-theoretic case. To find an example, let us return to Frankenian theory. Frankena does not think he can propose a simple recipe such as 'Always put justice first' for resolving conflicts of justice and beneficence.

But successful resolution of these conflicts on a case-by-case basis is possible, he believes: "One can only hope that, if we take the moral point of view, become clear-headed, and come to know all that is relevant, we will also come to agree on ways of acting that are satisfactory to all concerned" (1973, p. 53). Until such time as a case-specific solution is found, however, Frankenian moral agents can experience teleological disequilibrium of an intra-theoretic sort.

All too plainly, agents so afflicted may have a long wait. An individual moral agent or a group of moral agents might find that reflective disequilibrium resists their best efforts to surmount it. The reasons are not hard to find. As Sect. 5.5 pointed out, the Frankenian teleological directives 'You ought to be just' and 'You ought to be beneficent' are also means to a further end: to act morally. Because they are means, their relative priorities in a given situation could be justified in principle by their relative effectiveness in achieving their end. But the end of acting morally, as Sect. 5.5 also pointed out, is not as clear as we would like it to be. At a given moment in time, then, lack of clarity about what acting morally is may prevent an instrumental justification that is possible in principle from actually being carried out.

Even here, though, all is not lost. One way of summarizing Chap. 2's take on vagueness is to say that clarity is not given; it has to be won. Hence what is unclear at one point in time need not remain unclear at another. Nevertheless, I am not sanguine about the prospects for applying the analogical tactics of Chap. 2 to clarify "thin" concepts like acting morally. But there is another—perhaps longer—route. Frankena provides a clue:

> It seems to me that everyone who takes the moral point of view can agree that the ideal state of affairs is one in which everyone has the best life he or she is capable of. Now, in such a state of affairs, it is clear that the concerns of both the principle of justice or equality and the principle of beneficence will be fulfilled. If so, then we can see that the two principles are in some sense ultimately consistent, and this seems to imply that increasing insight may enable us to know more and more how to solve the conflicts that trouble us now when we know so little about realizing the ideal state of affairs in which the principles are at one. (1973, p. 53)

From the standpoint of the present, "when we know so little about realizing the ideal state of affairs" in which justice and beneficence are in equilibrium, Frankena gestures toward "increasing insight" that "may enable us to know more and more how to solve the conflicts that trouble us now."

How might we obtain this increasing insight? One technique would be to milk the social sciences and humanities for information on the consequences of pursuing diverse ethical goals. Deeper insight into the contrasting conceptions of justice embedded in capitalist and socialist societies could be immensely important for a more moral future. A full 55 years before the Bolshevik Revolution, Victor Hugo included a series of astoundingly prescient observations on the subject in *Les Miserables*:

> It will not surprise the reader that, for a variety of reasons, we do not here proceed to a profound theoretical examination of the questions propounded by socialism. We will simply indicate what they were.
> Problem One: the production of wealth.

> Problem Two: its distribution.
> Problem One embraces the question of labour and Problem Two that of wages, the first dealing with the use made of manpower and the second with the sharing of the amenities this manpower produces….
> England has solved the first of these problems. She is highly successful in creating wealth, but she distributes it badly. This half-solution brings her inevitably to the two extremes of monstrous wealth and monstrous poverty. All the amenities are enjoyed by the few and all the privations are suffered by the many, that is to say, the common people: privilege, favour, monopoly, feudalism, all these are produced by their labour. It is a false and dangerous state of affairs whereby the public wealth depends on private poverty and the greatness of the State is rooted in the sufferings of the individual: an ill-assorted greatness composed wholly of materialism, into which no moral element enters.
> Communists and agrarian reformers believe they offer the solution to the second of these problems. They are mistaken. Their method of distribution kills production: equal sharing abolishes competition and, in consequence, labour. It is distribution carried out by a butcher, who kills what he distributes. It is impossible to accept these specious solutions. To destroy wealth is not to share it.
> These two problems must be solved together if they are to be properly solved, and the two solutions must form part of a single whole. ([1862] 1982, iv.i.iv, pp. 722–723)

Note that Hugo's Problem One, the production of wealth, calls for an economic variety of beneficence, and his Problem Two, the distribution of wealth, demands an economic sort of justice. These two problems (and their generalized Frankenian forms) create teleological disequilibrium of an obdurate sort. Such disequilibrium is unlikely to be resolved by a single individual. But individuals can act in such a way that their efforts, if combined with others—many others, in all probability, over long periods of time—can facilitate the efforts of future generations to grapple with the problem more insightfully than we can. Hence even if we are unable to resolve intra-theoretic teleological disequilibrium at present, there are some prospects of being able to resolve it in the future. This, I submit, is a worthy answer to the Kantian question 'What may I hope?'.

6.5 Extra-Moral Disequilibrium

The preceding three sections have dealt with reflective disequilibrium as it arises in different moral strata. But disequilibrium can also arise as a conflict between moral and nonmoral concerns. Morality can appear to clash with self-interest, as in Cicero's grain merchant of Sect. 2.5.2. Moral and scientific ends may tug in opposite directions, as in the purposely deceptive experiments designed by Stanley Milgram (1974). Morality may diverge from legality, as those who violate the law on moral grounds well know. And moral endeavors like relieving famine and comforting the sick lose momentum whenever resources are diverted to personal pursuits like bungee jumping or reading Roethke.

One approach to resolving disequilibria between the moral and the nonmoral is to always put the nonmoral first. However popular this approach may be, it is clearly unacceptable. As a result, some have veered off in the opposite direction and urged that we put the moral first. The view that we should always do this is known as the overridingness thesis: whenever moral and nonmoral concerns conflict, the

moral should override the nonmoral. The overridingness thesis can take two different forms (Flanagan 1986, p. 51). The morally ideal form would require that what is morally ideal should always override the nonmoral; the demands of becoming a saint, say, should trump nonmoral goods like personal enjoyment. Less strenuously, the morally required form of the thesis would stipulate that what is morally required should always override the nonmoral.

The morally ideal version almost invariably receives short shrift. As Owen Flanagan remarks, "it is something of a mystery how the thesis of the overridingness of the morally ideal falls so easily to our realistic attitudes about persons while the thesis of the overridingness of the morally required stands so imperiously over moral life" (1986, p. 52). That the overridingness of the morally required stands imperiously over moral life is appropriate, I think. It appeals to us as a sensible mean between two repellent extremes: downgrading morality by elevating the nonmoral ('Always put the nonmoral first') and exaggerating morality by making the morally ideal obligatory ('Always put the morally ideal first'). To always put the morally required first is to grant that sometimes the moral should come first, while sometimes the nonmoral should.

I want to explore this intermediate position as a way of addressing the relative priorities of the moral and nonmoral. In addition to the concept of the morally required (or morally obligatory), the concept of supererogation is essential. Together, these concepts afford a straightforward general answer to the question about the relative priorities of the moral and nonmoral.[3] The main idea can be summarized in two conditionals:

C1: If the action is supererogatory, the nonmoral can (though need not) trump the moral.
C2: If the action is morally obligatory, the moral should take precedence over the nonmoral.

In the case of Cicero's grain merchant, for example, if the moral option would be to tell the Rhodians about the other ships and the nonmoral (economic) option would be not to tell, the merchant should tell if telling is morally obligatory but could refrain from telling if telling is supererogatory. But making this straightforward solution workable requires manageable concepts of supererogation and moral obligation. The following sections develop these concepts. Section 6.5.1 considers supererogation; 6.5.2 deals with moral obligation; and 6.5.3 undertakes a modest defense of the overridingness of the morally required.

6.5.1 Supererogation

The idea that some actions are morally good but not morally obligatory is ancient, going back at least to the parable of the Good Samaritan in the New Testament. The idea was developed theologically in the Roman Catholic separation of precepts,

[3] The Kantian alternative demarcates perfect duties, which give strict priority to the moral, from imperfect duties, which permit the nonmoral to win out some of the time (1785, Ak 4:421–423).

which are divine commands that always take precedence, and counsels on matters like renunciation of wealth and love of one's enemy, which are recommended but not obligatory. Protestant reformers bitterly opposed this distinction, objecting to the suggestion that divine counsels could ever be optional (Mellema 1991, pp. 49–54, 2004, pp. 109–110). Despite the theological furor, the concept of supererogation entered the philosophical vocabulary rather late. J. O. Urmson urged that, in addition to the traditional trichotomy of obligatory, permitted, and prohibited actions, an additional category is needed for saintly and heroic actions that are morally good yet not obligatory (1958). These exemplary actions have come to be known as supererogatory.

Let us take a moment to situate the supererogatory in conceptual space. Recall that the deontic modalities of permission, prohibition, and obligation are analogous in some ways to the alethic modalities of possibility, impossibility, and necessity. They are, for instance, interdefinable (von Wright 1983, p. 101). Suppose we take the idea that an action *a* is obligatory (O*a*) as primitive. Within the complementary class of non-obligatory actions, we notice two subcategories: non-obligatory and prohibited (–O*a* ∧ O–*a*), and non-obligatory and permitted (–O*a* ∧ –O–*a*). The subcategory of non-obligatory and permitted actions subdivides further. It includes actions that are morally neutral, such as tying your left shoe before the right one in ordinary circumstances, and actions that are morally positive. These morally positive actions include the supererogatory.

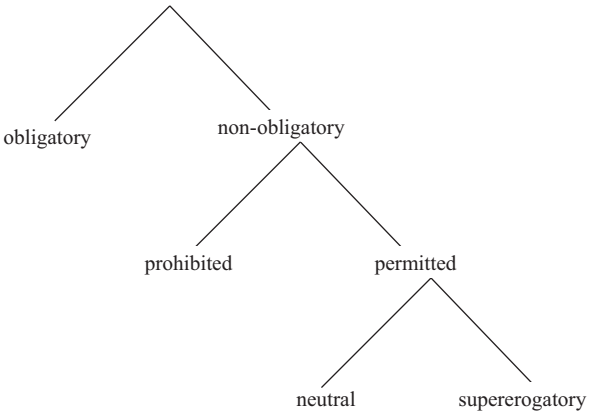
Whether anything beside the supererogatory belongs to this morally positive subcategory does not appear to be established by ordinary usage. On the one hand, we might consider minor actions like a kind response to a prickly question to be morally permitted and positive yet not supererogatory, for they lack the great self-sacrifice characteristic of saintly and heroic actions. Alternatively, we could consider them to be supererogatory, though to a lesser degree. The second option is preferable, I think. A kind response to a prickly question does require a small degree of self-sacrifice, but just as generosity admits of degrees, supererogatory actions need not all exhibit the same degree of self-sacrifice. This inserts useful middle ground between the moral mediocrity of those who merely observe moral obligations and the moral heroism of those who perform supererogatory acts of an extreme sort.[4] In addition, the second option has the advantage of simplicity; it permits an exhaustive division of the morally permitted into morally neutral and supererogatory actions. If we adopt this second option, the conceptual relations can be represented by Fig. 6.1.

The appeal to supererogation in the antecedent of C1 above raises at least two major problems. The first is whether we need the concept of supererogation at all. The issue was aired some years ago in an exchange between Elizabeth Pybus and Patricia McGoldrick. Pybus initiated the discussion by arguing that the saintly and heroic actions Urmson wanted to account for could be adequately conceptualized as moral obligations:

> I cannot at the same time say that something is a moral ideal, and feel that I have no sort of obligation to pursue it. Saying that something is a moral ideal is saying that it is something we have some obligation to pursue. (1982, p. 195)

[4] Here I am indebted to Renzo Llorente.

Fig. 6.1 Deontic modalities



According to Pybus, then, the concept of supererogation is superfluous.

In response, McGoldrick hewed to the Kantian line that we have duties to ourselves as well as to others. To claim that in some circumstances we could have a duty to others to throw ourselves on a hand grenade, say, would violate our duties to ourselves:

> For such a requirement would come into conflict with our obvious duty to recognize our own intrinsic worth, and judge our own aspirations, goals, and interests as no less endowed with value than the aspirations and interests of others. Heroic or saintly virtues can be judged morally praiseworthy and worthy of aspiration, but they cannot be demanded, of ourselves or others, as an imperative of duty without abrogating the intrinsic worth of the individual upon whom the imperative is laid. (1984, p. 527)

Thus the concept of supererogation is needed, according to McGoldrick, to differentiate the extraordinary actions of saints and heroes, which are freely performed despite not being morally required, from basic moral requirements, which are binding for all.

Pybus pursued the exchange by maintaining that duties of self-sacrifice for the well-being of others need not violate duties to self:

> I think it is possible consistently to believe that circumstances might arise in which someone might correctly judge that he ought to sacrifice his life for others, and to believe that it is wrong to allow others to abuse him or use him as means to their ends. If I perform a sacrificial action, it is after all *my* action. Other people cannot, even in killing me for their own ends, involve me in self-sacrifice. To believe that there can be duties of sacrifice is far from believing that others may treat me as they will. To believe that circumstances might arise when justice could require me to die is not to believe that I am a suitable object for use or abuse by others. (1986, p. 529)

Once again, Pybus implies, actions that others deem supererogatory are morally required.

To support the view that supererogation cannot be swallowed up by moral obligation, I would like to call attention to Pybus' just-quoted claim that self-sacrifice is consistent with not being a suitable object for abuse by others. This much is true, I think, but it also misses McGoldrick's point. McGoldrick was concerned about

violating duties to self, about abuse of one's own self rather than abuse by others. That someone who feels *obligated* to sacrifice her life—not someone who wants to *give* her life—can at the same time fulfill her duties to herself is far from clear. This consistency could be shown by the right kind of example, but that leads us to a deeper point still: Are there such cases?

Russell A. Jacobs, who comments on the Pybus-McGoldrick exchange, adduces two types of situations where someone might be morally required to sacrifice her life:

> [I]n some situations at least, it does not violate the moral worth of the individual agent to demand large sacrifices of him: suppose, for instance, that he has voluntarily accepted employment which he knows may involve great risk or sacrifice. Can he later justify failing to do what his job requires because it is dangerous, or requires sacrifice? Or suppose the agent has, deliberately or negligently, created the risk or the need for sacrifice: is he free to allow others to suffer the consequences of his acts because it would be costly for him to do so? If I become a fire-fighter, or a policeman, or a soldier, voluntarily accepting the risks and the rewards, can I subsequently escape the dangers and cost of my duties by claiming that "I am not morally required to make large sacrifices for others"? We may well have duties which are costly, and this despite the fact that we have intrinsic moral worth. (1987, p. 100)

According to Jacobs, then, occupational choices like that of a firefighter, policeman, or soldier and other choices that have created the risk or need for sacrifice may lead to an agent's moral obligation to sacrifice her life.

But Jacobs also argues that not all great self-sacrifice is morally required. A useful test case is Socrates' death. Pybus seems to think that it was both consistent with Socrates' duties to himself and morally required:

> Socrates' death was heroic, governed by considerations of justice, and an autonomous act. Being governed in such a way by moral considerations does not show a lack of autonomy or self-respect. It exhibits self-respect of the best kind, i.e., respect for oneself as a moral agent. (1986, p. 530)

Granted, Socrates' death was morally heroic; he chose not to escape out of fidelity to his vocation as public benefactor. And it does exhibit self-respect as a moral agent. But I do not think Socrates' decision was morally required. Moral blame would be inappropriate if, after devoting the greater part of a lifetime to the intellectual and moral improvement of Athens, Socrates had decided to escape in the interests of his wife, his three children, and himself. By the start of the trial, Socrates had amply fulfilled his commitment as gadfly to Athens. Having given so much, he was not required to give any more. But he did.

If, as I think, Socrates' death was not morally required but freely given, we need the concept of supererogation to describe such actions. I will proceed, therefore, under the assumption that the concept of supererogation has a point.

That leaves us with a second major problem, however. To make appeal to supererogation more than an in-principle solution to problems posed by moral-nonmoral disequilibria, we must have some way of identifying supererogatory actions. If we are going to say that some moral actions are required and some are not, how are we going to differentiate them? How are we going to draw the line between moral duty and that which is beyond the call of duty?

As a kind of trial balloon, I would like to float a theory-relative approach. Since Sects. 4.6 and 5.5 have argued for the superiority of Frankenian theory to Kantian and Benthamite theory, I want to explore the possibility of distinguishing the obligatory from the supererogatory within the ambit of Frankenian theory. As we have noted, Frankena recognizes fundamental principles of justice and beneficence. The principle of justice is egalitarian. The principle of beneficence breaks down into the four subsidiary principles detailed in Sect. 5.6:

(1) Do not cause evil.
(2) Prevent evil.
(3) Remove evil.
(4) Do good.

Might we demarcate what is morally required from what is supererogatory within this framework? Frankena himself may seem to go a long way toward doing so when he remarks "Of the four [subsidiary principles of beneficence], it is most plausible to say that (4) is not a duty in the strict sense" (1973, p. 47). We might then try to specify the content of what is morally required as follows:

(R1) Do not cause evil.
(R2) Prevent evil.
(R3) Remove evil.
(R4) Do not cause injustice.

Correlatively, we might attempt to describe what is supererogatory along these lines:

(S1) Do good.
(S2) Promote justice.

Although this proposal may have a certain appeal, it is evidently not satisfactory. Take (R2), the directive to prevent evil, for example. To prevent a death by drowning might be a duty for a lifeguard but beyond the call of duty for a neophyte swimmer. Similarly, (R3)'s charge to remove evil in the case of a stranger in need of medical assistance could be a duty for a properly-equipped doctor but beyond the call of duty for the Good Samaritan. In addition, there is the problem of overlapping descriptions. The Good Samaritan is both removing evil (R3) and doing good (S1) in treating the victim's wounds. This would seem to make the same action both morally required and not morally required. These problems are so serious that I see no way to make this principle-driven approach work. But if it will not work, is there anything that will?

There are two additional approaches that, I think, have a greater chance of success. One is simply to start with our paradigm cases and work outward. Beginning with actions like those of the Good Samaritan and the soldier who throws himself on a hand grenade, we can identify other actions as supererogatory because they are relevantly similar to these. As a matter of historical fact, this is how the concept of supererogation has grown. It is also the way that the concept can be expected to continue growing. Though our paradigm cases of supererogation involve great self-sacrifice for others, Jason Kawall has argued that the concept of supererogation can be extended to actions motivated by self-regarding reasons (2003).

The other promising approach can get underway once this process of analogical reasoning has produced a body of identifiably supererogatory actions. Moral philosophers can begin to discuss these results in an attempt to determine what general conditions supererogatory actions fulfill. David Heyd, for example, has proposed the following definition (1982, p. 115):

> An act is supererogatory if and only if
> (1) It is neither obligatory nor forbidden.
> (2) Its omission is not wrong, and does not deserve sanction or criticism—either formal or informal.
> (3) It is morally good, both by virtue of its (intended) consequences and by virtue of its intrinsic value (being beyond duty).
> (4) It is done voluntarily for the sake of someone else's good, and is thus meritorious.

Careful conceptual analyses like this one are invaluable, but I suggest that they be interpreted from a post-Wittgensteinian perspective. What we will find, I think, is that comparing one supererogatory action to others results in "a complicated network of similarities overlapping and criss-crossing: sometimes overall similarities, sometimes similarities of detail" (Wittgenstein 1953, p. 66). We can expect the network of similarities to turn out to be so complicated that whether or not an action is supererogatory will not always be clear. In other words, the concept of supererogation is vague. But its vagueness will not prevent us from unproblematically identifying many actions as supererogatory.

6.5.2 Moral Obligation

To adjudicate the relative priority of the moral and nonmoral, Sect. 6.5 stated the conditionals C1 and C2. C1, which was addressed in Sect. 6.5.1, relies on the concept of supererogation. But C2 appeals to moral obligation:

> If the action is morally obligatory, the moral should take precedence over the nonmoral.

Unfortunately, the concept of moral obligation suffers from both ambiguity and vagueness. The point about ambiguity can be made in more than one way. The term 'moral' alone has at least three distinguishable senses. The most fundamental is as primitive in the evaluative triple of moral, immoral, and amoral. 'Moral' in this sense affords first-order evaluations of actions and people. By contrast, there are at least two second-order uses. Narrow second-order usage picks out the discursive field characterized by predicates such as 'moral', 'immoral', and 'amoral' in their evaluative senses together with related predicates like 'just', 'honest', and 'generous'. In this narrow sense, moral discourse is set off from esthetic, legal, culinary, ludic, economic, and other sorts of discourse. By contrast, wide second-order usage admits no contrast between moral and other discursive fields. In this wide sense, any action (including mental acts of decision) that would count as an answer to 'What should I do?' or 'How should I live my life?' is moral. This is the sense

traditionally employed in the moral sciences as conceived by Hume, Adam Smith, and Dilthey. I will refer to these three senses of 'moral' as evaluative, narrow, and wide respectively.[5]

The wide sense appears to be particularly inviting to contemporary theorists. Marcia Eaton maintains that it is "at least sometimes a mistake" to separate the moral and the esthetic and that conflicts between them require "broad moral decisions" (1992, p. 226, 234). In this wide sense, morality becomes pervasive. Robert Louden develops the thesis of the pervasiveness of morality as follows:

> If we can convince ourselves (as many people before us have done) that our own character is what ethics is primarily about, then it becomes much more difficult to evade the question of moral assessment regardless of what one is doing. And if the development of one's moral character also entails the development of one's nonmoral character (subject to one's talents and resources, as well as to one's own choice of life plan), we have a second argument for pervasiveness. Third, if we accept the classical view (revived in our own era by virtue theorists) that ethics is primarily about how one lives one's life rather than what discrete acts one performs in moral quandary situations, we have yet another argument for the pervasiveness of the ethical. (1988, p. 374)

Attempts to flesh out this wider conception of morality might take an Aristotelian or a Wittgensteinian turn. Louden illustrates the Aristotelian approach by calling attention to the architectonic role of ethics and politics (*Nicomachean Ethics* 1094a26–b7): "*all* serious questions about what to do and how to live are (by definition, on this [Aristotelian] view) ethical and/or political questions" (1988, p. 375). But Alice Crary takes the Wittgensteinian route, arguing that acquiring a language is inseparable from acquiring a "practical orientation to the world [that] cannot help but encode a view of what matters most in life or how best to live"; that is, "learning to speak is inseparable from the development of an—individual—moral outlook" (2007, p. 43). Either way, we end up with a view of morality as pervasive.

In addition to these developments of the wide sense of morality, some have attacked the narrow sense. Cora Diamond refers to it as the "departmental conception of morality" and urges us "to reject the idea that moral thought is a *department* of thought, and moral discourse a *department* of discourse" (1996, p. 104). But I see no point in trying to outlaw either the narrow (departmental) or wide (pervasive) sense of morality. The main thing is to be clear about which one is in play. Though I will employ all three senses below, I will rely primarily on the narrow sense, taking it for granted that we have a moderately clear distinction between narrowly moral and, say, esthetic considerations. There are borderline cases, of course; admiration can be a mix of moral and esthetic considerations, for instance (Eaton 1992, p. 230). But we have no difficulty in classifying 'Gauguin's treatment of his family was cruel' as a moral judgment and 'Gauguin's painting has affinities to cloisonné' as an esthetic one.

[5] Some writers have proposed the adoption of different terms to keep these different senses straight. Marcia Eaton, for example, suggests 'moral' for assessing particular actions (the evaluative sense) and 'ethical' for "meaning-of-life" issues (the wide sense) (1992, pp. 237–238).

We have been reviewing the ambiguities of 'moral', but the compound term 'moral obligation' is likewise ambiguous. To say that an action is morally obligatory relies on evaluative first-order usage, while to refer to the language of moral obligation employs narrow second-order usage. Might 'moral obligation' also have a wide second-order sense? Describing an obligation as moral normally identifies it as an obligation of a special—moral—sort, and this invokes the moral-nonmoral distinction characteristic of narrow usage. Yet it does not violate sense to widen the term's meaning and say that moral obligation rests on all relevant considerations, narrowly moral and nonmoral alike.

The intimately related term 'ought' is highly ambiguous. We can distinguish the narrowly moral 'ought' and the nonmoral 'ought'. The instrumental 'ought', which recommends an action as a means, differs from the teleological 'ought', which advocates an end. The subjective 'ought', which can be mistaken, is not the same as the objective 'ought', which cannot (Ewing 1947, pp. 118–123, 1953, pp. 144–145; Parfit 1984, p. 25). 'Ought' can be past-regarding or future-oriented (Weirich 2004, pp. 128–129). Prima facie 'ought' diverges from actual (or absolute) 'ought' (Ross 1930, pp. 19–34). The 'ought' that occurs in statements of moral prescription is not the deliberative 'ought' of 'What ought I to do?' (Williams 1973, pp. 184–185). Likewise for prima facie 'ought' and overall 'ought' (Zimmerman 1996, p. 207).

In addition to the ambiguities just noted, the vagueness of moral obligation is painfully familiar to all. A much-discussed instance is Gauguin's conflicting obligations to his family and his art. Was Gauguin morally obligated to devote himself to his family? His art? Both? Neither? These are intricate questions, of course, and I will not pretend to offer definitive answers. But I do want to explore some of the relevant issues.

Some writers have suggested perspectives from which Gauguin's choice of his art can be seen as evaluatively moral. Brian Rosebury, for example, assumes that Gauguin's motives are altruistic: "the dilemma in its most crudely stated form is that if he stays with his family he contributes to their happiness, whereas if he leaves them and goes to the South Seas he contributes to the happiness of numerous art lovers, but he cannot do both" (1995, p. 517).

Arguing from a different perspective, Joseph Kupfer responds to Michael Slote's claim that Gauguin's single-minded passion was admirable but immoral by contending that his passion was not admirable but moral (Slote 1983; Kupfer 1992). Its morality, he thinks, can be shown on utilitarian grounds: "the cost was not so great relative to the value produced," and "it yielded great social good as only it could, in spite of its inseparability from harm—an unavoidable 'casualty of greatness'" (1992, p. 63, 66).

The two foregoing approaches do not, I think, show Gauguin's decision to be evaluatively moral. However helpful Rosebury's assumption of an altruistic Gauguin may have been in its original context, it does not seem to hold for the real Gauguin.[6] Gauguin's attitude toward his five children with Mette Gauguin and his

[6] Rosebury may have been emboldened by Bernard Williams' invention of a Gauguin whom we can describe "[w]ithout feeling that we are limited by any historical facts" (1981, p. 22).

four known children with Polynesian women appears to have been near-total neglect; his financial support was woefully meager and grudging, and he did not even bother to observe the birthdays or Christmases of the children of his marriage (Matthews 2001, p. 100, 164). In addition, his extra-familial relations were apparently cut from the same cloth. A journalist interviewing Tahitians who had known him remarked "Not one person with whom I spoke who had known him (with the solitary exception of his old mistress), had a good word to say concerning the painter" (quoted in Matthews 2001, p. 234). In short, Gauguin does not seem to have been an altruist in any sense. At best, he was driven by esthetic concerns; at worst, as his wife observed, by "ferocious egoism" (Matthews 2001, p. 221, 268).

Moreover, any utilitarian conclusion that the good produced by Gauguin's art outweighs the harm produced by abandoning his family is doubtful. Anyone who has read Mette Gauguin's surviving letters (Matthews 2001, Appendix) cannot help but be dismayed by blithe dismissals of the "(relatively) minimal cost" of Gauguin's decision (Kupfer 1992, p. 67). If we think in terms of the hedonistic calculus, Gauguin's decision to pursue his art is probably supported by the extent of the resulting esthetic pleasures but undercut by the duration and intensity of the consequent familial pains. Notoriously, though, the calculus provides no way of actually carrying out the calculation.

If neither of these defenses of the morality of Gauguin's decision holds, then was his decision evaluatively immoral? It would be presumptuous, I think, for distant observers like ourselves to attempt a definitive answer to this question, for definitive answers require full knowledge of the facts. But I will venture three further comments.

First of all, I will not deny the *logical* possibility of circumstances that would force a choice between the interests of Gauguin and those of his family. But I will suggest, second, that the default objective for dealing with binary dilemmas—neither one nor the other but both—does not seem to have been pursued very strenuously in Gauguin's case. Finally, there are general strategies that might be employed in dealing with career-and-family dilemmas and, more inclusively, with dilemmas characterized by moral-nonmoral conflict.

One such strategy is due to Adam Morton, who discusses several ways of coping with incomparable desires, that is, desires such that the agent neither prefers one to the other nor is indifferent between them. Since desires include ends as a special case (Sect. 5.3), Morton's discussion is applicable to incomparable career and family ends.[7] He proposes revaluing as a subtler alternative to the obvious strategies of abandoning one desire or compromising between desires. In the special case of conflicting career and family ends, revaluing would consist in finding "a pair of more precise values than just 'career success' and 'family life' such that you can hope to maximize both simultaneously" (1991, p. 48). Specifically, revaluing in such cases might work as follows:

> Suppose, for example, that advancement in your career, in the sense of occupying better-paid and more respected positions, does not matter crucially to you. What you do want from

[7] Morton himself refers to "career and family ambitions" (1991, p. 48).

> your work is that it provide a succession of interesting and rather different jobs, so that you can look forward to a change every few years, never get stuck in a rut. You may be able to arrange this without putting in the long hours and superior-stroking that conventional advancement requires. You might become your company's expert on rewriting defective software, or on handling morale crises in sub-departments, so that you would be transferred from one trouble-spot to another when your special skills were required. Or, on the other side of the balance, suppose that what matters to you about family life is not day-to-day contact but a sense of long-term involvement with the lives of your spouse and children. So you may be able to spend long hours at work, often not coming home in the evening, but devoting one day a week to a family excursion, chosen so as to allow real conversations and exchanges of feeling, and every year having a brief but intense and memorable family holiday. (1991, pp. 47–48)

The foregoing remarks have stressed the ambiguity and vagueness of our concept of moral obligation. I hope they will contribute to a sober estimate of the difficulties of applying the concept. Despite these difficulties, I also hope that we can continue to affirm C2:

> If the action is morally obligatory, the moral should take precedence over the nonmoral.

C2 is an endorsement of the overridingness of the morally required. Although Flanagan concedes the appeal of this version of the overridingness thesis, he objects even to it: "But not because the notion it expresses of more or less worth-while goods is unimportant, rather because qua philosophical thesis it lacks content and does little action-guiding or dispute-resolving work" (1986, p. 53).

I would like to conclude this section with a modest suggestion about how to improve the action-guiding and dispute-resolving power of the overridingness thesis. The suggestion is to begin by determining whether the action in question is not supererogatory. If it is not, it must then be morally neutral, prohibited, or obligatory (cf. Sect. 6.5.1, Fig. 6.1). Which of these three categorizations is appropriate can be determined in principle on instrumental grounds. If the action makes no difference to our moral goals, it is morally neutral; if its performance is necessary to these goals, it is morally obligatory; and if its nonperformance is necessary to the same goals, it is morally prohibited. No one will mistake this procedure for an algorithm, evidently. We can expect it to be plagued by borderline cases. Section 2.1 has already endorsed Peirce's position on vagueness: "*No* concept, not even those of mathematics, is absolutely precise; and some of the most important for everyday use are extremely vague" ([c. 1906, 6.496; emphasis in original). But to the extent that we have moral goals, the overridingness thesis can be action-guiding; to the extent that these goals are shared, the thesis can be dispute-resolving.

6.5.3 Overridingness

Beginning in the late 70s and early 80s, the overridingness thesis of Sect. 6.5 has been repeatedly challenged (e.g., Foot 1978; Williams 1981; Wolf 1982; Slote 1983). Though its critics have not always observed Flanagan's distinction between the overridingness of the morally ideal and the overridingness of the morally

required, I will focus their objections on what I take to be the only defensible version of the thesis: the overridingness of the morally required. One way of stating this form of the thesis is that what is morally obligatory is always obligatory overall. The critics of overridingness are therefore united by opposition to a universally quantified thesis, and their arguments typically resort to counterexamples. The counterexamples purport to show that the morally obligatory action is not always obligatory overall.

Philippa Foot, one of the earliest of these critics, instances several such examples. One involves a decision about whether to shake someone's proffered hand when the rules of etiquette prohibit handshaking. She supposes that shaking the hand would be motivated by the moral consideration of not hurting the other's feelings but that doing so would necessitate "the spending of a rather large sum of money; perhaps thousands of dollars or pounds." She also supposes that "the man who would spend the money doesn't need it, and is going to spend it on something for himself, not on some morally good cause" (1978, p. 183). She thereby structures the dilemma as a conflict between the moral consideration of not hurting the would-be handshaker and the financial consideration of spending a large sum of money. Her solution: "In face of a sizeable financial consideration a small moral consideration often slips quietly out of sight" (1978, p. 184). Hence what is morally obligatory is not always obligatory overall.

Given Foot's description of the situation, it does seem plausible that the financial consideration should come first. But the reasons, I believe, are not the ones Foot offers. Not hurting the overeager handshaker is indeed a moral consideration, but so is not hurting the reluctant spender, who is also a member of the moral community. Choosing either horn of the dilemma would cause pain to someone, but since that of the spender would be "sizeable" and that of the handshaker would be "small," the balance naturally falls in favor of the spender. Hence the situation is not properly characterized as a conflict between moral and financial considerations; it is a conflict between one moral consideration and another. So here it is not the case that the morally obligatory action and the overall obligatory action are distinct.

Another of Foot's examples cites "a rule of etiquette which operates on most people so strongly that it takes precedence even over a rather weighty moral consideration. There is, for instance, distinct resistance to the idea that a host or hostess might refuse to serve any more drinks when the guests have had as much as is good for them given that they must drive home" (1978, p. 184).

Unfortunately, this example suffers from three crippling deficiencies. The first is that even if it is a fact that "most people" place the etiquette of hospitality before the morality of drunken driving, this hardly shows that they *should* do so. Secondly, the example fails to consider the possibility of creative alternatives; in situations like these, the host with the most spares no effort in seeking alternative transportation for incapacitated guests. Finally, if push comes to shove and a better alternative cannot be found, is there any doubt about which of the two worst-case scenarios is worse, all things considered? A dissatisfied guest or carnage on the highway? Instead of refuting the overridingness thesis, this example actually turns out to support it.

Susan Wolf also objects to the overridingness thesis, claiming that "its being one's duty [that is, morally obligatory] to do something does not necessarily imply that, all things considered, one should do it" (1986, p. 145). One of her examples goes as follows:

> [I]t is reasonable to think that you have a duty to honor your commitments and that deciding to hold office hours on Tuesday afternoons involves a commitment to being in your office every week at that time. But imagine that one Tuesday on your way to the office, you are passing an apartment house on fire. There is a frail, elderly man inside, and the fire department has not yet arrived. It is uncertain whether the firemen will get there in time to save the man's life. Now, it is generally believed that one does not have a duty to rush into a burning building to save the life of a stranger. But surely it would be absurd to be deterred from this act by the thought that one *does* have a duty to keep office hours. So it would seem that sometimes it is better to violate one's duty than to forego the opportunity to do a good that is not one's duty. (1986, p. 139)

To evaluate this example, we note the dual possibility that attempting to save the elderly man may or may not be the best thing to do, all things considered. It may not be the best thing to do if you suffer from a lung condition that would probably be fatal in a burning building or if the chance that you and the man could both be trapped is high. In such cases, the duty to keep office hours would not be trumped by some higher good. But if, on the other hand, the attempt to save the elderly man is the best thing to do, it is also morally required by the duty to prevent evil (cf. Sect. 5.6). What we would have, then, is a conflict between two prima facie moral obligations: the duty to keep one's commitments and the duty to prevent evil. If the best thing to do really is to attempt to save the elderly man, the moral obligation to prevent evil wins out. Hence what is morally obligatory is also obligatory overall.

Another of Wolf's examples contrasts the duty to hold office hours with an unexpected opportunity to attend a lecture by your "philosophical heroine." Unfortunately, both activities are scheduled for the same time. "Should you go [to the lecture], or instead drive to your own office just in case some of your students want to complain about their grades?" (1986, p. 142).

Wolf plainly assumes that her academic readers will think it best to attend the lecture, and it would indeed be best in many circumstances that fit her template for the case. But why is this so? To sketch an answer to this question, I propose that we reflect briefly on the nature of professions, for Wolf's description strongly suggests that the philosophical heroine's lecture has considerable professional importance. According to Stephen Barker, the original medieval professions of medicine, law, theology, and higher education shared two central features:

> 1. The medieval professions had in common that each required mastery of an extensive body of book-learning, and this was to be achieved by years of university study.
> 2. Entrants into them were required to commit themselves to a distinctive ideal of service which imposed ethical demands to which ordinary citizens were not subject. (1992, p. 87)

Barker then proposes that "occupations today should be counted as professions to the extent that they resemble the medieval professions in these two basic respects" (1992, p. 87).

Contemporary higher education and, more specifically, contemporary higher education in philosophy clearly share these two features. Concomitantly, a vast majority of contemporary philosophers would surely concur in classifying their occupation as a profession. If philosophy is a profession, then philosophers are committed to "a distinctive ideal of service" that imposes special "ethical demands." What is this distinctive ideal of service? For philosophers who practice their profession within the context of higher education, the ideal of service includes service to students, and this imposes ethical demands to teach classes, grade papers, hold office hours, etc. But the ideal of service is not exhausted by reference to students; philosophical service includes service to the profession, that is, to the community of philosophers. Attending the lecture might well be a form of service to other philosophers, who would gain from an engaged interlocutor; to the philosopher's students, who would benefit from a more informed instructor; and to the philosopher herself, who is no less a member of the philosophical community and would profit from exposure to her philosophical heroine. In such situations, then, attending the lecture would probably be the best thing to do, all things considered. It would also be the morally best thing to do.

To exhaust the reader's patience with further criticism of counterexamples would be both easy and pointless, for even a great many of these exercises would not show conclusively that what is morally obligatory is always obligatory overall. So I will adduce no further counterexamples. But I will suggest that the failure of putative counterexamples like those just cited is already weighty evidence in favor of the overridingness of the morally required. Additional confirmatory evidence can be found in the inexhaustible stock of positive instances whereby the morally obligatory is also obligatory overall. Suppose, for example, that I am fond of the thrills of arson. If it is morally obligatory to respect others' property despite this fondness, then it is also obligatory overall to do so. Because of the number and variety of positive instances of this sort, the overridingness of the morally required appears to be highly confirmed. If so, it should continue to be regarded as such until such time as a convincing counterexample appears.

Then are the critics of overridingness all wrong? I do not think so. But there is a relatively straightforward explanation for their position's appeal. This explanation has two facets. The first recurs to the distinction between the overridingness of the morally ideal and the overridingness of the morally required. If this distinction is ignored, the overridingness thesis does turn out to be false, for nonmoral considerations can rightly override the morally ideal. As Wolf notes, we are required "to do something for the world but not everything" (1986, p. 134). The other facet of the explanation exploits the distinction between prima facie and overall moral obligation. If the morally obligatory is merely prima facie obligatory, then the overridingness thesis would again appear to be false, for prima facie moral obligations can be legitimately overridden. I suspect, however, that they can only be overridden by other moral obligations. If this is so, the overridingness thesis would appear to be false whenever the moral dimensions of the overriding obligations are ignored. This is in fact what happened with Foot's handshaking example, which overlooked the moral implications of causing financial pain, and Wolf's cases of the elderly man

and the lecture-inclined philosopher, which ignored the moral obligation to prevent evil and the moral demands of professional membership respectively.

6.6 Conclusion

If the concepts of the morally obligatory and the supererogatory are both vague, then there is a very fuzzy line between moral duty and that which is above and beyond it. The vagueness of the morally obligatory is especially exasperating when there are conflicting prima facie moral obligations. Along these lines lie many of our most recalcitrant moral dilemmas. This is the hard part, of course: here we labor to decide. But there is a brighter side as well. These are also moments of moral opportunity. They provide prospects for the kind of self-definition and value consolidation that Socrates achieved in his final days. If, on a given Sunday afternoon, the relative priorities of family and career are not clear, then there is no alternative to *creating* priority. Moments like these are malleable; they await the stamp of distinctively personal action.

John Keats once sketched a nexus of ideas that serves well, I think, as a coda to this work. He thought of human intelligences as generic "atoms of perception" that are not initially souls. Souls can be made, however, and they are made by acquiring identities. To acquire an identity, an intelligence requires a "vale of Soul-making," a series of personal trials: "Do you not see how necessary a World of Pains and troubles is to school an Intelligence and make it a soul? A Place where the heart must feel and suffer in a thousand diverse ways!" (Keats 1819, pp. 290–291). Applied to moral experience, these ideas suggest the value of moral struggle. In these moments of uncertainty, the decisions that we make turn out to make us as well. In the pain and trouble of moral decision making, we make the most of our limited but inescapable freedom; here, amid a range of reasonable possibilities, we take our bearings and set a course.

References

Anderson, John D., et al. 1998. Indication, from Pioneer 10/11, Galileo, and Ulysses data, of an apparent anomalous, weak, long-range acceleration. *Physical Review Letters* 81:2858–2861.

Anderson, John D., et al. 1999. Anderson et al. reply. *Physical Review Letters* 83:1891.

Anderson, John D., et al. 2002. Study of the anomalous acceleration of Pioneer 10 and 11. *Physical Review D* 65:082004.

Aristotle. *Nicomachean Ethics*. In *The complete works of Aristotle,* ed. Jonathan Barnes, Vol. II, 1729–1867. Princeton: Princeton University Press, 1984.

Barker, Stephen F. 1992. What is a profession? *Professional Ethics* 1:77–99.

Crary, Alice. 2007. *Beyond moral judgment*. Cambridge: Harvard University Press.

Diamond, Cora. 1996. "We are perpetually moralists": Iris Murdoch, fact, and value. In *Iris Murdoch and the search for human goodness,* eds. Maria Antonaccio and William Schweiker, 79–109. Chicago: University of Chicago Press.

Eaton, Marcia. 1992. Integrating the aesthetic and the moral. *Philosophical Studies* 67:219–240.
Elgin, Catherine Z. 1996. *Considered judgment*. Princeton: Princeton University Press.
Ewing, A. C. 1947. *The definition of good*. New York: Macmillan.
Ewing, A. C. 1953. *Ethics*. London: English Universities Press.
Flanagan, Owen. 1986. Admirable immorality and admirable imperfection. *The Journal of Philosophy* 83:41–60.
Foot, Philippa. 1978. Are moral considerations overriding? In *Virtues and vices and other essays in moral philosophy,* 181–189. Oxford: Basil Blackwell. (Berkeley: University of California Press).
Frankena, William K. 1973. *Ethics*. 2nd ed. Englewood Cliffs: Prentice-Hall.
Gauthier, David. 1986. *Morals by agreement*. Oxford: Clarendon Press.
Goodman, Nelson. 1979. *Fact, fiction, and forecast*. 3rd ed. Indianapolis: Hackett.
Heyd, David. 1982. *Supererogation: Its status in ethical theory*. Cambridge: Cambridge University Press.
Hugo, Victor. 1862. *Les misérables*. English edition: Hugo, Victor. 1982. *Les misérables* (trans: Denny, Norman). London: Penguin.
Jacobs, Russell A. 1987. Obligation, supererogation and self-sacrifice. *Philosophy* 62:96–101.
Kamm, F. M. 2007. *Intricate ethics: Rights, responsibilities, and permissible harm*. Oxford: Oxford University Press.
Kant, Immanuel. 1785. *Grundlegung zur metaphysik der sitten*. English edition: Kant, Immanuel. 1996. *Groundwork of the metaphysics of morals* (trans. and ed: Gregor, M. J.). In *Practical philosophy,* 41–108. Cambridge: Cambridge University Press.
Kawall, Jason. 2003. Self-regarding supererogatory actions. *Journal of Social Policy* 34:487–498.
Keats, John. 1819. To George and Georgiana Keats. In *Selected letters of John Keats: Based on the texts of Hyder Edward Rollins,* rev. ed., ed. Grant F. Scott, 254–301. Cambridge: Harvard University Press, 2002.
Kupfer, Joseph. 1992. Gauguin, again. *Pacific Philosophical Quarterly* 73:63–72.
Levi, Isaac. 1984. *Decisions and revisions*. Cambridge: Cambridge University Press.
Louden, Robert B. 1988. Can we be too moral? *Ethics* 98:361–378.
Maher, Patrick. 1993. *Betting on theories*. Cambridge: Cambridge University Press.
Matthews, Nancy Mowll. 2001. *Paul Gauguin: An erotic life*. New Haven: Yale University Press.
McGoldrick, Patricia M. 1984. Saints and heroes: A plea for the supererogatory. *Philosophy* 59:523–528.
Mellema, Gregory. 1991. *Beyond the call of duty: Supererogation, obligation, and offence*. Albany: State University of New York Press.
Mellema, Gregory. 2004. *The expectations of morality*. Amsterdam: Rodopi.
Milgram, Stanley. 1974. *Obedience to authority*. New York: Harper and Row.
Morton, Adam. 1991. *Disasters and dilemmas*. Oxford: Basil Blackwell.
Nussbaum, Martha C. 1990. *Love's knowledge: Essays on philosophy and literature*. New York: Oxford University Press.
Nussbaum, Martha C. 1997. Review of Gregory Vlastos, *Socratic studies*. *The Journal of Philosophy* 94:27–45.
Parfit, Derek. 1984. *Reasons and persons*. Oxford: Clarendon Press.
Peirce, Charles S. [c. 1906] 1931–1958. Answers to questions concerning my belief in God. In *Collected papers of Charles Sanders Peirce,* eds. C. Hartshorne, P. Weiss, and A. Burks, 340–355. Cambridge: Harvard University Press.
Putnam, Hilary. 1983. *Realism and reason*. *Philosophical papers,* Vol. 3. Cambridge: Cambridge University Press.
Putnam, Hilary. 1992. Beyond the fact/value dichotomy. In *Realism with a human face,* 135–141. Cambridge: Harvard University Press.
Pybus, Elizabeth M. 1982. Saints and heroes. *Philosophy* 57:193–199.
Pybus, Elizabeth M. 1986. A plea for the supererogatory: A reply. *Philosophy* 61:526–531.
Quine, W. V. 1960. *Word and object*. Cambridge: The MIT Press.

Rawls, John. 1971. *A theory of justice*. Cambridge: The Belknap Press of Harvard University Press.
Richardson, Henry S. 1986. *Rational deliberation of ends*. Dissertation, Cambridge: Harvard University.
Richardson, Henry S. 1994. *Practical reasoning about final ends*. Cambridge: Cambridge University Press.
Rosebury, Brian. 1995. Moral responsibility and "moral luck." *The Philosophical Review* 104:499–524.
Ross, W. D. 1930. *The right and the good*. Oxford: Clarendon Press.
Slote, Michael. 1983. Admirable immorality. In *Goods and virtues,* 77–107. Oxford: Clarendon Press. (New York: Oxford University Press).
Turyshev, Slava G., and Viktor T. Toth. 2010. The Pioneer anomaly. *Living Reviews in Relativity* 13:4. http://relativity.livingreviews.org/Articles/lrr-2010-4/. Accessed 30 April 2014.
United States v. Holmes. 1842. United States Circuit Court, Eastern District of Pennsylvania. 26 F.Cas. 360.
Urmson, J. O. 1958. Saints and heroes. In *Essays in moral philosophy,* ed. A. I. Melden, 198–216. Seattle: University of Washington Press.
Von Wright, Georg Henrik. 1983. *Practical reason: Philosophical papers*. Vol. 1. Oxford: Basil Blackwell.
Voorhoeve, Alex. 2009. *Conversations on ethics*. Oxford: Oxford University Press.
Weirich, Paul. 2004. *Realistic decision theory: Rules for nonideal agents in nonideal circumstances*. Oxford: Oxford University Press.
Williams, Bernard. 1973. Ethical consistency. In *Problems of the self,* 166–186. Cambridge: Cambridge University Press.
Williams, Bernard. 1981. Moral luck. In *Moral luck,* 20–39. Cambridge: Cambridge University Press.
Wittgenstein, Ludwig. 1953. *Philosophical investigations*. Oxford: Basil Blackwell.
Wolf, Susan. 1982. Moral saints. *The Journal of Philosophy* 79:419–439.
Wolf, Susan. 1986. Above and below the line of duty. *Philosophical Topics* 14:131–148.
Zimmerman, Michael J. 1996. *The concept of moral obligation*. Cambridge: Cambridge University Press.

Index

J. R. Welch, *Moral Strata,* Theory and Decision Library A:, 49,
DOI 10.1007/978-3-319-08013-0,