VALUE INCOMMENSURABILITY

ETHICS, RISK, AND DECISION-MAKING

Edited by
Henrik Andersson and Anders Herlitz

Value Incommensurability

Incommensurability is the impossibility to determine how two options relate to each other in terms of conventional comparative relations. This book features new research on incommensurability from philosophers who have shaped the field into what it is today, including John Broome, Ruth Chang, and Wlodek Rabinowicz.

The book covers four aspects relating to incommensurability. In the first part, the contributors synthesize research on the competing views of how to best explain incommensurability. Part II illustrates how incommensurability can help us deal with seemingly insurmountable problems in ethical theory and population ethics. The contributors address the Repugnant Conclusion, the Mere Addition Paradox, and so-called Spectrum Arguments. The chapters in Part III outline and summarize problems caused by incommensurability for decision theory. Finally, Part IV tackles topics related to risk, uncertainty, and incommensurability.

Value Incommensurability: Ethics, Risk, and Decision-Making will be of interest to researchers and advanced students working in ethical theory, decision theory, action theory, and philosophy of economics.

Henrik Andersson is a postdoc at Lund University. His research has had a focus on value theory and especially the phenomenon of value incommensurability. In his current research project, he applies recent results from value theory in order to address the hard choices we face when we aim to combat climate change.

Anders Herlitz is a researcher at the Institute for Futures Studies in Stockholm. His research focuses on comparability problems and rational choice, especially in relation to distributive theory. He is currently working on a monograph addressing how to distribute scarce health resources.

Routledge Studies in Ethics and Moral Theory

From Value to Rightness
Consequentialism, Action-Guidance, and the Perspective-Dependence of Moral Duties
Vuko Andrić

Practical Wisdom
Philosophical and Psychological Perspectives
Edited by Mario De Caro and Maria Silvia Vaccarezza

Virtues, Democracy, and Online Media
Ethical and Epistemic Issues
Edited by Nancy E. Snow and Maria Silvia Vaccarezza

Evil Matters
A Philosophical Inquiry
Zachary J. Goldberg

Desert Collapses
Why No One Deserves Anything
Stephen Kershnar

Neglected Virtues
Edited by Glen Pettigrove and Christine Swanton

Incomparable Values
Analysis, Axiomatics and Applications
John Nolt

Value Incommensurability
Ethics, Risk, and Decision-Making
Edited by Henrik Andersson and Anders Herlitz

Value Incommensurability
Ethics, Risk, and Decision-Making

Edited by Henrik Andersson and Anders Herlitz

Contents

Acknowledgments

Contributors

Henrik Andersson
Postdoctoral researcher
Lund University

Chrisoula Andreou
Professor of Philosophy
University of Utah

Gustaf Arrhenius
Director of the Institute for
    Futures Studies
Stockholm and Professor of
    Practical Philosophy
Stockholm University

John Broome
Emeritus White's Professor of
    Moral Philosophy
Honorary Professor
Australian National University

Krister Bykvist
Professor of Practical Philosophy
Stockholm University and Institute
    for Futures Studies

Ruth Chang
Professor of Jurisprudence
Oxford University and Professorial
    Fellow at University College
Oxford

Ryan Doody
Assistant Professor of Philosophy
Brown University

Luke Elson
Lecturer in Philosophy
University of Reading

Anders Herlitz
Researcher
Institute for Futures Studies
Stockholm

Mozaffar Qizilbash
Professor of Economics
University of York

Wlodek Rabinowicz
Professor of Practical Philosophy
    (emeritus)
Lund University

Katie Steele
Associate Professor of Philosophy
Australian National University

Introduction

Henrik Andersson and Anders Herlitz

Comparisons play an essential role in our lives.[1] When considering what patients should get treated first in the emergency room, we compare health conditions. We compare how promising different TV shows are when we decide how to entertain ourselves. We compare the menus of lunch restaurants to decide where to eat. We compare destinations before planning weekend trips. We compare career prospects before choosing what to study – and so on. Many comparisons like these are made without hesitation. To many, but, of course, not all, it is obvious that patients with gunshot wounds should be treated before those with sprained ankles, that *The Sopranos* is better than *Days of Our Lives*, that the local Italian restaurant is better than McDonald's, that a weekend in Paris is better than a weekend in Slough, and that the career prospects of a lawyer are better than those of a race car driver. However, some comparisons are not easy. Feelings of being at a loss and struggling when trying to determine what option is best are familiar to most of us. This book is about those hard comparisons when no option is at least as good as all the alternatives. It is about the situations in which it seems impossible to rank options in conventional ways and it seems as though conventional comparisons themselves are impossible—the book is about what can be broadly called "incommensurability."

Examples of incommensurability abound. When Sartre's student, during World War II, faced the choice of joining the French resistance in England to fight the occupying Germans or staying in France and take care of his elderly mother, the alternatives seemed incommensurable (Sartre 1975). When asked if Mozart or Michelangelo is the better artist, many think that they are incommensurable (Chang 2002), and when contemplating whether deafness reduces one's health more or less than muteness, the health conditions seem incommensurable (Hausman 2015). Even though the stakes and values involved may differ significantly, these comparisons have one thing in common: they are comparisons in which we cannot determinately judge which alternative is best nor do the alternatives seem equally as good. Furthermore, the phenomenon does not

seem to be merely an epistemic problem. The problem is not lack of knowledge. Studying the alternatives and asking friends or experts will not help determine whether patriotic duties are more important than filial duties, whether music is a higher art than sculpting, or whether hearing is more important than speaking. This book addresses how to make sense of examples such as these, incommensurability in general, its role in ethical theory, and its implications for decision-making.

There is ongoing debate about how to best account for examples such as the aforementioned. Some believe some items simply cannot be compared at all: they are incomparable (Raz 1986). Some believe they are examples of comparisons in which the items relate to each other in non-conventional ways (Chang 2002; Rabinowicz 2008). Some believe they are comparisons in which it is indeterminate which of the conventional rankings hold between the two items (Broome 1997). In this introduction, we do not side with any camp in this debate and therefore use the familiar term "incommensurability" to cover all these possibilities.

The theoretical interest in incommensurability is due to the fact that it indicates that not everything needs to be related by a "better than," "worse than," or "equally as good" relation. This is puzzling since often-used comparative predicates allow us to rank all things they apply to using the analogues of these relations. All things that have some length, weight, or age can, for instance, be pairwise compared in terms of longer/heavier/older than or shorter/lighter/younger than, and the comparison, at least theoretically, allows for the possibility of the things being equally as long/heavy/old. These three types of comparatives exhaust all possible relations between things that have length, weight, or age. The possibility of incommensurability indicates that some predicates are different: when two things are incommensurable, neither is determinately better than the other nor are they equally as good as each other. Hence, they are related in some other manner.

The possibility that the three standard value relations might not fully exhaust all possibilities actualizes some very general problems for axiology and normative theory. In axiology, the possibility of incommensurability would entail that conventional, one-dimensional value representations (e.g., real numbers) cannot accurately represent values and that conventional one-dimensional operators in betterness orderings must be complemented with additional ones. In normative theory, it is often assumed that the correct (or most plausible) normative view can always fully determine a course of action that is at least as good as (or at least as choice-worthy as) all alternatives. If courses of action can be incommensurable, that assumption seems misguided. The possibility of incommensurability might be most obviously worrying for consequentialists who believe that one ought to always act such that the act maximizes total goodness, but it is equally troublesome for non-consequentialists who espouse the idea

that one ought to do that which one has most reason to do. If two different acts, X and Y, lead to outcomes that are incommensurable in value, conventional consequentialism fails to provide guidance. If the reasons to do X are neither stronger than, weaker than, nor equally as strong as the reasons to do Y, non-consequentialist views that one should do what one has most reason to do fail to provide guidance. An early example of such potential incommensurability can be found in *The Methods of Ethics*, in which the author Henry Sidgwick (Sidgwick, 1981) discusses the dualism of practical reason. According to Sidgwick, what is morally right can sometimes conflict with what is prudentially right. Since our duty and self-interest are derived from different basic principles, it will be impossible to find a common scale on which the two kinds of right can be weighed. This gives rise to a dualism of practical reason that on some occasions can lead to conflicting requirements and incommensurability.

However, the possibility of incommensurability can also provide opportunities. The dilemma of Sartre's student can illustrate the importance of creating an identity and meaning in life. The fact that neither option was at least as good as the other means that the student could make the choice in another way: he could invent a solution. Thus, recognizing that moral codes sometimes fail to determine what one ought to do (and thus by recognizing incommensurability) creates a space for agency, individual freedom, and creativity. At a more general level, by accepting that options are sometimes incommensurable with respect to some criteria (e.g., what is morally and prudentially right), theoretical space for other criteria is created (e.g., what is most in line with one's previous choices).

In the rest of this introduction, we will provide some historical background to the philosophical work on incommensurability, as well as an overview of contemporary research on the subject. At the end of the introduction, we will provide summaries of the chapters in this volume.

1 The History of Incommensurability

Philosophical interest in comparability problems dates back at least to the Pythagoreans, a philosophical-mystical movement that was thriving in the sixth century BC. According to Western tradition, based largely on some remarks of Aristotle in his *Metaphysics* (Aristotle, 1984: 986a, 987b, 1080b), the Pythagoreans believed that intrinsic properties of numbers and their ratios (*arithmos*) could explain all there is in the world: "They assumed the elements of numbers to be the elements of everything, and the whole universe to be a proportion or number" (Aristotle, 1984: 986a). As a historian of mathematics put it, "the essence of all things, in geometry as well as in the practical and theoretical affairs of man, is explained in terms of *arithmos*, or intrinsic properties of whole numbers or their ratios" (Boyer and Merzbach 2011: 65). Emboldened by this general belief in the explanatory powers of mathematics, the Pythagoreans

took an interest in studying both relations between mathematical objects and more worldly phenomena. To their great chagrin, however, while studying geometry, they discovered that sometimes whole numbers and their ratios cannot describe all relations, the paradigmatic example being the ratio of the lengths of the side and the diagonal of a unit square, although it has been argued that they first discovered that all whole numbers and their ratios cannot describe all relations by studying pentagons (Von Fritz, 1970).[2]

In modern times, the term "incommensurability" has been used with two very different meanings. On one hand, it has come to play a prominent role in the philosophy of science and philosophy of language ever since Thomas Kuhn and Paul Feyerabend used the term to describe the possibility that certain propositions cannot be translated from one scientific framework (or conceptual schema) to another (Feyerabend, 1978; Kuhn 1962). If Kuhn and Feyerabend are right about the possibility of such incommensurability, there could, for instance, be propositions within the Pythagorean scientific framework that cannot be translated or even understood within our modern scientific framework. We would have reason to doubt that we could ever, even in principle, understand what the Pythagoreans meant when they described the world.[3]

The term "incommensurability" has also stood in the center of discussions of potential problems when comparing the value or choiceworthiness of different things in ethics, value theory, and political philosophy. This use of the term can in modern times be traced to Isaiah Berlin, who, in his 1958 lecture "Two Concepts of Liberty," suggested that the incompatibility between different types of liberties might be an instantiation of a clash between absolute and incommensurable values (Berlin 1969). Following Berlin, in the second half of the twentieth century, there is rich literature that outlines different suggestions as to what one might mean by claiming that values are incommensurable. For instance, Bernard Williams suggested that the claim could mean (i) denial of the idea that there is a unique currency that can be used to compare all values; (ii) denial of the idea that, for each value conflict, there is some independent value that can resolve the conflict; (iii) denial of the idea that, for each value conflict, there is some value an appeal to which can rationally resolve the conflict; or (iv) that there is no rational solution to any value conflict (Williams 1981: 77). Meanwhile, James Griffin suggested that there are five kinds of incommensurability: (a) *incomparability*—neither of two values is more valuable than the other, but they are not equally valuable either, (b) *trumping*—any amount of one value is more valuable than *any* amount of another value, (c) *weighting*—differently weighted values cannot be reduced to an unweighted common scale, (d) *discontinuity*—a certain amount of some value is more valuable than *any* amount of another value, and (e) *pluralism*—there are irreducibly many values (Griffin 1986: 77–92).

The contemporary notion of incommensurability, which is the focus of this book, can be said to originate in the work of Joseph Raz and his focus on the failure of transitivity and the identification of the "mark of incommensurability" (Raz 1986: 325–326). According to Raz, incommensurability (which he equates with incomparability) obtains when "neither of [two options] is better than the other nor are they of equal value" (Raz 1986: 326). He proposes that one can identify incommensurability with the following test:

> Two valuable options are incommensurable if (1) neither is better than the other, and (2) there is (or could be) another option which is better than one but is not better than the other.
>
> (Raz 1986: 325)

This passage is important for the development of research on incommensurability in two ways. On one hand, it expresses a focus on *valuable options* (as opposed to previous writings that oscillate between talking about options, values, and the general state of affairs). On the other hand, Raz identifies the "mark of incommensurability" by providing a test that makes salient the general conditions on which most contemporary research on incommensurability focuses. (However, much contemporary incommensurability research disagrees with Raz about whether this indicates incomparability or some other phenomenon such as indeterminacy or parity):

> We have here a simple way of determining whether two options are commensurate given that neither is better than the other. If it is possible for one of them to be improved without thereby becoming better than the other, or if there can be another option which is better than the one but not better than the other, then the two original options are incommensurate.
>
> (Raz 1986: 325–326)

Contemporary with Griffin's and Raz's discussions in the 1980s and inspired also by work in social choice theory (e.g., Sen 1970; Fishburn 1973), Derek Parfit discussed the possibility that the not worse than relation was not transitive. For Parfit, puzzling results in population ethics indicated the possibility that two things need not always be related with a *precise* better than, worse than, or equally as good relation (Parfit 1984: 431). Each of these three value relations could be "roughed up," according to Parfit. Much contemporary discussion of incommensurability continues in the tradition started by these three philosophers, and it is against this backdrop the contemporary discussions of incommensurability are placed.

2 Terminology

An important controversy in the background of philosophical discussions of the nature, importance, and implications of incommensurability has concerned what to actually call the phenomena one is interested in. As seen earlier, the term "incommensurability" has been used with many different meanings. This has led to different authors using different terms to refer to the same phenomenon, but it has also led authors to use the same term to refer to different phenomena. To readers not familiar with the literature, this can be very confusing.

The two most common terms in the discourse are "incommensurable" and "incomparable." Some, such as Joseph Raz, use these terms interchangeably (Raz 1986; Anderson 1993). Others, such as Ruth Chang, argue that "incommensurability" ought to be given its etymologically correct meaning: two things that cannot be placed on the same cardinal scale are incommensurable (Chang 1997, 2015). If two things are incommensurable, we cannot say how much better one is in comparison with the other in terms of some cardinal scale of units of value. It might, however, be possible to say that one of them is better than the other. In situations in which no positive characterization can be given of the comparative relation that holds between two objects, Chang uses the term "incomparable." She thinks that it is important to distinguish incommensurability from incomparability because she thinks that each notion captures a distinctly important phenomenon.

If a value bearer is neither better than, worse than, nor equally as good as the other, they have commonly been believed to be incomparable. However, as some argue, value bearers can be positively related in more ways than previously believed. For this reason, "incomparable" ought not to be defined in terms of "better than," "worse than," and "equally as good," but rather in terms of a lack of *any* positive, binary, generic comparative relation. Furthermore, it seems reasonable to claim that incomparability is a ternary relation. A is incomparable with B *with respect to C* if and only if no positive, binary, generic comparative relation holds between A and B with respect to C. The third item, C, is what Chang (1997) refers to as the "covering consideration." It does not make much sense to say that "Paris is better than Slough" *simpliciter*.[4] Such claims only make sense when we clarify what covering consideration we have in mind; for example, "Paris is better than Slough with respect to a romantic and cultural weekend destination." The same is true for incomparability claims: we need to specify in respect to what covering consideration the things are supposed to be incomparable.

Introducing the idea of covering consideration also allows us to distinguish *noncomparability* from incomparability. Some covering considerations do not apply to some value bearers, and this allows for noncomparability. For instance, the claim that the color blue is better than the number four with respect to taste seems like an absurdity; what

we have here is an example of noncomparability (Chang 1997). This somewhat absurd comparison problem seems, however, less worrying from the perspective of practical reasoning. Why would we ever need to compare a color and a number with respect to their tastiness?

Some follow Chang's use of the term "incomparable" but interpret "incommensurable" somewhat differently. They use the term to refer to a relation such that if it holds between two items, then neither is determinately better than the other nor are they determinately equally as good (Rabinowicz 2008). "Incommensurability" used as a catch-all term is thus compatible with, among other possibilities, incomparability and parity, and all non-conventional value relations fall under its umbrella. Other terms used to refer to this plethora of possible value relations are "non-determinacy" (Herlitz 2019, 2020) and "triply not true" (Boot 2017).

While we agree that Chang's use of the term "incommensurable" seems etymologically correct and that there is good reason to keep the ideas of not being cardinally measurable and not being comparable apart, since there is no standard umbrella term to denote the range of possible ways to understand hard comparisons, we will co-opt the term "incommensurability" for practical reasons to encompass all non-conventional comparative relations. We will thus use the term "incommensurable" in this introduction but encourage the reader to be aware of the different uses of the term by different authors throughout the book.

3 Reasons to Accept Incommensurability

People are attracted by the possibility of incommensurability for many different reasons. A straightforward reason to accept the prevalence of incommensurability is that some of the things we value (and have reason to value) are very different from one another. For instance, most people value pleasure, social relations, and achieving goals that they set for themselves. For those who accept that there is a plethora of practical reasons and values and adhere to some pluralistic account of reasons and value, it seems highly plausible that these can come into conflict in the sense that we cannot determinately judge how they relate to one another. At the very least, one needs an argument for the opposite view, the assumption that it is always, in principle, possible to determine which of two options is better than the other or that they are equally good. The burden of proof seems to lie with those who reject incommensurability.

In what follows, we will present what seem like the most convincing arguments for the possibility of incommensurability being instantiated.

3.1 Dilemmas

The possibility of incommensurability can be argued for by referring to the existence of moral dilemmas. Consider the situation Sophie faces in the novel *Sophie's Choice* (Styron 1979). When Sophie and her two

children arrive at a concentration camp in Nazi Germany, a guard forces her to choose which child gets to live and which child gets to die. If she does not choose, both children will be killed. In such a horrific situation the alternatives seem incomparable. It is, however, not clear whether this constitutes an example of incommensurability. Perhaps the alternatives are equally as bad and the dilemma instead ought to be understood in terms of conflicting obligations (Sinnott-Armstrong 1988). Or perhaps moral dilemmas only show that some values are irreplaceable. This could mean that even if one alternative is better than another, the value of the outcome cannot fully replace the loss that comes with not choosing the other alternative (Griffin 1997).

3.2 The Wrongness of Comparisons

Joseph Raz (1986) introduced the notion of *constitutive* incomparability. He argued that certain things *ought* not to be compared. For example, money ought not to be compared with spending time with one's spouse. Exchanging money for time with one's loved ones has symbolic significance. According to Raz (1986: 348), "Significant social forms, which delineate the basic shape of the projects and relationships which constitute human well-being, depend on a combination of incommensurability with a total refusal even to consider exchanging one incommensurate option for another." To have certain relationships one must thereby conclude that certain values are incommensurable: the relationship itself is partly constituted by the belief that it is not exchangeable for money.

Elizabeth Anderson (1993) has developed a similar argument in favor of incommensurability. Sometimes when we make a comparison, it makes no sense to compare the two goods, all things considered. If the value of an object is understood in terms of its being rational to value the object, then, since it is not rational to make an all-things-considered comparison, the objects must be incommensurable. Anderson outlines the following conditions under which she believes incommensurability (which she, following Raz, thinks of as incomparability) is present:

(1) The goods in question meet the standards measured by the scale in very different ways.
(2) There are no gross differences in the degree to which each good exemplifies its own way of meeting the standards.
(3) Meeting the standard in one way is not categorically superior to meeting it the other way.

(Anderson 1993: 55)

3.3 The Small Improvement Argument

The currently most influential argument in favor of incommensurability is the so-called small improvement argument. In the context of axiology

and normativity more generally, it has, among others, been advanced by Joseph Raz (1986), Derek Parfit (1984), Walter Sinnott-Armstrong (1985), and Ruth Chang (2002). However, a similar argument has been discussed in the context of preference relations since the 1950s (Savage 1954; Luce 1958; de Sousa 1974).

The argument, which is meant to establish that the "at least as good as" relation does not exhaust the domain of value relations, takes the following form:

> Premise 1: It is false that A is better than B and it is false that B is better than A.
> Premise 2: A+ is better than A.
> Premise 3: A+ is not better than B.
> Conclusion: It is false that A is better than B, it is false that B is better than A, and it is false that they are equally as good.

Premises 2 and 3 establish that A and B are not equally as good. Betterness is transitive across equal goodness, so if A and B are equally as good, then, if A+ is better than A, A+ must also be better than B. From this and Premise 1, the conclusion follows.

This is, however, only the structure of the argument. Examples that satisfy the structure described in the small improvement argument must be given in order to conclude that things can be incommensurable. The choice between two different holiday destinations could be such an example. If we compare a week in Iceland with a week in Seoul, it is likely that we will find neither to be better than the other, and if we slightly improve one of the alternatives, it will still not be judged to be better than the other. A small improvement of the meal during the flight to Reykjavik would not make the trip to Iceland clearly better than the week in Seoul. In the literature, many similar examples are given.

3.4 Solving Difficult Problems in Ethical Theory

Another powerful, but indirect, argument that some options are incommensurable can be derived from contemporary research on paradoxes in ethical theory. Accepting incommensurability seems to provide a way of avoiding or explaining certain paradoxes. Many paradoxes in ethical theory rely on the background assumption that all options are related by an "at least as good" relation and that this relation is transitive. That is, if A is at least as good as B, and B is at least as good as C, then A must be at least as good as C. If that assumption is valid, then it seems that paradoxes can be established. The prime example of such a paradox is the fact that there are compelling arguments in favor of the so-called repugnant conclusion, although this seems deeply counterintuitive. That is, consider two possible populations, X and Y, with perfect equality and in which

everyone has lives worth living. Everyone in population Y has slightly lower well-being than the people in population X, but the population in Y is much bigger so that with respect to the total sum of well-being, Y is better than X. It seems that for all such pairs of populations, the Y population is better than the X population. By transitivity, this means that the following deeply counterintuitive proposition is true:

> For any possible population of at least ten billion people, all with very high quality of life, there must be some much larger imaginable population whose existence, if other things are equal, would be better, even though its members have lives that are barely worth living.
>
> (Parfit 1984: 388)

Similarly structured arguments have been developed to support other deeply counterintuitive ideas – for instance, that there is a number of headaches that is so large so that it would be better to cure them than to save a life (Norcross 1997; Voorhoeve 2014; Herlitz 2018).[5]

One reason to accept the possibility of incommensurability is that it potentially provides a way of avoiding these paradoxes (Griffin 1986; Blackorby et al. 1996, 1997; Broome 2004; Qizilbash 2007; Chang 2016; Rabinowicz 2021). To establish the repugnant conclusion, all populations of X- and Y-type must relate to each other via a chain of "at least as good" relations. It is this fact that enables the series of inferences that leads one from saying that a population of 20 billion people with fantastic lives is better than a population of 10 billion people with slightly better lives to the conclusion that a population of people with lives barely worth living is better than the population of 10 billion people with amazing lives, as long as the population of people with lives barely worth living is large enough. If populations could be incommensurable, that series of inferences would be blocked. If some X- and Y-type populations were incommensurable, the repugnant conclusion would be avoided.

4 Places of Incommensurability

Incommensurability can be possible in many fields, in both ethics and other areas. Whenever one can make a small improvement argument, one can make a case for incommensurability, and this seems possible any time one makes multidimensional comparisons. Multidimensional comparisons are ubiquitous. In ethical theory, they are almost always present since plausible conceptions of well-being are multidimensional (Griffin 1986). At the very least, they typically ascribe value to both life extensions and improvements in the quality of life within time periods. All ethical theories that ascribe importance to individual well-being and its distribution thereby either must explain what is wrong with the small improvement argument or accept that alternatives that differ with respect

to how well they extend lives and how well they improve the quality of life within time periods can be incommensurable (Herlitz 2017).

As often is the case, discussions in ethics have close cousins in aesthetics. Examples meant to establish that there is incommensurability in the domain of morality are often examples that belong to aesthetics (this is probably because they are less likely to be offensive or dependent on substantive moral theories). It does not seem absurd to claim that the aesthetic values of two pieces of art—an archetypical type of multidimensionality—are incommensurable.

Multidimensional comparisons are also common outside of ethics and aesthetics. Anyone who accepts that both observations in support of the theory and simplicity count in favor of a scientific theory will face the possibility that one theory is supported by more observations while a competing theory is simpler, such that neither theory has more that counts in its favor. Will one additional minor observation necessarily tilt the scale and provide sufficient evidence to abandon one of the theories? Similar arguments can be presented in epistemology, where reasons to believe can be of different types.

Moreover, incommensurability should be relevant to all fields of research that somehow consider trade-offs. Roughly, trade-offs involve compensation – i.e., the loss of one good is compensated for by the gain of another. Justifications of trade-offs thus seem to presuppose that the goods can be compared and adequately ranked such that the trade-offs constitute a satisfactory compensation.

The relevance for decision theory is also obvious. This is a natural connection considering the close ties between axiology and decision theory, but there are also other reasons that speak in favor of the possibility of incommensurable options. A prime example of this is the possibility of incommensurable probabilities that was discussed by Keynes (1921). More recently, Wlodek Rabinowicz (2017, 2020) has considered this possibility and connected the discussion even more firmly to incommensurability within axiology.

5 Accounts of Incommensurability

A significant part of contemporary research on incommensurability has focused on how to explain the phenomenon. What explains the presence of the mark of incommensurability?

5.1 Parity

While the small improvement argument seems to suggest that value bearers can be incomparable, it has been questioned whether this must be the case. It has been argued that value bearers that stand in the relation described by the argument may nevertheless be comparable. Intuitively,

there is something to this suggestion. Parfit, for example, argues that things that are related as described by the small improvement argument can be roughly comparable (1984: 431). In his example, a certain novelist may be neither better than, worse than, nor equally as good as a certain poet, but they can be roughly comparable.[6]

Chang has presented a formal argument that is meant to establish that those value bearers that are related as described by the small improvement argument can nevertheless be comparable. Her so-called chaining argument can roughly be described as follows:

Imagine that we make a small unidimensional change to the novelist. After a series of such changes, we could end up with a very bad novelist. This very bad novelist is clearly worse than the poet. Yet it also seems to be the case that "between two evaluatively very different items, a small unidimensional difference cannot trigger incomparability where before there was comparability" (Chang 2002: 674). This means that if the very bad novelist is comparable to the poet, then so must the original novelist be.

Chang argues that value bearers related in the manner described by the small improvement argument are neither better than each other, equally as good as, nor incommensurable but "on a par." According to Chang, this is a fourth *sui generis* and previously overlooked value relation. The soundness of her chaining argument has, however, been questioned (Boot 2009; Elson 2014a; Andersson 2016b).

5.2 Vagueness

The claim that there is a fourth positive comparative relation has been challenged by those who hold a more conservative view of comparative relations. It has been argued that there is no need to assume that value bearers can be on a par or even incomparable. Epistemic explanations may suffice to account for some of the alleged cases of parity and incomparability (Regan 1997), whereas other cases could possibly be accounted for in terms of vagueness. This view has been defended by John Broome (1997). He presented his collapsing argument to show that vagueness crowds out non-conventional comparative relations. The argument is meant to establish that if we assume that there is vagueness, then there is no room left for comparative relations such as incomparability and parity. His argument, however, depends on the collapsing principle, which many find hard to accept (Carlson 2004, 2013; Elson 2014b; Gustafsson 2018; Andersson and Herlitz 2018).

This vagueness approach is, however, not dependent on the collapsing argument. Others have argued that the possibility of vagueness allows us to be skeptical of claims about parity and incomparability (Andersson 2017; Elson 2017). A consequence of vagueness is that we cannot be

determinately certain that the result of the small improvement argument holds, and for that reason, we should be conservative and not embrace the possibility of further comparative relations (Wasserman 2004; Klocksiem 2010; Gustafsson 2013).

5.3 *Formal Accounts of Incommensurability*

Presenting examples of purported non-conventional comparative relations raises the question of how they should be analyzed. Some have argued that parity can be understood in terms of the three standard comparative relations (Gert 2004; Carlson 2010). Similarly, Nien-hê Hsieh (2005) has argued that non-conventional relations are to be understood in terms of "clumpy values": values do not provide fine-grained orderings but rather place items in clumps, so equally as good items can appear to be, for example, on a par.

Chang (1997) has suggested that the evaluative difference between two things that are on a par can be understood as a nonzero unbiased difference. It is nonzero because it has a magnitude—just as when one thing is better than another there is a nonzero difference—and it is unbiased because the difference does not have a valence—in the same way as the difference between two things that are equally as good is unbiased.

Inspired by Joshua Gert's account, Wlodek Rabinowicz (2008, 2012) has provided a fitting attitudes account (henceforth FA) of non-conventional value relations. The FA accounts for value in terms of a normative component (e.g., fittingness) and an attitudinal component (e.g., favor). Goodness could consequently be analyzed in terms of fittingness to favor. If we acknowledge that there can be two levels of normativity – i.e., requirement and permissibility – this account can also encompass conventional and non-conventional value relations. For example, X is better than Y if and only if it is rationally required to prefer X to Y, and X and Y are on a par if and only if it is rationally permissible to prefer X to Y and also rationally permissible to prefer Y to X. The two levels of normativity and preferences, including preference gaps, can be combined in 15 different ways, which entails that the FA have the conceptual space for as many as 15 possible value relations.

6 Problems for Normative Theory

Incommensurability poses problems for normative theory. If incommensurability is possible, it is possible that there is no available option that is at least as good or choiceworthy as all the alternatives. It might be the case, for instance, that two options are better than all other alternatives, but they are not equally good. Some have suggested that this means that it is impossible to act rationally.

6.1 No Optimal Option, Seemingly Solved by Changing to Maximal

According to standard rational choice theory, choosing an option is rational if and only if that option is at least as good as every alternative (i.e., it is an option that is *optimal*). If incommensurability is possible, it will be possible that no option is optimal, which would mean that it is impossible to make a rational choice. For instance, if an agent faces the choice between an apple, a pear, and an orange, and if the apple is incommensurable with both the pear and the orange while the pear is better than the orange, there is no optimal option. Neither the apple nor the pear is at least as good as all the alternatives since they are incommensurable with each other, and for an option to be at least as good as another option it must be either better than or equally as good.

This appears to be a significant problem. However, there are some proposed solutions to it. Somewhat depending on how one explains incommensurability, one can revise the criterion for what makes a choice rational (Herlitz 2019). For instance, one can follow Amartya Sen (1997) and say that choosing an option is rational if and only if that option is not worse than any alternative – i.e., it is an option that is *maximal*. With this definition of what it is to choose rationally, it follows that it is rational to choose either the apple or the pear. The orange is worse than the pear and is thereby not a rational choice, but neither the apple nor the pear is worse than any alternative (they are both incommensurable with one available alternative—each other—but being incommensurable with is not the same as being worse than). This works well if incommensurability is some non-conventional comparative relation and similar moves can be made by those who believe that incommensurability is vagueness. If the incommensurability is due to vagueness (i.e., on one permissible sharpening of the covering concept, the apple is better than both the pear and the orange and on another the pear and the orange are better than the apple while the pear remains better than the orange), one can revise the rational choice criterion and say that choosing an option is rational if and only if that option is at least as good as all alternatives on at least one permissible sharpening (Broome 2009). Similar moves are available to proponents of the FA approach (Rabinowicz 2008; Herlitz 2019). Since the apple is optimal on one permissible sharpening and the pear optimal on another, choosing either is rational.

This raises an interesting question. Although revising the conception of rational choice so that all options that are maximal are rational solves some initial problems, it is possible that something has gotten lost on the way. Chang has argued that this is the case, that maximality is not a sufficient criterion for a choice being rational (Chang 2013). But if it is not sufficient, one might wonder if it is *never* sufficient. It could be the case that maximality sometimes is enough, but that there are situations in which it is not (Andersson and Herlitz 2021). This is a topic that deserves more attention.

6.2 *Dynamic Inconsistency*

Although the most obvious decision-theoretical problem has some rather simple technical solutions, there are other such problems that pose more serious technical difficulties. One of the biggest problems is that acceptance of incommensurability in combination with revisions of the rational choice criterion leads to dynamic inconsistency (McClennen 1990; Andreou 2016). When an agent faces a sequence of choices, it is possible for her to form a series of maximal choices but end up with an option that is worse than some alternative that she could have had if she had chosen differently. To see this, consider an agent that faces a sequence of choices such that she at t_1 chooses between the pear and the apple, and regardless of what she chooses, she is offered the opportunity to change to the orange at t_2. If choosing a maximal option is all that rationality requires, in this situation, it would be rational to choose the apple at t_1 and then change to the orange at t_2. At t_1, both the apple and the pear are rational choices, so it is rational to choose the apple. At t_2, both the apple and the orange are rational choices, so it is rational to choose the orange. However, the orange is worse than the pear, and she could have had the pear if she had chosen differently at each choice node. By making choices that in isolation appear rational (in virtue of being maximal), the agent has ended up with an option that appears irrational (in terms of not being maximal) when the sequence of choices is considered.

The dynamic inconsistency also enables a weak kind of money-pump argument. To see this, consider a case in which at t_1, the agent has a choice between the apple and the pear, and then, if she chooses the apple, she is at t_2 offered the chance to change the apple to the orange. If she accepts this exchange, she is at t_3 offered the chance to change the orange to the pear if she pays a small amount (which she is willing to do since the pear is better than the orange). These three exchanges get the agent to the pear, but she could have had the pear without cost if she chose it at t_1.

There are various suggestions for how to avoid this problem in the literature, but no solution seems unproblematic. For instance, it has been proposed that it is rational to be "resolute," to make a plan and stick to it (McClennen 1990). But why would it be irrational to revise plans and act on how the alternatives relate to each other? It has also been suggested that rational decision-makers are "sophisticated" and use backward induction to foresee the problems that await them when they face a sequence of choices (e.g., Rabinowicz 1995). But that solution can rule out as irrational options that are maximal. How to avoid dynamic inconsistency while accepting incommensurability remains a debated topic in contemporary research.

6.3 Basic Contraction Consistency and Cycles

According to some views of how to act rationally when some maximal options are incommensurable with respect to a primary decision criterion, one ought to invoke a second decision criterion and choose in accordance with how the options relate to this. For instance, if the apple and the pear are incommensurable with respect to how good they taste, one compares them with respect to how much they weigh, the lighter the better. These approaches risk violating rationality requirements. To see this, consider the possibility that the orange is lighter than the apple and the apple lighter than the pear. If one faces all three options at once, this kind of two-step approach suggests that one should choose the apple. The orange is discarded since it is worse than the pear, and of the pear and the apple, the apple is lighter. However, if one faced a choice between only the apple and the orange, the approach says that one should choose the orange since the apple and the orange are incommensurable with respect to taste, and the orange is lighter than the apple. This means that the approach tells one to choose one option in one set but choose a different option in a subset of that set in which the first option is available. So-called "basic contraction consistency" is violated (Herlitz 2019). If one instead uses the approach to make pairwise comparisons of the options, one ends up with a cyclical ranking of them: the orange is better than the apple, the apple is better than the pear, and the pear is better than the orange (Herlitz 2020).

   This illustrates how different approaches to forming choices when maximal options are incommensurable risk violating rationality requirements. There are possible ways to avoid this, but they require imposing additional formal constraints on what kind of decision-making is acceptable when some options are incommensurable (Herlitz 2019).

6.4 Risk

Another set of complications arises once one considers what to do in situations that involve uncertainties (Hare 2010). Consider again the choice between an apple and a pear, but instead of choosing fruit, one can only choose one of two lotteries, the outcome of which depends on a coin toss. In the first lottery, one wins an apple on heads and a pear on tails. In the second lottery, one wins a pear plus ten cents on heads and an apple plus ten cents on tails. Now, in light of the small improvement argument, assume that ten cents is a small enough improvement that it does not change the comparative relation between the apple and the pear. An apple plus ten cents is incommensurable with a pear, and a pear plus ten cents is incommensurable with an apple. This gives us the puzzling situation in which the possible outcomes in each state are incommensurable yet the acts do not seem incommensurable since the latter lottery is certain to lead to the additional ten cents.

It is not obvious what one ought to do in this situation. Whereas some have argued that it is only rational to choose the latter lottery (Hare 2010; Bader, 2018), others have said that it can be rational to choose either (Schoenfield 2014; Bales et al. 2014; Doody 2019).

7 Opportunities

Although incommensurability has primarily received attention due to the problems it causes, there are also opportunities that arise once one accepts incommensurability. If one assumes that all items can always be compared in conventional ways so that it will always be the case that one item is at least as good as the other with respect to the criterion one applies, the criteria in question may seem to be ascribed too much importance. Accepting incommensurability is one way in which one can reduce the importance of different criteria, something which may seem very desirable in certain contexts.

7.1 *Chang and Going Hybrid*

Chang has argued that we ought to accept that some items are sometimes incommensurable (in her terminology, "on a par") with respect to given reasons and that, in these situations, we must resort to reasons grounded in the will (i.e., voluntarist reasons) to determine what to do (Chang 2013). For Chang, given reasons are important because she believes they are reasons that, independently of the will of the agent, determine that things such as sadistic torture are wrong. However, by accepting that these given reasons admit of incommensurability, her theory does not ascribe to given reasons the importance of always fully determining what one ought to do, something which might seem exigent since it seems to reduce agents to automatons whose task is mainly to respond to reason. By accepting incommensurability, Chang can propose that reasons grounded in the will of the agent must sometimes complement given reasons, thereby reducing the importance of given reasons and providing a role for human agency in her metaethical theory. This illustrates how incommensurability provides a theoretical opportunity to avoid certain problems.

7.2 *Law*

In the study of law, accepting incommensurability has been suggested to be one way of providing space for different valuations in society and for making progress in legal theory (Sunstein, 1994). By rejecting the idea that there is a single one-dimensional (or monistic) correct valuation of goods such as those provided by utilitarians and instead embracing pluralism and the idea that there are many different and equally valid valuations, what is at stake in legal disputes can be elucidated. Rather

than seeing legal disputes as disagreements about facts about which there is, fundamentally, a correct answer, legal disputes can be conceived of as disagreements about options that are fundamentally incommensurable. This, in turn, allows one to think differently about the role of the law and about how to make progress in hard cases: it shows that one needs a substantive theory of law (Sunstein, 1994).

7.3 Prerogatives

Accepting incommensurability is also a way of providing space for prerogatives, both individual and collective. Sometimes, it seems clear that it is up to the agent, be it an individual or a collective such as a state, to settle normative questions. One way of explaining this in a normative-theoretical framework is to accept incommensurability with respect to normative principles. In clinical ethics, a plausible normative principle says that clinicians should promote health, seeking to improve the health of their patients to the greatest extent possible. However, it also seems plausible that patients ought to have some influence over what treatments they receive. It has been suggested that patient prerogatives can be given a role in clinical decision-making if one accepts that treatment options can be incommensurable with respect to health (Herlitz 2017). If treatments A and B are incommensurable with respect to health (perhaps because one adds more years to the patient's life and the other improves the health-related quality of life more), it seems plausible to say that the patient ought to choose between them, and if one accepts incommensurability, one does not have to accept that there is a conflict between patients' wills and universal normative principles such as health promotion. Similarly, religious freedom is a plausible political–philosophical principle. Meanwhile, it seems desirable to say that it should be up to the society in question to determine whether this principle requires that one grant permission for the building of a minaret in a neighborhood where many of the residents are refugees who have experienced traumas due to violence in the name of Islam (Herlitz and Sadek 2021). If one accepts that different policies (e.g., grant or do not grant permission to build a minaret) can be incommensurable with respect to principles such as religious freedom, it is easy to explain collective prerogatives.

8 Summaries of the Chapters

The book covers four aspects relating to incommensurability:

In the first part of the book, new research on how to best explain incommensurability is presented. In contemporary debate, two competing explanations stand out. According to some, such as Ruth Chang (2002) and Wlodek Rabinowicz (2008, 2012), the fact that no conventional

comparative relation holds between some options can sometimes be explained by the fact that there are non-conventional comparative relations. The most widely discussed non-conventional comparative relation is "parity." On this view, things need not be related by a "better than," "worse than," or "equally as good as" relation but may also be on a par. On the competing view, espoused by, for example, John Broome (1997), Luke Elson (2017), and Henrik Andersson (2017), there is no need to introduce a fourth comparative relation to account for incommensurability. Instead, it is argued that incommensurability is in fact indeterminacy due to vagueness. According to this view, one of the three conventional relations does in fact hold but it is indeterminate which one.

In Chapter 1, John Broome argues that incommensurability is vagueness. While he has previously argued for this possibility by referring to the so-called collapsing principle, he now provides a new argument. Central to his novel argument is the claim that competing views cannot adequately account for our intuitions about hard cases. Broome concludes that it is only vagueness that can give a satisfactory account of the phenomenon.

This can be contrasted with Ruth Chang's contribution in Chapter 2. Chang is well-known for her seminal argument that incommensurability ought to be understood in terms of a fourth *sui generis* value relation that she calls "parity." In Chapter 2, she explains why the phenomenon of interest—what we have here called "incommensurability"—is not vagueness but is instead the holding of a fourth *sui generis* value relation that she calls "parity."

In Chapter 3, Chrisoula Andreou argues that there might be even more value relations beyond "equally as good," "better than," "worse than," and "on a par." According to her, the two terms "roughly equal as" and "on a par," which are often used to refer to the same value relation, actually denote two different phenomena. With this contribution, she expands the space for value relations and provides an explanation as to how the value relations differ.

In the second part of the book, the role of incommensurability in dismissals of so-called spectrum arguments and in population ethics is discussed. Population ethicists such as Derek Parfit (2016), Mozaffar Qizilbash (2007), and Wlodek Rabinowicz (2021) have recognized that accepting incommensurability can help one avoid perennial problems in population ethics. It has been argued that incommensurability can be invoked to avoid the repugnant conclusion and the mere addition paradox and to block so-called spectrum arguments.

In Chapter 4, Mozaffar Qizilbash contrasts Derek Parfit's suggestion as to how to avoid the repugnant conclusion with James Griffin's. Central to both of their suggestions are phenomena that can be called "incommensurability." Parfit invokes the concept of imprecision, which seems to share many characteristics with parity, and Griffin invokes the

concept of discontinuity. While the latter concept is sometimes called a form of incommensurability, it can nevertheless be accounted for in terms of the three conventional value relations. The details of their insightful views are expounded and the merits of their views are thoroughly discussed.

In Chapter 5, Henrik Andersson discusses a specific spectrum argument that has been advanced by Larry Temkin (2012). Temkin's conclusion is that the betterness relation is not transitive. Andersson rejects this conclusion by arguing that certain assumptions made by Temkin can be questioned and that the possibility of indeterminacy and so-called Millian inferiority can show why we need not accept Temkin's conclusion. The unwanted conclusion of the spectrum argument can thus be avoided without referring to incommensurability.

In Chapter 6, Gustaf Arrhenius explores the idea that one can avoid impossibility results in population axiology by accepting incommensurability. Arrhenius illustrates that although some problems in population axiology can be avoided by accepting incommensurability, other problems remain.

In the third part of the book, the role of incommensurability in decision theory is addressed. It is well-known that allowing for incommensurability causes dynamic choice problems. Relatedly, it has recently been shown that certain decision strategies that seem appealing when dealing with situations characterized by incommensurability violate basic requirements of rationality. The decision strategies might violate basic contraction consistency, and they might lead to cyclical evaluations of alternatives.

The potential problems for rational decision-making are thoroughly discussed in Chapter 7, in which Anders Herlitz illustrates how different plausible decision strategies for choice situations characterized by incommensurability risk violating rationality requirements. He argues that unless one introduces formal constraints on how to choose between incommensurable options, one risks forming cyclical evaluations or violating basic contraction consistency. The question then arises as to what grounds these formal constraints. Herlitz does not settle this issue but presents some alternative ideas: they might be grounded in the comparative relation (in incommensurability), in the domain of reasons, or in rationality itself.

In Chapter 8, Krister Bykvist presents a novel contribution on the topic of rational decision-making, developing an account of so-called cross-categorical comparisons. The comparison "this table is a better table than Trump is a president" would constitute an example of such a comparison. An account of these comparisons is important since they can, for example, allow us to compare well-being across people and make cross-theory comparisons of value when we face evaluative uncertainty.

In Chapter 9, Luke Elson contrasts two influential views of the consequences of incommensurability for agency. More precisely, he provides a detailed account of the consequences of Chang's hybrid view of rational choice and of Joseph Raz's classical view. Elson concludes that by adopting a vagueness account of incommensurability we are not forced to accept such radical views of agency.

In the fourth and final part of the book, incommensurability and risk is investigated.

In Chapter 10, Wlodek Rabinowicz addresses a puzzling problem presented by Caspar Hare (2010) concerning rational choice. Rabinowicz, however, considers it in the domain of formal axiology: how can one action be better than another despite the fact that their outcomes are bound to be incommensurable? By invoking a fitting attitudes analysis of value, he argues that even if the outcomes are on a par, it is possible for one action to be better than another. Whether this also is possible if the outcomes are bound to be incomparable (which is a radical form of incommensurability) is left as an open problem.

In Chapter 11, Katie Steele considers the possibility of ignoring criteria under which options are incommensurable. Steele focuses on a non-dominance relation that is often appealed to in discussions of incommensurability and uncertainty. She concludes that such a principle is not plausible in the multi-criteria setting that she considers and that we ought not to ignore criteria under which options are incommensurable.

In Chapter 12, Ryan Doody presents an impossibility result for decision theories under risk. More specifically, he argues that no plausible decision theory for agents with incomplete preferences can both satisfy a dominance principle and accept a transitivity requirement.

Notes

1 We are grateful to Ruth Chang, Mozaffar Qizilbash, Wlodek Rabinowicz, and Joseph Raz for the valuable feedback they have given us on this introduction. We also wish to thank the participants at the Philosophy, politics and economics seminar at the Institute for Futures Studies in Stockholm at which a draft of this introduction was presented.
2 The Pythagoreans, of course, admitted that a comparison could be made, as the diagonal of a square is clearly longer than its side. It should also be stressed that the Pythagorean view of commensurability is a narrow conception, as the relation between the diagonal and the side can be expressed by an irrational number.
3 For a rather devastating criticism of this idea, see Davidson (1973).
4 This is similar to the claims made by, for example, Peter Geach (1956) and Judith Jarvis Thomson (1997). Geach famously argued that there is no such thing as goodness *simpliciter*, and Thomson expressed a similar view, explicitly discussing betterness *simpliciter*. For the connection between these discussions and value incommensurability, see Andersson (2016a).
5 For an in-depth discussion of arguments of this kind, see Temkin (2012).

6 Similarly, Griffin (1986: 80–1) argued that they may be "roughly equal," but he probably had a view similar to the vagueness view in mind, according to which it could be indeterminate how value bearers relate.

References

Anderson, E. (1993), *Value in Ethics and Economics* (Harvard University Press).

Andersson, H. (2016a), 'Vagueness and Goodness Simpliciter', *Ratio* 29/4: 378–394.

Andersson, H. (2016b), 'Parity and Comparability, a Concern Regarding Chang's Chaining Argument', *Ethical Theory and Moral Practice* 19/1: 245–253.

Andersson, H. (2017), *How It All Relates: Exploring the Space of Value Comparisons*. Ph.D. dissertation, Lund University

Andersson, H. and Herlitz, A. (2018), 'A More Plausible Collapsing Principle', *Theoria* 84/4: 325–336.

Andersson, H. and Herlitz, A. (2021), 'Classifying Comparability Problems in a Way That Matters', *The Institute for Futures Studies. Working Paper* 2021:22: 239–262.

Andreou, C. (2016), 'Dynamic Choice', in E. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Center for the Study of Language and Information Stanford University).

Aristotle. (1984), Nicomachean Ethics, in J. Barnes (ed.), *The Complete Works of Aristotle: The Revised Oxford Translation* (Princeton University Press).

Bader, R. M. (2018), 'Stochastic Dominance and Opaque Sweetening', *Australasian Journal of Philosophy* 96/3: 498–507.

Bales, A., Cohen, D. and Handfield, T. (2014), 'Decision Theory for Agents with Incomplete Preferences', *Australasian Journal of Philosophy* 92/3: 453–470.

Berlin, I. (1969), 'Two Concepts of Liberty', in his *Four Essays on Liberty* (Oxford University Press).

Blackorby, C., Bossert, W. and Donaldson, D. (1996), 'Quasi-Orderings and Population Ethics', *Social Choice and Welfare* 13: 129–151.

Blackorby, C., Bossert, W. and Donaldson, D. (1997), 'Critical-Level Utilitarianism and the Population-Ethics Dilemma', *Economics & Philosophy* 13/2: 197–230.

Broome, J. (1997), 'Is Incommensurability Vagueness?', in R. Chang (ed.), *Incommensurability, Incomparability, and Practical Reason* (Harvard University Press).

Broome, J. (2004), *Weighing Lives* (Oxford University Press).

Broome, J. (2009), 'Reply to Rabinowicz', *Philosophical Issues* 19/1: 412–417.

Boyer, C. B. and Merzbach, U. C. (2011). *A History of Mathematics* (John Wiley & Sons).

Boot, M. (2009), 'Parity, Incomparability and Rationally Justified Choice', *Philosophical Studies* 146/1: 75–92.

Boot, M. (2017), *Incommensurability and Its Implications for Practical Reasoning, Ethics and Justice* (Rowman and Littlefield).

Carlson, E. (2004), 'Broome's Argument against Value Incomparability', *Utilitas* 16/2: 220–224.

Carlson, E. (2010), 'Parity Demystified', *Theoria* 76/2: 119–128.

Carlson, E. (2013) 'Vagueness, Incomparability, and the Collapsing Principle', *Ethical Theory and Moral Practice* 16/3: 449–463.

Chang, R. (1997), 'Introduction', in R. Chang (ed.), *Incommensurability, Incomparability, and Practical Reason* (Harvard University Press).

Chang, R. (2002), 'The Possibility of Parity', *Ethics* 112/4: 659–688.

Chang, R. (2013), 'Grounding Practical Normativity: Going Hybrid', *Philosophical Studies* 164/1: 163–187.

Chang, R. (2015), 'Value Incomparability and Incommensurability', in I. Hirose and J. Olsson (eds.), *The Oxford Handbook of Value Theory* (Oxford University Press).

Chang, R. (2016), 'Parity, Imprecise Comparability and the Repugnant Conclusion', *Theoria* 82/2, 182–214.

Davidson, D. (1973), 'On the Very Idea of a Conceptual Scheme', *Proceedings and Addresses of the American Philosophical Association* 47: 5–20.

De Sousa, R. (1974), 'The Good and the True', *Mind* 83/332: 534–551.

Doody, R. (2019), 'Parity, Prospects, and Predominance', *Philosophical Studies* 176/4: 1077–1095.

Elson, L., (2014a), 'Heaps and Chains: Is the Chaining Argument for Parity a Sorites?', *Ethics* 124/3: 557–571.

Elson, L., (2014b), 'Borderline Cases and the Collapsing Principle', *Utilitas* 26/1: 51–60.

Elson, L. (2017), 'Incommensurability as Vagueness: a Burden-Shifting Argument', *Theoria* 83/4: 341–363.

Feyerabend, P. (1978), *Science in a Free Society* (New Left Books).

Fishburn, P. C. (1973), *The Theory of Social Choice* (Princeton University Press).

Geach, P. T. (1956), 'Good and evil', *Analysis* 17/2: 33–42.

Gert, J. (2004), 'Value and Parity', *Ethics* 114/3: 492–510.

Griffin, J. (1986), *Well-Being: Its Meaning, Measurement and Importance* (Clarendon Press).

Griffin, J. (1997), 'Incommensurability: What's the Problem?', in R. Chang (ed.), *Incommensurability, Incomparability, and Practical Reason* (Harvard University Press) 35–51.

Gustafsson, J. E. (2013), 'Indeterminacy and the Small-Improvement Argument', *Utilitas* 25/4: 433–445.

Gustafsson, J. E. (2018), 'Does the Collapsing Principle Rule Out Borderline Cases?', *Utilitas* 30/4: 483–492.

Hare, C. (2010), 'Take the Sugar', *Analysis* 70/2: 237–247.

Hausman, D. M. (2015), *Valuing Health: Well-Being, Freedom, and Suffering* (Oxford University Press).

Herlitz, A. (2017), 'Comparativism and the Grounds for Person-Centered Care and Shared Decision Making', *Journal of Clinical Ethics* 2017/4: 269–278.

Herlitz, A. (2018), 'Committing to Priorities: Incompleteness in Macro-Level Health Care Allocation and Its Implications', *The Journal of Medicine and Philosophy* 43/6: 724–745.

Herlitz, A. (2019), 'Nondeterminacy, Two-Step Models, and Justified Choice', *Ethics* 129/2: 284–308.

Herlitz, A. (2020), 'Nondeterminacy, Cycles and Rational Choice', *Analysis* 80/3: 443–449.

Herlitz, A. and Sadek, K. (2021), 'Social Choice, Nondeterminacy, and Public Reasoning', *Res Philosophica* 98/3: 1–25.

Hsieh, N. H. (2005), 'Equality, Clumpiness and Incomparability' *Utilitas* 17/2: 180–204

Keynes, J. M. (1921), *A Treatise on Probability* (Macmillan and Company)

Klocksiem, J. (2010), 'In Defense of the Trichotomy Thesis', *Acta Analytica* 25/3: 317–327.

Kuhn, T. (1962), *The Structure of Scientific Revolutions* (University of Chicago Press).

Luce, R. D. (1958), 'A Probabilistic Theory of Utility', *Econometrica* 26/2: 193–224.

McClennen, E. F. (1990), *Rationality and Dynamic Choice: Foundational Explorations* (Cambridge University Press).

Norcross, A. (1997), 'Comparing Harms: Headaches and Human Lives', *Philosophy & Public Affairs* 26/2: 135–167.

Parfit, D. (1984), *Reasons and Persons* (Oxford University Press).

Parfit, D. (2016), 'Can We Avoid the Repugnant Conclusion?', *Theoria* 82/2: 110–127.

Qizilbash, M. (2007), 'The Mere Addition Paradox, Parity and Vagueness', *Philosophy and Phenomenological Research* 75/1: 129–151.

Rabinowicz, W. (1995), 'To Have One's Cake and Eat It, Too: Sequential Choice and Expected-Utility Violations', *The Journal of Philosophy* 92/11, 586–620.

Rabinowicz, W. (2008), 'Value Relations', *Theoria* 74/1: 18–49.

Rabinowicz, W. (2012), 'Value Relations Revisited', *Economics and Philosophy* 28/2, 133–164.

Rabinowicz, W. (2017), 'From Values to Probabilities', *Synthese* 194/10, 3901–3929.

Rabinowicz, W. (2020), 'Are Probabilities Values?', in I. Niiniluoto and S. Pihlström (eds), *Normativity—The 2019 Entretiens of Institut International de Philosophie, Acta Philosophical Fennica* 96: 67–99.

Rabinowicz, W. (2021), 'Can Parfit's Appeal to Incommensurabilities in Value Block the Continuum Argument for the Repugnant Conclusion?', in J. McMahan, T. Campbell, J. Goodrich, and K. Ramakrishnan (eds.), *Ethics and Existence: The Legacy of Derek Parfit* (Oxford: Oxford University Press).

Raz, J. (1986), *The Morality of Freedom* (Clarendon Press).

Regan, D. (1997), 'Value, Comparability, and Choice', in R. Chang (ed.), *Incommensurability, Incomparability, and Practical Reason* (Harvard University Press).

Sartre, J.-P. (1975), 'Existentialism Is a Humanism', in W. Kaufman, *Existentialism from Dostoevsky to Sartre* (Meridian-New American).

Savage, L. J. (1954), *The Foundations of Statistics* (Wiley).

Schoenfield, M. (2014), 'Decision Making in the Face of Parity', *Philosophical Perspectives* 28, 263–277.

Sen, A. (1970), *Collective Choice and Social Welfare* (Holden Day).

Sen, A. (1997), 'Maximization and the Act of Choice', *Econometrica* 65/4: 745–779.

Sidgwick, H., (1981), *The Methods of Ethics*, seventh edition, first published 1874 (Hackett Publishing Company).

Sinnott-Armstrong, W. (1985), 'Moral Dilemmas and Incomparability', *American Philosophical Quarterly* 22/4: 321–329.

Sinnott-Armstrong, W. (1988), *Moral Dilemmas* (Basil Blackwell).

Styron, W. (1979), *Sophie's Choice* (Random House).

Sunstein, C. (1994), 'Incommensurability and Valuation in Law', *Michigan Law Review* 92: 779–861.

Temkin, L. S. (2012), *Rethinking the Good: Moral Ideals and the Nature of Practical Reasoning* (Oxford University Press).

Thomson, J. J. (1997), 'The Right and the Good', *The Journal of Philosophy* 94/6: 273–298.

Von Fritz, K. (1970), 'The Discovery of Incommensurability by Hippasus of Metapontum', in D. Furley and R.E. Allen (eds.), *Studies in Presocratic Philosophy* (Routledge and Kegan Paul).

Voorhoeve, A. (2014), 'How Should We Aggregate Competing Claims?', *Ethics* 125/1, 64–87.

Wasserman, R. (2004), 'Indeterminacy, Ignorance and the Possibility of Parity', *Philosophical Perspectives* 18/1: 391–403.

Williams, B. (1981), *'Conflicts of Values', in His Moral Luck* (Cambridge University Press).

Part I

Accounts of Incommensurability

1 Incommensurateness Is Vagueness

John Broome

1 Introduction

When we have to make a decision, it often seems indeterminate what would be the best thing to do.[1] The alternatives seem "incommensurate" with each other, to use the term I favor. ("Incommensurable," "incomparable," "on a par," and other terms are often used instead.) The classic example is the choice faced by a student of Jean-Paul Sartre's in wartime France (Sartre, 2007). The student could choose to stay at home to look after his mother, who badly needed him, or he could try to escape to Britain and join the Free French Forces. Sartre took these alternatives to be incommensurate with each other. The aim of this chapter is to try to persuade you that incommensurateness like this is nothing more than vagueness.

We are interested in what is the best thing for the student to do, which is a matter of comparative value. In other words, it is a matter of betterness: what is better than what? Betterness is a comparative relation. When a monadic property *F* comes in degrees, it has a comparative *more F than* or *Fer than*, which is a dyadic relation. Betterness is the comparative of the monadic property of goodness.

Among comparative relations, betterness is especially important because of its connection with normativity – with what we ought to do. This chapter concentrates on it for this reason. However, betterness shares the common features of comparatives in general. Many other comparatives exhibit incommensurateness, for example. The analysis of incommensurateness is therefore not in itself particularly a topic within the philosophy of value; it is a topic concerned with the structure of comparatives in general. We can ask in general whether incommensurateness is vagueness.

That is where this chapter starts. The next three sections examine incommensurateness in general. Section 2 explains that there are two alternative sorts of it, which I call soft and hard incommensurateness, respectively. Soft incommensurateness is simply vagueness; hard incommensurateness is something else.

In Sections 3 and 4, I shall describe two formal considerations that argue in favor of soft incommensurateness, though not conclusively.

Then in Section 5, I turn to incommensurateness of betterness specifically. This is a particular bone of contention because of the connection between betterness and normativity. I describe a well-known normative puzzle that arises from the incommensurateness of betterness.

The rest of the chapter examines solutions to this puzzle. Sections 6, 7, and 8 describe solutions that can be provided on the basis of hard incommensurateness. Section 1.9 describes a solution that can be provided on the basis of soft incommensurateness. It argues that this solution is the best.

2 Hard and Soft Incommensurateness

I say two things are incommensurate in respect of a comparative *Fer than* when neither is definitely Fer than the other and also they are not definitely equally F.

As an example, I shall use the comparative relation *redder than*, which I call "redderness." Redderness is the comparative of the monadic property of redness. Figure 1.1 illustrates a part of the redderness relation as it holds among colors. Sadly, I have not been offered color printing in this volume, so I have to ask you to use your imagination. To keep things simple, I hold one color constant and examine how other colors are related to this one. My constant is the color A in the diagram, which is a reddish orange. I compare this with the range of colors B, which extend from pure red at the top to bluish purple at the bottom. Figure 1.1 indicates which colors in this range are redder than A.

The colors toward the top are definitely redder than A, and the colors toward the bottom are definitely not redder than A. In the middle are
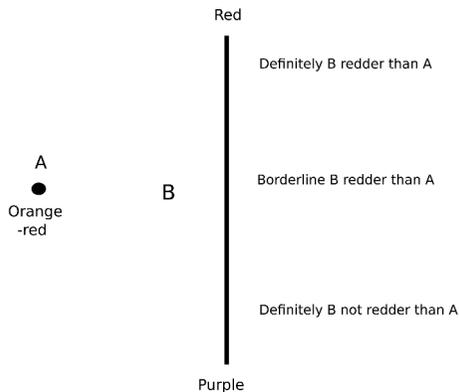

Figure 1.1 Colors redder than A.

colors that lie on the borderline between being redder than A and not being redder than A. The property of being redder than A is a monadic property and it is plainly vague. Some colors definitely possess it, and others definitely do not. In between are borderline colors, as the figure shows.

The vagueness of redderness is a sort of incommensurateness. Borderline colors are not definitely redder than A, nor definitely not redder than A, nor definitely equally as red as A. I call this "soft incommensurateness."

In this chapter, when I need an analysis of vagueness, I shall assume supervaluation (Fine, 1975). According to supervaluation, a vague property is a package of sharp properties called its "sharpenings." A proposition involving a vague property can be sharpened by replacing the vague property with one of its sharpenings. The original proposition is definitely true if and only if every one of its sharpenings is true.

In Figure 1.1, each color within the range B, if it is in the borderline between colors that are definitely redder than A and those that are definitely not redder than A, can represent a particular sharpening of the redderness relation. According to this sharpening, any higher color is redder than A and any lower color is not redder than A. By the same token, a color in the borderline also represents a sharpening of the monadic property *redder than A*. Any color that lies above the borderline is, therefore, redder than A according to every sharpening, so it is definitely redder than A. Correspondingly, any color that lies below the borderline is definitely not redder than A.

Figure 1.1 illustrates a part of the redderness relation, and Figure 1.2 illustrates a different part of it. The same constant color A is compared with the same range B, but the comparison is reversed. This diagram identifies which colors A is redder than. A is definitely redder than colors toward the bottom of the range. It is definitely not redder than colors toward the top of the range. In between are borderline colors.
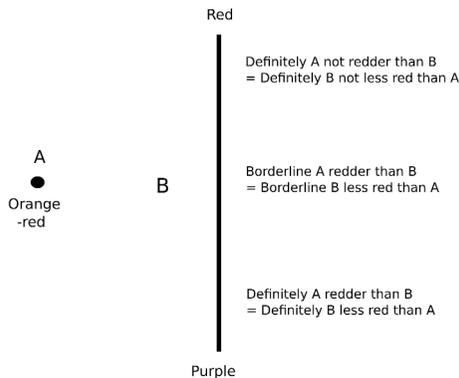
Figure 1.2 Colors A is redder than.

Figure 1.4 Overlapping vagueness.

Figure 1.5 Non-overlapping vagueness.

shall argue that it does not. This implies that all the incommensurateness in betterness is vagueness.

How can the question be settled? It may be that some comparatives have hard incommensurateness and others do not. There are some formal considerations to take into account that apply to all comparatives. I shall mention two. Neither of them is conclusive, but they constitute a *prima facie* case against hard incommensurateness in general.

3 The Collapsing Principle

The first consideration is this. If you look at the color B_2 in Figure 1.4 or Figure 1.5, you will see something puzzling about it. A is definitely not redder than B_2. On the other hand, it is borderline whether or not B_2

is redder than A. So the status of B_2's redness in comparison to A is not symmetrical: B_2 is better off than A with regard to redness. It is therefore puzzling why it is not definitely redder than A. You would think that possessing a higher status with regard to redness would be enough to make B_2 definitely redder. Yet according to those diagrams, this is not so.

This casts suspicion on the configurations shown in those two diagrams. This suspicion attaches to the formal configuration and has nothing to do with the particular example of redderness. It suggests that hard incommensurateness may not exist at all.

I once formulated this suspicion as something I called "the collapsing principle" (Broome, 1998). It may be expressed this way:

> If X is definitely not Fer than Y but Y is not definitely not Fer than X, then Y is definitely Fer than X.

I have now been convinced that the collapsing principle is not true in every case. Luke Elson has developed some convincing counterexamples to it.[2]

Here is one of them. Elson tells us that, when he plans a holiday, he prefers to visit a large country. It is not that he prefers one country to another if and only if it is larger. Among large countries, he has no preference on grounds of largeness, nor does he among countries that are not large. But he prefers any large country to any country that is not large. The relation that constitutes the counterexample is *better than as a holiday destination for Elson*, which I shall write as just *better than*. It is the comparative of the monadic property *good as a holiday destination for Elson* or *good*. A country is definitely good if and only if it is definitely large.

Elson tells us that China is definitely large, Ireland is definitely not large, and France is borderline large. Consequently, Ireland is definitely not better than France, but France is not definitely not better than Ireland. By the collapsing principle, therefore, France is definitely better than Ireland. On grounds I shall explain next, Elson argues that this can be so only if France is definitely good, which is to say definitely large.

By parallel reasoning, comparing France with China, we may also conclude that France is definitely not large, which implies it is not definitely large. The collapsing principle therefore implies a contradiction, so it is false. That is Elson's conclusion.

How does Elson derive "France is definitely good" from "France is definitely better than Ireland"?[3] We cannot, in general, derive "X is definitely F" from "X is definitely Fer than Y." For example, we cannot derive "Andorra is definitely large" from "Andorra is definitely larger than Monaco". So we need an argument.

The derivation is valid for sharp monadic properties. We can validly derive "5 is prime" from "5 is more prime than 4." But Elson is not interested in sharp properties because the collapsing principle is trivially true

of them. A sharp monadic property F has a sharp comparative Fer than. This means that when X is not definitely not Fer than Y it is automatically definitely Fer than Y.

Elson needs his monadic property to be vague, and indeed *good* is vague. France is borderline good because it is borderline large. But the collapsing principle implies France is definitely better than Ireland, and Elson claims this is inconsistent with France's being only borderline good. It has to be definitely good.

This can be demonstrated by supervaluation applied to the monadic property goodness. Every sharpening of goodness identifies a particular size as the boundary between countries that are good and those that are not good. According to any sharpening, all the good countries are equally as good as each other, and all the not good countries are also equally as good as each other. That is the way goodness as a holiday destination for Elson works.

Now suppose France was only borderline good. Then there would be at least one sharpening of goodness in which France is not good. But in this sharpening, France would be equally as good as Ireland, which is also not good. Consequently, in this sharpening, France would not be better than Ireland. Since there would be at least one sharpening in which France is not better than Ireland, France would not be definitely better than Ireland as the collapsing principle implies. So the collapsing principle implies that France is definitely good.

I think Elson's is a sound counterexample to the collapsing principle. However, it works only because of an exceptional feature of the property in question. Both the monadic property of goodness and its comparative betterness are vague. Each has a number of sharpenings. The special feature is that in every sharpening of both, no country that is not good is better than any other country that is not good. Elson's highly contrived property is designed to have this feature. I know of no more natural properties that do.

Moreover, although Elson's is a counterexample to the collapsing principle, it is not an example of hard incommensurateness. True, the borderlines of the two vague properties *better than France* and *worse than France* do not overlap. But between them lie only countries that are the same size as France, and these are all definitely equally as good as France. Any countries that are definitely not better than France and definitely not worse than France are definitely equally as good as France.

Erik Carlson (2013) and subsequently Andersson and Herlitz (2018) have pointed out that this weaker version of the collapsing principle is immune to Elson's example:

> If X is definitely not Fer than Y but Y is not definitely not Fer than X, then Y is definitely either Fer than X or equally as good as X.

Furthermore, these authors show that this weak principle, like the stronger version, implies there is no hard incommensurateness. But I see no independent attraction in it. If the collapsing principle itself is refuted, there are no grounds for clinging on to this weak version of it.[4]

I conclude that Elson has shown that the collapsing principle is not universally true. Moreover, other counterexamples have been described in the literature (e. g. Carlson, 2004 and Gustafsson, 2018). They are not as watertight as Elson's but are nevertheless effective. The collapsing principle is subject to severe doubt, therefore. But it remains intuitively attractive and may well be true for common cases. Where it is true, it implies that matching vagueness is the correct account of incommensurateness. So it provides some evidence in favor of matching vagueness. *A fortiori*, it provides some evidence against hard incommensurateness.

4 Gradation

I shall say no more about the collapsing principle. This chapter concentrates on the second formal consideration that opposes hard incommensurateness. It is this. Intuitively, incommensurateness has a graded structure. As we move through the range B from bottom to top, we move from colors that are definitely less red than A to ones that are definitely redder than A. In between is a zone of colors that are neither. In moving up through this incommensurate zone, we of course come to colors that are progressively redder. Intuitively, it is also true that their comparative redderness in comparison to A progressively increases. It is easy to make sense of this intuition of gradation in terms of vagueness. But if there is hard incommensurateness, it is not so easy.

Vagueness supplies gradation through this principle:

> *Greatervaluation*. X is Fer than Y if X is F according to every sharpening of F in which Y is F, and X is F according to some sharpening of F in which Y is not F.

In Broome (1998: 83), I offered greatervaluation as a general truth. However, Henrik Andersson and Ruth Chang have shown me it is not one. Whereas vagueness can induce gradation on a property through greatervaluation, many properties have a prior gradation on some different basis. For instance, there are degrees of redness even among colors that are definitely red, and this gradation cannot derive from the vagueness of *red*. For some properties, this prior grading on a different basis can override grading on the basis of vagueness. But I am now using greatervaluation only in order to explain how grading can be based on vagueness.

Let F be the property *redder than A*. Take two colors B_2 and B_3 in the borderline of this property, where B_2 is above B_3. Each sharpening of *redder than A* is represented by a color within the borderline. According to

all sharpenings represented by colors below B_3, both B_2 and B_3 are redder than A. According to all sharpenings represented by colors between B_3 and B_2, B_2 is redder than A, but B_3 is not. So by greatervaluation, B_2 is redder than A more than B_3 is redder than A. This supports the intuition of gradation within this borderline.

Next, let F be the property *less red than A*. Take two colors B_4 and B_5 in the borderline of this property, where B_4 is above B_5. Each sharpening of *less red than A* is represented by a color within the borderline. According to all sharpenings represented by a color above B_4, both B_4 and B_5 are less red than A. According to all sharpenings represented by colors between B_4 and B_5, B_5 is less red than A, but B_4 is not. So by greatervaluation, B_5 is less red than A more than B_4 is less red than A. This supports the intuition of gradation within this borderline.

So there is increasing redderness in the borderlines. But in the zone of hard incommensurateness in Figure 1.5, which lies between the two borderlines, redderness apparently does not increase as we move up through the zone. Any color in this zone is definitely not redder than A and definitely not less red than A, and this is equally so for any color in the zone. Hard incommensurateness seems to create a hiatus in the gradation of redderness, between the zones of vagueness in which redderness is graded.

At least, in the zone of hard incommensurateness, no gradation results from greatervaluation. A gradation could be supplied by other means but demonstrating gradation would require more assumptions and more theoretical apparatus. This is further inconclusive evidence against hard incommensurateness.

5 A Puzzle about Betterness

I turn now from redderness to the more important topic of betterness. The example of Sartre's student is illustrated in Figure 1.6. The student has a choice between A, staying in France with his mother or B, leaving for Britain to join the Free French Forces. Is either better than the other? This depends on the details of each. For example, it depends on how likely it is that he will get to Britain if he tries and on how much his mother needs him. The question is illustrated in Figure 1.6. For the sake of analysis, I shall once again hold one of the options constant. For this role, I have chosen the option of staying in France, which I assume will be in the conditions described by Sartre. B is a range of different versions of the option of leaving. High up are cases where it is easy to get to Britain, and the student's contribution to liberating France is likely to be great. For these, I assume that leaving is definitely better than staying. Low down in the range are cases where travel is very dangerous, and he is likely not to succeed. Staying is definitely better than leaving in one of these cases. Since the considerations that favor either option are very different from each other, it is plausible that there is a range of cases in

B: leave

B₁   Definitely better than A

B₃

A: stay   Incommensurate zone

●

B₄

B₂   Definitely worse than A
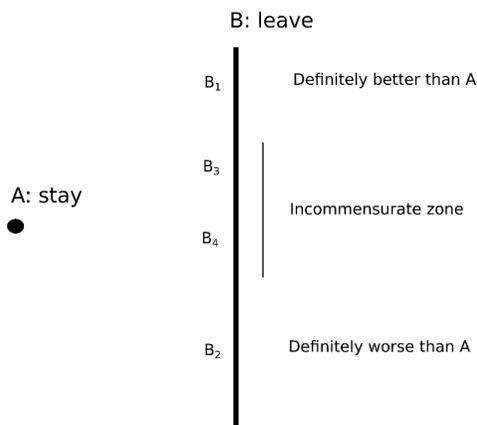
Figure 1.6 Sartre's student.

between where staying is not definitely better than leaving, and leaving is not definitely better than staying. These are incommensurate cases.

Betterness, like redderness, is undoubtedly vague. One possibility is that all the incommensurateness is soft, which is to say that it consists entirely of vagueness. In that case, the whole incommensurate zone consists of the borderlines of the properties *better than A* and *worse than A*, which overlap to some extent. The two borderlines may even coincide exactly; this is the case of matching vagueness.

It is also possible that within the range of B, there is a zone where the incommensurateness is hard. In a zone of hard incommensurateness, staying is definitely not better than leaving, and leaving is definitely not better than staying, and staying and leaving are definitely not equally good.

Which is the correct account of the situation? Is there hard incommensurateness or not? This is the same question for betterness as for any other comparative. The same formal considerations arise as the ones I described for redderness, stemming from the collapsing principle and from gradation. Both provide some evidence against hard incommensurateness.

The question assumes special importance for betterness because of the connection between betterness and normativity. This connection also provides a different perspective on the question, besides the formal considerations. We can ask what sort of incommensurateness provides the best account of normative phenomena that need explaining. Since, of course, betterness is vague, we automatically have soft incommensurateness as a resource for explaining the phenomena. We can ask whether we also need hard incommensurateness as a further resource.

I shall argue we do not. It has been recognized for a long time that incommensurateness throws up a particular normative puzzle (see Chang, 1998: 11). I shall argue that, not only do we not need hard incommensurateness

to solve it, but soft incommensurateness provides a better solution to it than hard incommensurateness does. It will turn out in Section 9 that this is because of one of the formal considerations: a good solution requires betterness to be graded.

As a preliminary to describing the puzzle, I need to specify one small part of the connection between betterness and normativity. Betterness affects what you ought to do. In many cases, it does not fully determine what you ought to do, but in some cases it does. Take a case where it does and suppose that in such a case a person has a choice between just two options. Then she ought to choose one of them if and only if it is better than the other. This is a minimal assumption about the connection.

As I am going to use "permissible," it is permissible for you to choose an option if and only if it is not the case that you ought not to choose it. Given there are only two options, you ought not to choose one if and only if you ought to choose the other. So the minimal assumption implies that it is permissible for you to choose an option if and only if the other is not better than it.

Let us assume the example of the student is a case where betterness fully determines what you ought to do. To justify this assumption, we shall have to recognize a wide range of values, perhaps including national honor and the performance of filial duties. Let us recognize all the goods we need to and incorporate all of them into betterness. So, for instance, if the student has a choice between A and B_1 in Figure 1.6, he ought to choose B_1, and if he has a choice between A and B_2, he ought to choose A.

Now concentrate on two options B_3 and B_4 that are both incommensurate with A. If there is a zone of hard incommensurateness, assume they are both in that zone. If there is no hard incommensurateness, they are both softly incommensurate with A.

These two options can illustrate the normative puzzle I mentioned. Suppose the student is offered two choices on successive days. Today, he is offered a choice between A, staying in France, and B_3, which is travel to Britain by a reasonably safe means. He chooses A. The next day he is offered a choice between A and B_4. By then the safer means of travel has shut down, and B_4 involves less safe travel. This time the student chooses B_4. On the face of it, there seems to be nothing wrong with either of his two choices since he does not choose a worse option out of those he is offered. Yet he ends up with B_4 when he could have had B_3, and B_4 is definitely worse than B_3. Had he been offered all three options at once, he definitely ought not to have chosen B_4. This is the puzzle. If the student makes the choices I have described, does he do something impermissible, and if so, what? It seems he should have some normative protection against this bad outcome. What can it be?

What solution to this puzzle can be offered by different accounts of incommensurateness?

6 Hard Incommensurateness: The First Solution

Let us first assume the incommensurateness is hard and see what follows. In that case, definitely, neither A nor B_3 is better than the other. Suppose the student has a choice between these two options only. My minimal assumption about the connection between betterness and normativity implies it is definitely permissible for him to choose A and definitely permissible for him to choose B_3. Similarly, in a choice between A and B_4, it is definitely permissible for him to choose A and definitely permissible for him to choose B_4. So both the choices he makes are definitely permissible, yet a bad outcome results. That is the puzzle.

Various solutions are available given the assumption of hard incommensurateness. A solution I once offered myself (Broome, 2001) is to say that actually, the student does nothing wrong. There is no normative protection against the bad result. However, causal processes may well protect him from it. When he chooses A on the first day, he forms an intention to stay with his mother. A person's intentions typically persist until they are fulfilled, and they typically cause the person to take means toward their fulfillment. On the second day, the student can be expected still to have the intention he formed on the first day. A means toward its fulfillment is to reject the option B_4. If he does that, he avoids the bad outcome. It is prevented just by the causal tendency of intentions to persist. Sadly, this protection against the bad outcome is weak because the student might easily change his mind on the second day and give up his intention.

More recently (Broome, 2013: Section 10.1) I came to think that the causal tendency for an intention to persist has some rational support. It is not just that intentions typically do persist; under certain conditions, rationality requires them to persist. Among the conditions is that the person does not reconsider the intention. I claim it is not rational just to give up an intention you have without at least thinking about it. If the student thoughtlessly accepts the offer of B_4 on the second day, that would be irrational. He would go wrong at that point. Nevertheless, if he does think about it, it would not be irrational for him to accept the offer of B_4 since it is a permissible choice for him to make. So, even reinforced with this rational support, causal protection against the bad outcome is weak.

The bottom line is that the student may do nothing wrong even though he comes to a bad outcome.

7 Hard Incommensurateness: The Second Solution

A second solution can be developed from an idea of Ruth Chang's (2017, 2021). I do not believe Chang herself has explicitly offered this idea as a solution to this particular puzzle, but it is easily turned to that purpose. Chang thinks that a person can create reasons for herself by making commitments. For example, the student might commit himself to his mother,

and this commitment gives him a new reason to stay and look after her. Chang deals in reasons, but the point can be made for values too. A person can create values for herself by making commitments. For example, by committing himself to his mother, the student can make looking after her a particular value for him.

Moreover, Chang thinks that a "hard choice," such as the one the student makes, creates a special opportunity for making a commitment of this sort. When considering his choice on the first day, wondering whether to stay at home or leave for Britain, the student might decide to commit himself to his mother. If so, this makes the option of staying at home more valuable to him than it previously was. Whereas the option A of staying at home was previously not definitely better than B_3, leaving for Britain, it may now become so. If it does, A is then also definitely better than B_4. The value created by the commitment has consequences that extend beyond the particular choice where it was made. The student ought now to choose A on the second day too. If he chooses B_4, he goes wrong. That is a solution to the puzzle.

However, it is only a partial solution because, whatever values a person creates by making commitments, she may still face an incommensurate choice. Self-made values have to be weighed against each other and against other values, and they may easily be incommensurate with them. It would take a total commitment to a single value to make this impossible, and that would at best provide a very rare and extreme solution to the puzzle. So now I want you to read the student example as having already taken into account all the student's self-made values, including all those he makes when considering his first choice. Assume that these commitments leave the student with a choice between incommensurate options, as we have been assuming up to now. In this case, we still have a puzzle. The commitments do not solve it.

Chang recognizes this possibility. She recognizes that a person faced with a "hard choice," as she calls it, may not settle the choice by making a commitment. Instead, she may drift into a decision (Chang, 2017: 18–19). I take Chang's word "drifting" to mean making the choice in some way that does not involve a commitment that projects a value beyond the particular choice. In choosing A over B_3 on the first day, the student is not necessarily making a commitment. So it does not follow that, on the second day, he ought not to choose B_4 over A. The student may do nothing wrong at all.

We could deny that drifting is possible. We could claim that a person could not make a decision – or perhaps could not rationally make a decision – between two incommensurate options without making them commensurate by means of a commitment. But that would be entirely implausible. In order to make a rational choice between two options, you do not have to think one of them is better than the other. If you happen not to think so, rationality does not paralyze you. Buridan's ass was

paralyzed in such a case and died as a result, but that was a failure of its rationality rather than a consequence of it. True, rationality requires you to choose to do whatever you believe you ought to do. But it does not require you not to choose to do anything that you do not believe you ought to do. Chang is right not to make this claim.

Another idea is that the decision between options itself creates a value so that any decision between options is itself a commitment of Chang's sort. But that is false, and Chang (2013) rejects it. She thinks that you can in effect decide you ought to do a particular act by making a commitment. Nevertheless, she recognizes that deciding to do something and deciding you ought to do it are different acts.

The bottom line is that the student may do nothing wrong, even though he arrives at a bad outcome.

8 Hard Incommensurateness: The Third Solution

Wlodek Rabinowicz (2012: 145) has a different way of solving the puzzle. He deploys a version of the fitting attitude account of value. This is a metaphysical theory about the nature of goodness. For something to be good is for it to be fitting to have a positive attitude toward it. Rabinowicz extends this account to betterness by saying that for one thing to be better than another is for it to be fitting to prefer the first to the second.

When neither of two things is better than the other and nor are they equally good, in typical cases Rabinowicz thinks it permissible to prefer one to the other and permissible to prefer the other to the one. He would think it permissible for the student to prefer A to B_3 and permissible for him to have the opposite preference, and also permissible for the student to prefer A to B_4 and permissible for him to have the opposite preference. But he must prefer B_3 to B_4 since B_3 is better. So it is not permissible for the student to prefer A to B_3 and also prefer B_4 to A, even though both these preferences are individually permissible. This is because a permissible preference relation must not be cyclic, and these two preferences would form a cycle with the student's preference for B_3 over B_4.

Let us assume the student should make choices in accordance with his preferences. So if he chooses A rather than B_3 and also B_4 rather than A, he does something impermissible. Either he does not choose in accordance with his preferences, or he prefers A to B_3 and also prefers B_4 to A, which is impermissible. Rabinowicz's solution to the puzzle is to say that the student does indeed go wrong, and this is where.

This solution depends on assuming that permissible preferences must not be cyclic. Why should that be? Various answers are available, but none yields a very satisfactory solution to the solution to the puzzle.

One is to make a normative connection between permissible preferences and betterness. Since betterness is acyclic, we might be able to derive

the conclusion that permissible preferences are acyclic. This answer does not suit Rabinowicz because he defines betterness in terms of permissible preferences, which he takes to be metaphysically prior to betterness. For him, if betterness is acyclic, that would have to be because permissible preferences are acyclic, rather than the other way around.

In any case, this answer does not help with the puzzle of the student. We might make the connection between permissible preference and betterness in either of two ways but neither succeeds. We might say it is permissible to prefer X to Y if and only if X is better than Y. Then indeed permissible preferences will be acyclic. But we are assuming that B_3 is neither better nor worse than A, so it will be impermissible for the student to prefer B_3 to A and impermissible for him to prefer A to B_3. Likewise, it will be impermissible for the student to prefer B_4 to A and impermissible for him to prefer A to B_4. Since he has to make choices, they cannot be constrained by his preferences, so we cannot use preferences to explain where he goes wrong.

Alternatively, we might say it is permissible to prefer X to Y unless Y is better than X, and then it is impermissible to prefer X to Y. But then permissible preferences may be cyclic. It is permissible for the student to prefer A to B_3, B_4 to A, and B_3 to B_4. So this first explanation of acyclicity leaves the puzzle unanswered.

A second explanation is the "money-pump" argument. Rabinowicz subscribes to this argument. It has been much debated, and Rabinowicz's own version of it is designed to overcome objections that have been raised against it (Gustafsson & Rabinowicz, 2020). But here I shall not question the validity of the argument; I am concerned only with its broad structure, so I shall describe only a simple version of it. Suppose you make choices in accordance with your preferences, and your preferences are cyclic. Suppose you prefer Y to X, Z to Y, and X to Z. Imagine you start off with X and are then offered the chance of swapping from X to Y. Because you prefer Y to X, you accept this offer and take Y. Next suppose you are offered the chance of swapping from Y to Z. You accept this offer too and end up with Z. But you prefer X, where you started, to Z, where you ended up. So you have arrived at a bad outcome. A course of action that leads to this outcome cannot be entirely permissible. It is presumably permissible to make choices in accordance with your preferences. So it must be your cyclic preferences that are not permissible.

This argument rests on the bad outcome of getting Z when you could have had X. This is just the sort of bad outcome that Sartre's student comes to in my version of his story. We are pursuing the question of where the student goes wrong. Rabinowicz's answer is that his choices imply he has cyclic preferences, and he goes wrong because cyclic preferences are impermissible. But now it turns out that cyclic preferences are impermissible just because they can lead to a bad outcome of this sort. So

we have not found a real explanation of where the student goes wrong. True, we have learned that his problem belongs to a recognized set of problems. But for a real explanation, we need an independent explanation of why cyclic preferences are impermissible. This money-pump explanation is incomplete.

Rabinowicz (2012: 152) does offer an independent explanation. It works by treating a person's preference between two things as a relation between the person's favoring attitudes toward the two individual things. For the person to prefer X to Y is for her favoring of X to be greater than her favoring of Y. The *greater than* relation is necessarily acyclic. Therefore, preferences understood this way are necessarily acyclic. It follows that permissible preferences are necessarily acyclic.

However, because it implies that the student's preferences are necessarily acyclic, this explanation takes us off course in solving his puzzle. The student prefers B_3 to B_4; this is what makes it better. Because his preferences are necessarily acyclic, he, therefore, cannot prefer A to B_3 and also B_4 to A. Yet he chooses A over B_3 and B_4 over A. How come? One possibility is that his choices are not in accordance with his preferences, which is to say not in accordance with his favorings. This is certainly a way to explain how he goes wrong, but it is not the sort of explanation we need. If, in making a choice, you choose the option you favor less, that must be some sort of accident. We need a better explanation of the student's puzzle than to say he chooses wrongly by accident. We can presume the student is careful and deliberate in his choices.

The only alternative is to suppose he changes his preference between one day and the next.[5] That is to say, he changes his favorings. If we are to conclude he goes wrong in some way, we shall have to say that this change of mind is impermissible. This is a matter of the persistence of an attitude. I considered the persistence of an intention in Section 6, and I take the same view about the persistence of a preference. Firstly, attitudes including preference have a causal tendency to persist, and this gives the student some causal protection against the bad outcome. Furthermore, it is not generally rational to change a preference without thinking about it, so there is some rational protection against the bad result. But both the causal protection and the rational protection are very weak: the student might rationally reconsider his preference and change it. We therefore cannot conclude that his change of mind is impermissible. This alternative, then, takes us back to something like my own first solution, which is weak. The student may do nothing wrong.

The bottom line is that it seems the student may do nothing wrong even though he arrives at a bad outcome. However, there might yet be a different explanation of why permissible preferences are acyclic. If so, this third solution might show that the student does indeed do something wrong.

9 Soft Incommensurateness: The Best Solution

We have been assuming that the student faces hard incommensurateness. On that assumption, we have not been able to conclude that the student necessarily does something wrong, even though he arrives as a bad outcome.

The underlying reason for this failure is that hard incommensurateness does not provide any grading of betterness. So far as their betterness relative to A is concerned, B_3 and B_4 have exactly the same standing, even though B_3 is better than B_4. The choice between B_3 and A is entirely neutral, there is no normative inclination either way. The same is true of the choice between B_4 and A. There is no more normative inclination toward A in the second choice than in the first. The result is that the two choices may go in the way that leads to the bad outcome. But intuitively, there ought to be more normative inclination in the second choice because B_4 is worse than B_3.

Soft incommensurateness – vagueness – provides this greater inclination because it grades betterness. So now I turn to soft incommensurateness. Suppose now that there is no hard incommensurateness between A and the range of options B. B_3 and B_4 are in a zone of vagueness rather than hard incommensurateness. Then, although neither of them is definitely better than A or definitely worse than A, B_3 is more better than A or less worse than A than B_4 is.

With vagueness, we have degrees of permissibility. If the student has a choice between A and some version of B within the incommensurate zone, it is not definitely permissible for the student to choose either. Moreover, the better is B within this zone, the more permissible it is for the student to choose B, and the less permissible it is for him to choose A. So the student should in some way be more inclined to choose A over B_4 than he is to choose A over B_3. This is as it intuitively should be.

For more detail, remember that there are actually two zones of vagueness: the vagueness of *better than A* and the vagueness of *worse than A*. If the vagueness is matching, as I defined it in Section 2, then either choice within the incommensurate zone is only borderline permissible. Neither is definitely permissible. Where the vagueness is overlapping, either choice within the overlap is borderline permissible. Outside the overlap, one choice or the other is definitely permissible, but for no option in the range B, it is definitely permissible to choose A and definitely permissible to choose B. So there is subtlety in the permissibility of the student's choices.

Moreover, supervaluation applied to vague betterness provides a clear explanation of why the student goes wrong when he makes the choices I described. My minimal assumption about the connection between betterness and normativity tells us the student ought not to choose an option over a better one. According to every sharpening of the betterness

relation, either B_3 is better than A or else A is better than B_4. So according to every sharpening, either the student ought not to choose A over B_3 or else he ought not to choose B_4 over A. (This is a disjunction of oughts; it is not an ought governing a disjunction.) By supervaluation, therefore, definitely, either he ought not to choose A over B_3 or else he ought not to choose B_4 over A. He does both, so one of the choices he makes is impermissible. That is why he goes wrong. In this way, we can clearly conclude that the student necessarily does something wrong.

You might ask which of his two decisions is the impermissible one, but you will normally receive no answer to that question. It is well-known (see Fine, 1975) that supervaluation can imply that a disjunction is definitely true without either of the disjuncts being so. However, because permissibility is graded, the impermissibility can be distributed between the two choices. Neither is definitely impermissible, but one may be less permissible than the other.

You might also ask what is the normative position after the student makes his first choice of A over B_3. If he then chooses B_4 over A, he will definitely have done something impermissible. He will have acted impermissibly, either in his first choice or in his second. Can we, therefore, conclude that he ought not to choose B_4 over A since doing so will ensure that he does something he ought not to do? No, we cannot.

An argument could be deployed to say we can. Here is one argument made by Luke Elson (unpublished). Elson starts from the premise that the student ought either not choose A over B_3 or not choose B_4 over A. This is not the premise I have already stated. It is a wide-scope ought governing a disjunction. It does not follow from my minimal assumption about the connection between betterness and normativity. But it nevertheless seems plausible, and I shall accept it for the sake of argument. It may be reformulated to say that the student ought, if he chooses A over B_3, not to choose B_4 over A, where the clause "if he chooses A over B_3" is within the scope of "ought." This formulation is arrived at by substituting logical equivalents with the scope of "ought" – a rule of deontic logic that is hard to reject (see Broome, 2013: 122).

Now suppose the student chooses A over B_3 and his choice is irrevocable. This choice is then necessary in a temporal sense; it cannot be altered. Another rule of deontic logic that is hard to reject is "necessary detachment" as I call it (Broome, 2013: 123–125). It says that from "ought if X then Y" and "necessarily X," it follows that "ought Y." So it seems we can detach the conclusion that the student ought not to choose B_4 over A. That is Elson's conclusion.

But this argument has a problem. Elson assumes that the premise remains true even after the student has made his first choice, but normative truths can alter with the circumstances. This premise is true so long as the student's first choice is open, but once that choice is made and irrevocable, it may no longer be true.

Think about this case. For a moment, drop the assumption that B_3 and B_4 are in the incommensurate zone. Assume instead that they are both definitely better than A, though B_3 remains better than B_4. Presumably, if it is true in the original case that the student ought either not choose A over B_3 or not choose B_4 over A, it is true in this case too. There is no reason why it should not be. Now suppose the student irrevocably chooses A over B_3, which he ought not to do. Can we conclude by necessary detachment that he ought not to choose B_4 over A? Of course not. B_4 is definitely better than A and it is definitely not the case that he ought not to choose it. Indeed he ought to choose it. He went wrong in his first choice, and he ought to correct it as well as he can. Necessary detachment is hard to reject, so this example shows that the premise is no longer true once the student has made a choice. We cannot assume the premise remains true once a choice has been made. Because Elson's argument makes this assumption, it is invalid.

Revert now to the case where both B_3 and B_4 are in the incommensurate zone and let us look at the student's situation in more detail. Suppose for the sake of argument that B_3 is near the top of this zone. Then, when the student faces his first choice between A and B_3, it is not definitely permissible for him to choose A. It is borderline impermissible; indeed it may be close to being definitely impermissible. We are assuming that, all the same, the student chooses A. When he comes the next day to choose between A and B_4, he may regret his previous choice of staying in France. He may think of choosing B_4 – leaving for Britain – as a way to reverse the choice he now regrets, albeit at a cost. When he makes the second choice of B_4, he ends up having done something impermissible, but the impermissibility may be located largely in his previous choice rather than his second one. By the time he comes to his second choice, he can do nothing about the previous one.

Intuitively, this way of thinking and choosing seems perfectly permissible. I find it intuitively incorrect to conclude that, once the student has chosen A over B_3, he ought not to choose B_4 over A. When you are facing a difficult choice between incommensurate alternatives, the fact that you have previously made a particular choice does not necessarily place a normative constraint on what you should choose this time.

In sum, I think that the implication of soft incommensurateness for the puzzle of the student is exactly right. Here, soft incommensurateness gives a better account of normativity than hard incommensurateness does.

10 Conclusion

Soft incommensurateness supplies a better account of the puzzle of the student than hard incommensurateness does. I know of no normative phenomenon that hard incommensurateness can explain better than

soft incommensurateness. For betterness at least, we should give up on hard incommensurateness, and recognize that incommensurateness is vagueness.

Notes

1 I am very grateful to Luke Elson for his comments on this chapter, and for a conversation on these topics that has continued for several years. In particular, this chapter has been influenced by his "Unsharpness and forbidden compound actions". I have also received extremely valuable comments from Ruth Chang, Wlodek Rabinowicz, Henrik Andersson, and Anders Herlitz, which have saved me from significant mistakes. Research for this chapter was supported by ARC Discovery Grant DP180100355.
2 Elson (2014). Although he rejects the collapsing principle, Elson nevertheless agrees with my conclusion in this chapter that incommensurateness is vagueness. See Elson (2017).
3 Johan Gustafsson (2018: 289) apparently assumes Elson has no good answer to this question. He therefore denies the validity of Elson's counterexample to the collapsing principle. He misses the argument from supervaluation that follows.
4 Here I agree with Carlson (2013: 456–457). Andersson and Herlitz (2018: 331–332) take the opposite view, but I think their argument is mistaken.
5 Thanks here to Luke Elson.

References

Andersson, H. and Herlitz, A. (2018), 'A more plausible collapsing principle', *Theoria* 84: 325–336.
Broome, J. (2001), 'Are Intentions Reasons? And How Should We Cope with Incommensurable Values?', in C. Morris and A. Ripstein (eds), *Practical Rationality and Preference: Essays for David Gauthier* (Cambridge University Press) 98–120.
Broome, J. (1998), 'Is Incommensurability Vagueness?', in R. Chang (ed.), *Incommensurability, Incomparability, and Practical Reason* (Harvard University Press), 67–89. Reprinted in Broome, J. (1999), *Ethics Out of Economics*, (Cambridge University Press).
Broome, J. (2013), *Rationality Through Reasoning* (Wiley–Blackwell).
Carlson, E. (2004), 'Broome's Argument against Value Incomparability', *Utilitas* 16: 220–224.
Carlson, E. (2013), 'Vagueness, Incomparability, and the Collapsing Principle', *Ethical Theory and Moral Practice* 16: 449–463.
Chang, R, (2013), 'Commitments, Reasons, and the Will', *Oxford Studies in Metaethics* 8: 74–113.
Chang, R. (2017), 'Hard Choices', *Journal of the American Philosophical Association* 92: 586–620.
Chang, R. (1998), 'Introduction', in R. Chang (ed.), *Incommensurability, Incomparability and Practical Reason* (Harvard University Press) 1–34.
Chang, R. (2021), 'What Is It to Be a Rational Agent?', in R. Chang and K. Sylvan (eds), *The Routledge Handbook of Practical Reason* (Routledge).

Elson, L. (2014), 'Borderline Cases and the Collapsing Principle', *Utilitas* 26: 51–60.

Elson, L. (2017), 'Incommensurability as Vagueness: A Burden-Shifting Argument', *Theoria* 83: 341–363.

Elson, L. (unpublished), 'Unsharpness and Forbidden Compound Actions'.

Fine, K. (1975), 'Vagueness, Truth and Logic', *Synthese* 30: 265–300.

Gustafsson, J. and Rabinowicz, W. (2020), 'A Simpler, More Compelling Money Pump with Foresight', *Journal of Philosophy* 117: 578–589.

Gustafsson, J. (2018), 'Does the Collapsing Principle Rule Out Borderline Cases?', *Utilitas* 30: 483–492.

Rabinowicz, W. (2012), 'Value Relations Revisited', *Economics and Philosophy* 28: 133–164.

Sartre, J.P. (2007), *Existentialism is a Humanism* (Yale University Press), (originally 1946).

2 Are Hard Cases Vague Cases?

Ruth Chang[1]

In a hard case of comparison between two items, it seems that neither is better or worse than the other and yet nor are they equally good. If you are comparing careers in investment banking and interior design, you might judge that the banking career is better in some relevant respects, the design career better in other relevant respects, and yet neither is at least as good as the other overall. Or if you are comparing the evidence for believing that there is a God and the evidence for disbelieving (or withholding judgment), you might judge that the case for belief is better in some respects, worse in others, and yet the evidence for each attitude is not at least as warrant-providing or justifying as the other. As these cases illustrate, hard cases occur in both the practical and theoretical domains. In their simplest form, they are normative comparisons of items, A and B, with respect to a "covering consideration," "V," such as "goodness as a career" or "epistemic warrant," in which it seems that none of the usual trichotomy of relations, "better than," "worse than," and "equal to," hold.

How should we understand such cases? There are four familiar "i" explanations: hard cases are hard (i) because we are ignorant of relevant normative or nonnormative facts; (ii) because the items are incommensurable – i.e., their normative merits are not cardinally measurable; (iii) because the items are incomparable; or (iv) because the relevant concepts or facts are indeterminate or vague. I have previously argued that these four explanations, whether taken jointly or severally, are inadequate explanations of hard cases, that is, when taken together or considered individually, they fail as an adequate explanation of all hard cases (Chang 1997, 2002a, 2012, 2017, forthcoming).

Although I have argued against indeterminacy as an explanation of hard cases, my arguments have been less directed since it has always seemed to me that indeterminacy is the least plausible way of explaining hard cases. In holding this view, however, I am in disagreement with a large number of philosophers (Griffin (1986); Broome (1997, this volume); Wasserman (2004); Qizilbash (2005, 2007, 2014); Sugden (2009); Klocksiem (2010); Constantinescu (2012, 2016); Elson (2014), Andersson (2015); Williams

(2014, 2016); Dougherty (2016); and Thomas (forthcoming)). Indeed, most of those writing in the area suggest that the right way to think about hard cases is in terms of indeterminacy, and in terms of vagueness in particular.[2]

In this chapter, I provide further explanation of why I think understanding hard cases in terms of vagueness is a mistake. Vagueness both *under-delivers* in that it fails to explain a critical feature of hard cases and *over-delivers* in that it imposes a feature that hard cases do not – and should not – have. I then offer my own Goldilocks explanation of hard cases, with specific attention to how it solves the two difficulties that plague vagueness accounts. I end by offering a nudge to proponents of vagueness; given their insights about hard cases, the distance they need to travel to accept my explanation of hard cases – and thereby avoid the two problems with their account – is not as great as they may think.

Two points of clarification. First, my focus will be on a subclass of hard cases, what we might call *paradigmatic* hard cases – those in which the items being compared are qualitatively different with respect to the covering consideration, and it seems that neither is better than the other or that they are equally good: qualitatively different careers, say, one in investment banking and the other in interior design with respect to goodness as a career; qualitatively different paintings, say, an abstract impressionist painting vs. a photograph with respect to aesthetic merit; qualitatively different types of theory, say, evolutionary theory vs. quantum mechanics with respect to explanatory power; qualitatively different types of evidence, say, statistical vs. forensic with respect to epistemic warrant – where neither seems at least as good. Hard cases involving qualitatively different items are rife in human life, are the most difficult to explain, and, I believe, hold the greatest philosophical interest.[3] Going forward, by "hard cases" I mean cases of this paradigmatic variety.

Second, I take "better than," "worse than", and "equally good" as *master normative relations* that can, when relativized to an appropriate covering consideration, underlie certain other normative relations, such as being normatively stronger than, normatively less significant than, and equally warrant-providing. So being a stronger normative reason can be expressed as being better with respect to normative strength and having more epistemic warrant can be expressed as being better with respect to epistemic warrant. Sometimes "better than" is understood as confined to comparisons within axiology, but master relations, like "better than," "worse than," "equally good," are not so confined; when we say that two things are "equally good" with respect to V-ness, we are saying that they are normatively speaking equally V, where V-ness may be a nonaxiological consideration. Stipulating that there are such relations in terms of which all comparative relations can be expressed allows us to investigate the structure of normativity as a whole, without restriction as to relata or covering consideration.

1 The Problem of Resolutional Remainder

"A is better than B with respect to V" is linguistically vague if there is vagueness somewhere in the predicate "better than with respect to V" or in the concept it expresses (or strictly in "A" or "B" or their associated concepts, but I put that possibility aside here) such that it is indeterminate how the predicate or concept applies to A and B. In hard cases of comparison, we might say that the indeterminacy is in which of the master predicates, "better than," "worse than," and "equally good," when relativized to V, applies to A and B.[4] Just as it is "hard" to say whether Jason Statham is bald, it is "hard" to say whether investment banking is a better career than interior design. The hardness in hard cases, so the suggestion goes, is a matter of vagueness in our language.

Linguistic vagueness, however, is *prima facie* pretty implausible as an explanation of hard cases. The hardness in comparing a career in investment banking and one in interior design is substantive and normative, not something that disappears through linguistic stipulation. When it comes to vagueness in language, the difficulty is whether to apply "bald" to Jason Statham or "red" to a colored patch, a problem about whether to *call* Statham bald or the patch red, a linguistic matter that can be settled through linguistic stipulation. Hard cases are not a matter of whether to *call* one career better than another but of figuring out the substantive normative matter of how the careers normatively relate (Chang 2002a; Schoenfield 2015). That hard cases are not made easy through linguistic stipulation seems a compelling reason to reject linguistic vagueness as an explanation of such cases. But language and substantive reality may be closer than this quick argument supposes. And there is always the possibility, though highly controversial (e.g., Sainsbury 2010), that the vagueness is metaphysical – that is, in the world rather than in our language, which raises problems that do not necessarily disappear with linguistic stipulation. The arguments I propose against vagueness accounts apply to vagueness in both their linguistic and metaphysical varieties.

Metaphysical vagueness is vagueness not in language but in the world. A relation or property might be thought to hold indeterminately of some items; does the property of being part of a particular cloud hold of a particular water droplet in the sky? It may be thought to be metaphysically indeterminate whether it does. Following Akiba (2004), Barnes (2010), and Williams (2008), we might think that metaphysical vagueness holds when there are multiple fully determinate worlds – some in which, say, a given water droplet is part of the cloud and some in which it is not – but it is indeterminate which of those worlds is the actual world. In hard cases of comparison, we might say that it is metaphysically indeterminate whether A is better than B with respect to V if in some worlds A is better than B and in other worlds it is not, and it is indeterminate which world is actual. Just as whether a given water droplet is part of a cloud

is "hard," so too is whether investment banking is a better career than interior design. The hardness of hard cases, so the suggestion goes, is a matter of vagueness in properties or relations in the world.

Now there is an intuitive sense in which it is appropriate to resolve vagueness through stipulation. If it is vague whether Jason Statham is bald, there is an intuitive sense in which we can "resolve" the question of whether he is simply by arbitrarily stipulating that he is (or isn't). We might say that even though there is an answer to the question of whether he is bald – namely, that it is indeterminate whether he is, the question nevertheless admits of a *resolution* as to whether he is, and we can always resolve the question by arbitrary stipulation.

By "resolution," I mean the thin, neutral idea of knowable application or lack of application of the predicate, where the identity of the predicate remains the same although the associated concept or property may change through resolution. My interest is in *intrinsic* resolution – that is, a settling of whether the predicate applies or not solely on the basis of facts about how the items relate, and not on extrinsic factors, such as a million-dollar reward you will get if you resolve the vagueness one way rather than another.[5]

Consider a supervaluational account of semantic vagueness (Fine 1975). Here the problem is in our language; we have a word, "bald," and it is indeterminate whether "bald" applies to Statham. The word "bald" and the concept of baldness it expresses are not sufficiently precise determinately to answer the question of whether Statham is bald. So we can sharpen up the concept of being bald by offering different neighboring concepts of being bald, each of which is represented by a different sharpening, and arbitrarily stipulate that one of those concepts is expressed by the predicate "is bald." There is nothing in our language or concepts that favor choosing one sharpening over another; we can arbitrarily stipulate a predicate and its associated concept as applying to Statham or not.

The same holds for metaphysical vagueness. Here the problem is with the world; there are multiple fully determinate worlds and indeterminacy in which of those worlds is actual. So we can sharpen up the target actual world by choosing one of the fully determinate worlds as actual. There is nothing in the world that favors choosing one determinate world over another; we can arbitrarily stipulate a determinate world in which Statham is bald (or not) as the actual world. As R. J. Williams, one of the leading proponents of metaphysical vagueness, puts it, the resolution of metaphysical vagueness is a matter of "randomly and groundlessly" making a judgment call (R. J. Williams 2016: 429).

Hard cases are different. In a hard case, arbitrary fiat never intrinsically resolves the case. If, in attempting to resolve the hard comparison between investment banking and interior design, you arbitrarily stipulate that the investment career is better, you have not settled the matter but are left with "resolutional remainder" – that is, the substantive question

of what normative relation holds between the careers has not been closed but remains an open question. Or consider a hard case involving statistical evidence that John committed a tort and individualized evidence that he did not. If you arbitrarily stipulate that the statistical evidence is better with respect to providing epistemic warrant, you do not thereby settle the matter of whether it is – the question of which evidence is weightier remains an open question. In cases of vagueness, by contrast, arbitrary resolution does close the question. Resolutional remainder is a state in the world, not in us, concerning a normative matter. It obtains when that normative matter remains open after an attempt at resolution, however we might think or feel about the case. Vagueness fails to explain why there is resolutional remainder after arbitrary stipulation and thereby under-delivers as an explanation of hard cases (see also Chang 2002a, forthcoming).

It might be thought that "resolutional remainder" in hard cases can be explained away in other terms. When you arbitrarily stipulate which of two careers in a hard choice is better, the matter of what normative relation holds between them remains open perhaps because we can never be sure that we followed the correct epistemic procedure in judging that neither is better, or because we are uneasy about our stipulated resolution since we recognize that our peers may stipulate differently (Williams 2016), or because we are filled with angst, regret or uncertainty in "high stakes" cases, like those involving comparisons of careers (Constantinescu 2012; Williams 2016). These are all extrinsic features of hard cases to which indeterminists have helped themselves in attempting to account for resolutional remainder.

But these suggestions misunderstand resolutional remainder in two ways. First, it is intrinsic to hard cases, not a downstream consequence of arbitrary stipulation that is contingent on the circumstances. Arbitrary stipulation in a hard case leaves resolutional remainder as an intrinsic feature of such cases, regardless of the circumstances surrounding the choice. It is not, for instance, a feature of only "high stakes" cases, like comparisons between careers, places to live, or people to marry. A primary school teacher might face a hard case in judging which of two qualitatively very different finger paintings is more beautiful. If she arbitrarily stipulates that Timmy's is more beautiful than Tommy's, her arbitrary stipulation does not settle the matter; the question of how the paintings relate with respect to beauty remains open, although answering that question may be of little intrinsic or extrinsic importance. Nonarbitrary stipulation in a hard case can be always be made on extrinsic grounds. The fact that Tommy won a prize for a pencil drawing last week may be an extrinsic ground for stipulating this week that Timmy's is more beautiful. Being low stakes, not having adequate time to consider the matter, achieving some benefit, and so on, can be extrinsic grounds for stipulating one resolution rather than another. But arbitrary stipulation on intrinsic grounds will always leave resolutional remainder.

Second, resolutional remainder is a metaphysical, not a psychological phenomenon.[6] Arbitrary stipulation in cases of vagueness settles the question of whether the predicate applies, while in hard cases, after arbitrary stipulation, the metaphysical question of how the items normatively relate remains open. Whether we feel angst, uncertainty, or unease about the matter is not at issue; after all, you might be a confident stipulator who is unfussed about the fact that your epistemic peers could have stipulated differently or that the case is high stakes. The issue concerns the metaphysical upshots of arbitrary stipulation in answer to the question of what relation holds – does arbitrary stipulation close the matter or not? In cases of vagueness, it does; in hard cases, it does not.

Could the fact that hard cases are *normative* be why arbitrary stipulation yields resolutional remainder? Perhaps it is the feature of being a *normative* hard comparison that makes arbitrary stipulation in such cases different from arbitrary stipulation in nonnormative cases of vagueness (Constantinescu 2012). The problem with this suggestion is twofold. For one thing, if arbitrary stipulation resolves vagueness in nonnormative hard cases but not in normative hard cases, an account is owed of why this should be. One way such an account could be developed is by exploring whether all normative hard cases are what Gallie (1956) calls "essentially contested," that is, by their nature always open to further substantive debate, while nonnormative ones are not. Perhaps the proponent of vagueness could develop an account along these lines. But note that in doing so she would be abandoning what we currently understand as *vagueness* and proposing that there is some new phenomenon according to which arbitrary resolution leaves resolutional remainder. I suggest a way to think about this new phenomenon in the last section of this chapter.

But there is a deeper worry. If the hardness of hard cases cannot be resolved by arbitrary stipulation according to the vagueness theorist, it might be wondered what work vagueness does in explaining such cases. After all, if attempts to resolve the putative vagueness in such hard cases must fail to provide an intrinsic resolution in such cases, why should we think that vagueness is what explains the case in the first place? An appeal to vagueness would be otiose in normative hard cases. Sharpening the question gets us nowhere; we are left where we began, namely, with a substantive, normative question about how two items normatively relate in a hard case.

None of this is to deny that vagueness holds in the normative domain or even in hard cases.[7] Vagueness most plausibly holds when the items at stake are quite similar and we need arbitrarily to "draw a line" among these similar items to determine what counts as being V. Excrement, for example, smells disgusting. If we alter a steaming pile of it ever so slightly by adding one drop of Chanel No. 5, it will still smell disgusting. But if we add another drop and then another and another, at some point it will

no longer smell disgusting. Exactly where we draw the line and call the smell no longer disgusting is an arbitrary matter because it is vague whether something smells disgusting. The same goes for determining whether a life is good. Fixing all the features that go into making a life good, it may be an arbitrary matter where we draw the line in calling such lives good enough to count as "good lives." The notion of a good life may well be vague. But the problem of comparing two qualitatively different lives with respect to "goodness as a life" in hard cases is not a matter of drawing a line among similar lives to determine which lives are good enough to count as "better as a life" than other lives. A property can be vague (e.g., "long") while its comparative is not (e.g., "longer than").

In sum. Vagueness can be resolved by arbitrary stipulation, but in hard cases, attempts to resolve the hardness by arbitrary stipulation leave resolutional remainder. Vagueness cannot accommodate resolutional remainder unless it is taken to be a phenomenon that is different from how it has so far been understood. If vagueness is understood in this new way, there is reason to think that, as an explanation of hard cases, vagueness thereby becomes otiose.

2 The Problem of Normative Leakage

Hard cases raise an interesting puzzle for rational choice (Chang 1997). Suppose, as many do, myself included, that a comparison between alternatives A and B with respect to V determines the rational choice between A and B, where V is what matters in the choice between them.[8] Suppose, moreover, that as the vagueness theorist would have it, in a hard case of comparison, it is always rational arbitrarily to stipulate that one is better than the other. Putting these suppositions together, we might accept the following "bridge principles" connecting comparisons with rational choice:

Bridge Principles

(1) If A is better than B with respect to V, and V is what matters in the choice between them, then one should rationally choose A.

(2) If A is worse than B with respect to V, and V is what matters in the choice between them, then one should rationally choose B.

(3) If A and B are equally good with respect to V, and V is what matters in the choice between them, then it's always rationally permissible to flip a coin to determine which to choose.

(4) If the comparison between A and B with respect to V is hard, and V is what matters in the choice between them, then the choice is hard, and it's always rationally permissible to arbitrarily stipulate that one is better than the other and choose that option.

(Although both 3 and 4 involve arbitrariness along the route to choice, only the latter involves arbitrary stipulation that one alternative is better with respect to V.) Going forward, let us assume that these bridge principles hold.

Given these principles, hard cases permit what we might call "normative leakage"; they can lead to the loss of normativity for a chooser through a series of rational choices.[9] Suppose you are contemplating whether to pursue investment banking or interior design and that the comparison of those careers, and thus the choice between them, is hard. Following bridge principle 4, you arbitrarily stipulate that interior design is better and rationally choose it. I now offer you the opportunity to consider a banking career instead, one that is identical to the one you forewent except worse in that there will be no year-end bonus. The comparison between the two careers is once again hard and so is the choice. Employing bridge principle 4, you arbitrarily stipulate that the banking career-sans-bonus is better than the design career and choose accordingly. But now you have ended up with a career – banking-sans-bonus – that is worse than a career you could have had a moment ago – banking – through a series of rational choices involving hard cases. If arbitrary stipulation is always permissible in hard cases, then hard cases make us liable to normative leakage.

If we understand hard cases as cases of vagueness, can normative leakage be blocked? Since the vagueness theorist maintains that arbitrary stipulation is always permissible in the face of a hard case, it may seem that vagueness permits normative leakage. But the opposite is true. Vagueness theorists hold as an integral part of their theory "penumbral truths," such as those given by the ordering properties of the usual trichotomy of relations, like transitivity and consistency. This is precisely to block untoward consequences that would otherwise ensue, such as normative leakage. (Strictly speaking, some minority views, e.g., degree theorists, reject penumbral truths and have been uniformly criticized as inadequate accounts of vagueness on that score (e.g. Fine 1975)). Since all mainstream accounts of vagueness accept penumbral truths as key components of the account, I set minority views aside.[10]

Put supervaluationally, there can be no sharpening of "better than as a career" that permits the three judgments that interior design is better than banking, that banking is better than banking-sans-bonus, and that banking-sans-bonus is better than interior design. Each alternative meaning of "better than as a career," then, will obey consistency constraints and the transitivity of "better than." Thus there is no sharpening on which normative leakage will be permissible. Moreover, "switching" between sharpenings in cases of putative normative leakage would involve equivocation across choice situations. You might judge that design is *better1* than banking, that banking-sans-bonus is *better2* than design, and that banking is *better3* than banking-sans-bonus, but it would be odd to

make choices across different possible careers on the basis of such equivocal judgments. Broome sums things up nicely on behalf of vagueness theorists: an agent who normatively leaks "necessarily does something wrong." (Broome, this volume: 46).

The fact that vagueness blocks normative leakage might be thought to count in its favor as an explanation of hard cases. After all, if hard cases should block normative leakage and understanding them in terms of vagueness allows them to do so, then vagueness meets that adequacy condition of an explanation of hard cases.

But should hard cases block normative leakage? I want to suggest that it is an *intrinsic* feature of hard cases that they make normative leakage in the hard choices they determine rationally permissible. If hard cases rationally permit normative leakage, then the problem of normative leakage is not the problem of explaining hard cases and the choices they underwrite in a way that blocks such leakage. On the contrary, the correct explanation of hard cases would show how such leakage in the context of rational choice is rationally permissible. Since vagueness prohibits normative leakage, it over-delivers as an explanation of hard cases, imposing on them a feature that is alien to them.

To see why the rational permissibility of normative leakage is an intrinsic feature of hard cases, we need to ask what role hard choices play in rational life. To tackle this question, we can start by imagining a world without hard cases – in particular, a world in which it never seems that one of the standard trichotomy of relations fails to hold between two items. In Easy World, it is always evident which of the standard trichotomy of relations holds, and what's more, we are always right. We can immediately, reliably, and veridically see that banking is better than interior design, living in the country is better than living in the city, marrying Adam is better than marrying Brian, and so on. Perhaps we have evolved to have unfailing normative instincts that give us direct access to normative facts so that all comparative normative truths hit us like basic perceptual truths we can discover by just looking. What would such a world be missing?

Easy World has two striking features. First, normativity would be a dictatorial overlord, always determining in every possible set of circumstances whether one thing is at least as good as another and, given our first three bridge principles, what you should think, feel, and do in that situation. In a world with only easy choices, there would always be a best career, place to live, person to marry, number of kids to have, car to buy, and so on since between any two options, one would be at least as V as the other. Sometimes normativity would determine that two options are equal bests, in which case it wouldn't matter which you choose, and so you could rationally flip a coin between them. If all choices were easy, normativity would dictate every aspect of rational life, including when you should flip a coin between alternatives.

Second, rational agency would be fundamentally *passive*. In a world with only easy choices, rational agency would be a passive capacity – an instinct or a normative perception – to discover orders dictated by normativity concerning what to believe, feel, and do in any situation. To exercise rational agency, all you would need to do is sit back, relax, and let normative truths tell you how you ought rationally to respond. There would be no room in Easy World for *active* rational agency – that is, exercises of agency in which the rational agent herself determines for herself what she should rationally think, feel, and do.

Of course, there is a version of a world without hard choices in which we don't immediately perceive normative comparative truths but have to engage our rational capacities to discover those truths. This "Challenging World" may be thought to be our actual world: one thing is always at least as good as another, but it often requires exercises of rational agency to determine that this is so. Challenging World, however, is like Easy World in the respects of interest: although it takes rational effort to discover what normativity demands of us, normativity is nevertheless a dictatorial overlord, always telling us what is at least as good as what. And although discovering normative truths is hard work, involving the exercise of our rational capacities, those exercises are passive in the sense of interest: they permit us to discover existing truths, not give us the power to determine or create normative truths.

This thought experiment suggests that hard cases play two important and distinctive roles in our world. First, they give rise to junctures in human life in which normativity "runs out" or is "silent" as to what a rational agent should think, feel, or do.[11] Second, because normativity does not determine a rational response in such cases, they make space for the exercise of *active* rational agency in which a rational agent can determine for herself what it would be rational for her to think, feel, and do.

This space for the exercise of active rational agency allows agents *rationally* to *change their minds* about what they should think, feel, and do and to do so at a normative cost. Faced with a hard case involving banking and interior design, you might rationally choose design. In the next moment, you might be faced with another hard case involving design and banking-sans-bonus. You might rationally choose banking-sans-bonus. This series of rational choices in hard cases leaks normativity; you end up choosing a career that is worse than one you could have had moments before. But this is what rational human life is like. There is nothing rationally to regret; leaking normativity in the course of human life is a feature of rational human life. Human rationality makes space for rational agents to change direction, switch gears, pursue a new path, try out something new, and so on, even though doing so entails leaking normativity over the course of their life. Hard choices are the junctures at which such leakage is rationally permissible. We might say that the freedom to leak normativity is what hard cases are *for*.

The rational permissibility of normative leakage in hard cases is what allows rational agents to determine the direction of their lives. I believe that this self-determination involves the exercise of a robust normative power that lies at the heart of an understanding of rational agency as active (Chang 2021). Vagueness, which makes normative leakage a mistake, misunderstands the role of hard cases in human life.

A proponent of vagueness can of course appeal to extrinsic factors to explain how leakage in hard cases is rationally permissible. Those extrinsic factors may override the rational mistake inherent in normative leakage if hard cases are cases of vagueness. But if hard cases play the distinctive role in rationality that I have suggested, the rational permissibility of such leakage is intrinsic to them.[12]

In sum. Hard cases provide junctures in life in which agents can actively determine what they should think, feel, and do, despite the normative costs. They are cases in which normative leakage is rationally permissible. Vagueness, which makes normative leakage a rational mistake, is therefore inadequate as an explanation of such cases.

3 Parity and Hard Cases

Hard cases offer us a path to a new way of thinking about the structure of normativity. They give us reason to reexamine an unreflective assumption we make about normativity: when we make normative comparisons with respect to some V, normativity permits only three possibilities: A is better than B, worse than it, or they are equally good with respect to V. Normativity is assumed to be "trichotomous" in structure. This assumption can be seen to derive from an analogue in the nonnormative domain; when making nonnormative comparisons, say, with respect to length, the nonnormative domain permits only three possibilities: one item is longer, less long, or equally as long as the other. "Trichotomy," the view that between two comparable items with respect to some covering consideration, only one of three relations corresponding to or analogous to the trichotomy of relations "more than", "less than," and "equal to" could hold between them, has been assumed by theorists working in the normative and nonnormative domains alike. It is foundational to much work in the social sciences, especially economics and decision theory but is also widely assumed in the humanities and physical sciences.

I have argued elsewhere that Trichotomy does not hold in the normative domain. This is because the significant *qualitative* differences in normativity between items make Trichotomy too crude a view to capture the full range of possible normative relations among items (Chang 2016a). When we compare two careers with respect to goodness as a career, two policies with respect to justice, two reasons with respect to normative strength or significance, two sets of evidence with respect to warrant, or two scientific theories with respect to explanatory power, we

cannot assume that these normative covering considerations – goodness as a career, justice, normative significance, warrant, explanatory power – are like nonnormative covering considerations, such as length, weight, and volume, that permit only trichotomous rankings. There is a fourth, sui generis, master relation by which things can be normatively related, what I call "on a par." Two careers can be on a par with respect to goodness as a career: neither is better than the other but nor are they equally good. They are nevertheless comparable; they are qualitatively different and yet in the same neighborhood of goodness as a career, they are on a par. Similarly, statistical evidence that the defendant committed the tort may be neither stronger, less strong, nor equally as strong as forensic, individualized evidence arguing the contrary. It isn't that we can't compare the epistemic warrant of each kind of evidence; we're not trying to compare epistemic and practical reasons, for instance, or a weight and a volume. The statistical evidence is on a par with the individualized evidence. In short, when making normative comparisons, we should reject Trichotomy and accept instead Tetrachotomy, the view that between two comparable items with respect to some covering consideration, only one of four relations – "better than," "worse than," "equally good," and "on a par" – could hold between them.

Hard cases are, I suggest, cases in which the items are on a par. Normativity is tetrachotomous, not trichotomous, in structure.[13] There are four, not three master relations in the normative domain.

There are of course many questions that arise in the face of such a seemingly radical proposal, many of which I try to tackle elsewhere, and I won't repeat my arguments here. My aim instead is to describe how parity solves the two problems we have raised for vagueness theorists.

The first, concerning resolutional remainder can be straightforwardly explained. On the assumption of Trichotomy, arbitrarily stipulating that, say, investment banking is better than interior design leaves resolutional remainder because Trichotomy is false: the careers are on a par. The question of what normative relation holds between the careers cannot be settled by arbitrarily stipulating that one is better if the normative truth is that they are on a par. Arbitrary stipulation leaves resolutional remainder in hard cases because we assume that Trichotomy holds.

The second problem is to explain how hard cases, understood as cases of parity, rationally permit normative leakage. To solve this problem, we must ask what bridge principle connects the fact that options are on a par with how rationally to choose between them. That is, we need a bridge principle that replaces principle 4 mooted earlier for hard cases. I have suggested that the bridge principle is disjunctive: if options are on a par, it is rationally permissible to *commit* to (a feature) of one, thereby *creating* normativity in its favor and perhaps now having most reason to choose it *or* to *drift* into one of the options – that is, intentionally choosing it on the basis of some consideration that counts in its favor.

Bridge Principle for Hard Cases

> 4* If A and B are on a par with respect to V, and V is what matters in the choice between them, then it is always rationally permissible to *commit* to (a feature of) one of them, thereby creating a reason to choose it, which may then give one most all-things-considered reasons to choose it, *or* to *drift* into one option – that is, to intentionally choose it on the basis of already-existing reasons but without committing to any of its features.

The difference between committing and drifting is of the greatest importance. When you commit to something and thereby create a reason to choose it, you are engaging in the volitional activity of putting yourself behind that consideration. It is this putting yourself behind something, say, the lucre you will earn as an investment banker, that creates normativity in favor of the banking career. Since creating reasons is an active exercise of rational agency, hard cases, understood in terms of parity, rationally permit you to determine for yourself what you have most reason to think, feel, and do. Normativity doesn't dictate what you should think, feel, and do in every possible circumstance; it leaves you with "hard choices" – that is, cases in which the items are on a par, and you have the freedom to create for yourself a reason to pursue one thing over another, a freedom that is itself not governed by normativity. Committing to something is something rational agents *do* as an exercise of their rationality that is *not* guided by reasons (Chang 2021). In a hard choice, then, you can create a new "will-based" reason for yourself to pursue banking-sans-bonus, thereby making it better, we can suppose, than interior design. There is no mistake of rationality even though a moment ago you could have had banking *with* a year-end bonus. This is because hard choices allow you space to change direction in your life by creating normativity that favors new life paths even at the cost of leaking normativity. Moreover, how much normatively it is permissible to leak is constrained by the requirement that the options are in fact on a par. Rational life allows you to change direction at normative cost without making a rational mistake, but too much leakage – for example, leakage when options are not on a par – will undermine rational agency.

None of this is to say that hard cases, understood in terms of parity, must always permit normative leakage. If today you commit to the autonomy afforded by a career in interior design, your doing so confers normativity in favor of the design career that wasn't there before and which could now give you most all-things-considered reasons to become an interior designer over a banker. Tomorrow, when you are offered the chance to reconsider your choice of career, this time being offered the same banking career but without a year-end bonus, the design career may be better than both the original banking career and the banking

career without the bonus because of your commitment. In this way, what you chose previously can constrain what you can rationally choose subsequently; commitment in one choice can block what would have been normative leakage in a subsequent choice (Chang 2005).

Understanding hard cases in terms of parity also provides us with a diagnosis of the hardness in hard cases. Hard cases are hard because we assume too crude a view of the structure of normativity – that all normatively significant qualitative differences can be forced to fit within the usual three categories of relation corresponding to "more," "less," and "equal" in the nonnormative domain. Once we allow for the possibility of parity, determining which relation holds between two items in a hard case will be (relatively) easy: the items are on a par. But there is a hardness that remains in hard choices. In hard choices, we are rationally permitted to commit or to drift, and which we do is not a choice guided by reasons but a volitional activity. The hardness is volitional: can we commit or is the will content to drift? Committing to something is hard; it is a matter of putting your very self behind something. We commit in our friendships, love relationships, and personal projects. But commitment beyond these spheres is not yet recognized as part of what it is to be a rational agent. Instead, we are taught that when we face a hard choice one option is at least as V as the other and that our job as rational agents is to discover which it is. We are taught that rationality is a matter of discovery of reasons, not their creation through our commitments.

In sum. If we understand hard cases as cases of parity and not vagueness, we can explain both why arbitrary resolution of hard cases leaves resolutional remainder and why normative leakage is rationally permissible. Hard cases and the hard choices they underwrite are junctures in human life in which rational agents can change direction in their lives despite the normative costs of doing so.

* * *

I end with a nudge for vagueness theorists – at least of the traditional, non-epistemicist variety. Although I believe that indeterminacy understood as vagueness is the least plausible of the "i" explanations of what I have called "paradigmatic" hard cases, I also believe that those who appeal to vagueness to explain such cases are very close to the truth of these cases, i.e., that they are *almost* partisans of parity. This is because, they share with defenders of parity two key insights about hard cases – namely, (i) that they are *not* cases of incomparability and (ii) that it is implausible to think that in all hard cases between qualitatively different items, one is always at least as good as the other but we just don't know which. Surely, we can compare two careers, places to live, and human lives, even if they are qualitatively different. And is it really plausible to think that between every pair of qualitatively different careers, such as banking and interior design, there is always a precise truth about whether

increasing the salary of one by a dollar thereby makes it better, but we are too stupid to discover that truth? Maybe such precision holds of baldness and being a heap, but it is hard to believe it holds of goodness as a career, well-being, justice, epistemic warrant, and the like.

I wonder whether proponents of vagueness have come to vagueness as an explanation of hard cases largely by default. Indeed, indeterminacy sometimes seems to serve as a dumping ground for unclear or difficult normative determinations. Once the vagueness theorist accepts that not all hard cases can be explained by incomparability or ignorance, then on the assumption of Trichotomy, indeterminacy is the only remaining explanation. As I have argued in this paper, however, indeterminacy understood as vagueness is not the right kind of phenomenon to explain hard cases.

To get on the path to parity, the vagueness theorist needs only to take two steps. First, she needs to recognize that the assumption of Trichotomy is unreflectively built into certain, especially economic, approaches to hard cases and that this assumption requires examination and defense. Second, she needs to investigate the nature of the normative considerations she is attempting to model, without simply assuming that they are amenable to the same modeling fit for nonnormative considerations, such as length, weight, and volume. I have argued that qualitative differences in normativity make space for a fourth, sui generis way in which items can be normatively related (Chang 2002a, 2013a, 2013b, 2016a). The structure of normativity is not like the structure of nonnormativity. We need parity to explain hard cases.

The history of philosophy is filled with attempts to explain difficult phenomena in familiar terms. When familiar tools strain to explain very real and important phenomena, we make philosophical progress by adding new tools to our explanatory toolbox. Hard cases call on us to reexamine deep and unreflective assumptions we make about the structure of normativity. By adding parity to our explanatory toolbox, I believe that we put ourselves on a path to a deeper understanding of hard cases and the nature of normativity.

Notes

1 Thanks to the editors of this volume and to the audience at the UK ALPP conference, especially Matt Kramer, Rae Langton, and Re'em Segev, for comments that led me to make some useful clarifications, and to Kit Fine for discussion about the varieties of vagueness that saved me from many infelicities and helped me to simplify my arguments.

2 If my arguments are correct, the positive arguments offered by proponents of vagueness are not. The most trenchant of these arguments are provided by John Broome (1997, this volume). If Broome's arguments are correct, then two ways of understanding value relations that make room for both vagueness and the determinate failure of the usual trichotomy of relations offered

by Wlodek Rabinowicz (2007, 2012, see also 2009) and myself (2002b, 2016a) are mistaken.

I believe that Broome's arguments depend on a controversial assumption. They assume that *all* banking careers (or what not featuring in vagueness) can be arrayed on a spectrum such that it makes sense to say of each career on the spectrum that it is (trichotomously) "closer" or "further" from being a version of the banking career that is definitely better (or worse) than the interior design career. I have suggested (Chang 2002a) that there are at least two reasons we cannot expect that an item like a career can always be tweaked with small, successive normative changes to generate a spectrum of such careers where each career is increasingly good as a career. First, a small normative change can trigger a new aspect of goodness of career that wasn't relevant before, thus making the new career not better with respect to goodness as a career but only with respect to some additionally smuggled in covering consideration. I call this the Hegelian Proviso. Second, a small normative change in a career can make a career not better but worse than it was before because of organic unities. I call this the Aristotelian Proviso. These provisos show that we cannot *assume* that a spectrum of the sort on which Broome's arguments rely is always available. (Although I rely on a similar spectrum in my arguments for parity, my burden is to show that there is at least one such spectrum involved in a hard case, not that all hard cases are amenable to treatment in terms of the kind of spectrum that Broome envisages).

Other proponents of vagueness have offered mostly negative arguments against my arguments that hard cases are not cases of vagueness. Some telegraphic worries about such arguments that could not be addressed explicitly in this chapter are as follows:

Wasserman's (2004) objections suggest (a) that there is a difference in phenomenology involving vagueness of monadic as opposed to polyadic predicates, but I can see no difference in phenomenology in the relevant cases; my point is that in hard cases, it *seems determinately* that none of the usual trichotomy of relations holds (that the one career is *not better* than the other) and we are not inclined to judge that it's unclear whether 'better than' holds, as we might in cases of vagueness so the phenomenology is different; (b) maintains that the idea of resolutional remainder begs the question against a previous argument that ignorance as to which of the usual trichotomy holds is not always in play in hard cases since the idea of a resolutional remainder presupposes that one of the usual trichotomy of relations holds but we don't know which. This worry itself, it seems to me, begs the question against the possibility of parity and misunderstands the idea of resolutional remainder, which does not entail the claim that one of the usual trichotomy holds even after arbitrary stipulation but that the substantive question of what relation holds between items remains open.

Klocksiem (2010) rightly points out that the Small Improvement Argument alone does not establish incomparability but is consistent with vagueness, which is why there is a need to argue explicitly against the possibility that one or both of the Small Improvement Argument and the Chaining Argument trade on vagueness, which I try to do in Chang 2002a.

Sugden (2009) rightly argues that if we assume trichotomy and frame the question about hard cases as one about modelling them by the relation "at least as R as," then we might as well think about them as cases of vagueness since he fiats no space for contemplating a normative reality that defies these assumptions. However, the *philosophically* interesting question is about normative reality of hard cases, not how we can impose neat decision-theoretic

models on them that may not accurately reflect their features (see also Qizilbash 2014).

Elson's (2014) interesting and complex argument that my arguments for parity trade on vagueness wholly depends on the thought that "is comparable" (or being comparable) (with respect to V) can be vague without any one of "better than," "worse than," or "equally good" (with respect to V) being vague, assuming trichotomy. I don't see how this is plausible though it is a logical possibility. As a tetrachotomist, I would deny that "is comparable" (with respect to V) can be vague without at least one of "better than," "worse than," "equally good," or "on a par" (with respect to V) being vague (where the vagueness of course could derive wholly or partly from V-ness). If comparability is constituted by a certain range of relations, its vagueness seems to be a function of the vagueness of one of its components. Moreover, in Elson's (2017) it is suggested that the fact that there can be a quandary over how to trade off multiple different components of the covering value makes for a prima facie case in favour of vagueness, thereby shifting the burden to opponents of vagueness to explain why vagueness does not explain these cases. But a quandary over the "rate of trade-off" across components of, say, justice, is the bread and butter of substantive normative theorizing; it is this quandary, which is *substantive*, that accounts for many of the competing "conceptions" of justice that are each substantive and not simply sharpenings of the vague term or property of being just. Simply having the form of there being multiple legitimate ways to relate different criteria is not *ipso facto* grounds, it seems to me, for accepting vagueness as the default explanation of hard cases.

Andersson (2015) suggests that my assumption that a small unidimensional difference in an item cannot trigger incomparability where before there was comparability already begs the question of whether the items are comparable because "all differences [can] be understood to be composed of small uni-dimensional differences" (p. 252). However, this principle, which I understand normatively, not nonnormatively as Andersson supposes in his main text (p. 673 though he claims the arguments apply even to a normative version of the principle), does not beg the question since there are normative differences between incomparable items, too, which presumably can be broken down into small unidimensional differences if normative differences can. As Andersson himself notes, I say very explicitly that the intuition that a small normative change in one dimension of the covering consideration in one item is not sufficient to trigger incomparability with some other item if before those items were comparable holds *in a demarcated set of cases* and not universally, and I lay out two kinds of cases in which it does not hold. So the question is whether my appeal to the principle holds in the cases I have in mind. Here, all I can do is urge the reader to think of such cases for her or himself – laying out any single case in all the necessary detail would be not only too onerous but ultimately controversial for at least some readers. I suggest that the Mozart and Michelangelo case is such an example.

3   The dialectical point of focusing on paradigmatic hard cases is to exclude cases involving two nearly identical items about which the "hardness" in comparison, if indeed there is hardness, *may* not be of the same sort as the hardness involved in hard cases involving two qualitatively different items. (In other work, I refer to this amorphous class as "superhard.") Throughout this chapter, I assume that ignorance (as opposed to epistemic vagueness) is not the problem, though exactly how to distinguish ordinary ignorance from the ignorance that is putatively vagueness is a difficult question. (Again, in previous work, I explore arguments against the idea that in hard cases "at

least as good as" holds between items, but we just don't know which in the above (Chang 2002a).) I also assume that incommensurability, the idea that items cannot be measured on the same cardinal scale representing the covering consideration, is a nonstarter since the hard cases of interest do not require commensurability for comparability.

 4 We could hold instead that the vagueness is not in which of the usual trichotomy of relations holds but in whether the options are comparable/incomparable. For this indeterminacy to hold, however, there must be indeterminacy in *what* relation holds, or so I will suppose.

 5 Extrinsic factors are, of course, always available to resolve vagueness in nonarbitrary ways. Tom Dougherty (2016) suggests that social conventions can step in to solve indeterminacies in morality. It may be worth noting that views of vagueness that appeal to extrinsic features to ground nonarbitrary resolution of vagueness need special arguments to explain why arbitrary resolution would not be permitted if, for example, the sharpenings are split 50-50 or the degree of truth of each proposition is the same.

 6 I suspect I am to blame for some of my interlocutors treating resolutional remainder as a psychological phenomenon. In Chang 2002a, I try to elucidate the metaphysical idea that the question of what relation holds remaining open by talking of the "perplexity" over what relation holds "persisting." Here I meant "perplexity" as a conundrum, difficulty, puzzle, not a psychological state of being puzzled or perplexed.

 7 Indeed, the presence of vagueness and some of the other "i" phenomena may help to obscure what is really explains hard cases. We are, to be sure, ignorant in hard cases, but it is a mistake, I have argued, to think that ignorance explains why the case is hard.

 8 In Chang 2016b, I argue for a strong and tight connection between comparisons and rational choice: that if we (i) work with master comparative relations, R ("better than," "worse than," "equally good," and "on a par" (to be explained in the text that follows)); (ii) make the relata, A and B, include anything that can be said to be an object of choice; and (iii) allow covering considerations, V, to be whatever might matter in a choice between items, comparative facts of the form "A R B with respect to V" provide the *grounds* of all rational choice: they are that in virtue of which a choice is rational. According to "comparativism," comparisons occupy center stage in understanding practical normativity, whether you are a consequentialist, deontologist, virtue theorist, perfectionist, etc., because they are that in virtue of which choices, intentions, and actions are rational or normatively justified. Moreover, they provide a unified framework within which we can conduct debates within both practical and theoretical normativity – comparative facts provide the ground not only of rational choice but also rational belief.

 9 The idea of normative leakage is a cousin to the familiar idea in decision theory that cyclical preferences are disallowed on the pragmatic grounds that they would allow agents to money pump. We might say that cyclical comparisons in conjunction with bridge principles lead to the possibility of "normativity pumping."

10 In any case, even degree theorists are subject to the argument from resolutional remainder; insofar as degrees can be equal, arbitrary stipulation resolves the case but it does not do so in hard cases without resolutional remainder.

11 As we will suggest in the next section, normativity "runs out" only on the assumption that "better than," "worse than," and "equally good" are the only master relations.

12 Broome (this volume) cleverly notes that, given certain (controversial) assumptions, vagueness theorists can account for regret over a past choice by "distributing" the rational error involved in normative leakage. If you change your mind in the careers case, the rational mistake can be put mostly on your initial choice to go for interior design. But even if the rational error in normative leakage can be "distributed" in the way that Broome suggests, normative leakage always involves making a rational mistake somewhere in one's choices. Hard choices allow normative leakage without there being any rational error. (The controversial assumptions of Broome's argument are, first, that all banking careers can be arrayed along a spectrum of trichotomously increasing goodness as a career so that at some point along the spectrum, there is a banking career that is definitely better than the interior design career and, second, that there are trichotomous degrees of rational permissibility in choosing a career that are isomorphic with this spectrum of trichotomously increasing degrees of goodness as a career).

13 I have implied that trichotomy holds in the nonnormative domain and that tetrachotomy holds in the nonnormative domain. But this is not quite right. There is some reason to think that both domains are structured tetrachotomously – that is, that within each domain, there are properties or covering considerations that permit of tetrachotomous orderings. Within the normative domain, for instance, some normative covering considerations, such as "goodness of number of lives saved," where the goodness is measured by the number of lives saved, have a trichotomous structure – saving five lives must be better, worse, or equal to saving one life since their normative relation is determined by how many lives are saved and numbers can only be greater, lesser, or equal to another. Similarly, some nonnormative covering considerations may admit of tetrachotomous ordering. Which is more bulky, a bicycle or a 2 x 4 wall stud? Bulkiness is a multi-component covering consideration with qualitative aspects that arguably need not relate items trichotomously. Similarly, we might wonder whether an orangey-red patch must be redder, less red or equally as red as a purply-red patch. A bicycle and piece of lumber may be on a par in bulkiness, and two qualitatively different-looking red patches may be on a par in redness. These are controversial claims that I leave aside here (but see Chang 2002b).

References

Akiba, Ken. (2004), 'Vagueness in the World' *Nous* 38/3: 407–429.

Andersson, Henrik. (2015), 'Parity and Comparability—a Concern Regarding Chang's Chaining Argument', *Ethical Theory and Moral Practice* 19/1: 245–253.

Barnes, Elizabeth. (2010), 'Ontic Vagueness: A Guide for the Perplexed', *Nous* 44/4: 601–627.

Broome, John. (1997), 'Is Incommensurability Vagueness?' In R. Chang (ed.), *Incommensurability, Incomparability, and Practical Reason* (Cambridge, MA: Harvard University Press), 67–89.

Broome, John. (2021), 'Incommensurateness is Vagueness', in H. Andersson and A. Herlitz (eds.), *Value Incommensurability: Ethics, Risk, and Decision-Making* (Routledge).

Chang, Ruth. (1997) 'Introduction'. In Chang (ed.), *Incommensurability, Incomparability, and Practical Reason* (Cambridge, MA: Harvard University Press), 1–34.

Chang, Ruth. (2002a), 'The Possibility of Parity' *Ethics* 112:659–688.

Chang, Ruth. (2002b), *Making Comparisons Count*, (London: Routledge).

Chang, Ruth. (2005), 'Parity, Interval Value, and Choice,' *Ethics*, 114: 331–350.

Chang, Ruth. (2012), 'Are Hard Choices Cases of Incomparability?' in *Philosophical Issues*, 22/1: 106–126.

Chang, Ruth. (2013a), 'Grounding Practical Normativity: Going Hybrid,' in *Philosophical Studies*, 164/1: 163–187.

Chang, Ruth. (2013b), 'Commitments, Reasons, and the Will,' in Shafer-Landau (ed.), *Oxford Studies in Metaethics*, 8: 74–113.

Chang, Ruth. (2016a), 'Parity: An Intuitive Case', *Ratio* 29/4:395–411.

Chang, Ruth. (2016b), 'Comparativism: The Grounds of Rational Choice'. In *Weighing Values*, edited by Errol Lord and Barry Maguire (New York: Oxford University Press).

Chang, Ruth. (2017), 'Hard Choices', *Journal of the American Philosophical Association*, 92: 586–620.

Chang, Ruth. (2021), 'What Is It to Be a Rational Agent?' in Ruth Chang and Kurt Sylvan (eds.), *The Routledge Companion to Practical Reason* (New York: Routledge), 95–110.

Chang, Ruth. (forthcoming), 'How to Avoid the Repugnant Conclusion', In Jeff McMahan, Tim Campbell, James Goodrich, and Ketan Ramakrishnan (eds.), *Ethics and Existence: The Legacy of Derek Parfit* (Oxford: Oxford University Press).

Constantinescu, Cristian. (2012), 'Value Incomparability and Indeterminacy' *Ethical Theory and Moral Practice* 15/1: 57–70.

Constantinescu, Cristian. (2016), 'Vague Comparisons', *Ratio* 29/4: 357–377.

Dougherty, Tom. (2016), 'Moral Indeterminacy, Normative Powers and Convention' *Ratio* 29/4: 448–465.

Elson, Luke. (2014), 'Heaps and Chains: Is the Chaining Argument for Parity a Sorites?' *Ethics* 124/3: 557–571.

Elson, Luke. (2017), 'Incommensurability as Vagueness: A Burden Shifting Argument', *Theoria* 83: 341–363.

Fine, Kit. (1975), 'Vagueness, Truth, and Logic', *Synthese* 30: 265–300.

Gallie, W.B. (1956), 'Essentially Contested Concepts', *Proceedings of the Aristotelian Society* 56: 167–198.

Griffin, James. (1986), *Well-Being* (Oxford: Oxford University Press).

Klocksiem, Justin. (2010), 'In Defence of the Trichotomy Thesis'. *Acta Analytica* 25/3: 317–327.

Qizilbash, Mozzafar. (2005), 'Transitivity and Vagueness', *Economics and Philosophy* 21:109–131.

Qizilbash, Mozzafar. (2007), 'The Mere Addition Paradox, Parity, and Vagueness', *Philosophy and Phenomenological Research* 75/1: 129–151.

Qizilbash, Mozzafar. (2014), '"Incommensurability" and Vagueness: Is the Vagueness View Defensible?' *Ethical Theory and Moral Practice* 17/1:41–54.

Rabinowicz, Wlodek. (2007), 'Value Relations' *Theoria* 74/1: 18–49.

Rabinowicz, Wlodek. (2009), 'Incommensurability and Vagueness', *Aristotelian Society Supplementary Volume* 83: 71–94.

Rabinowicz, Wlodek. (2012), 'Value Relations Revisited', *Economics and Philosophy* 28/2:133–153.

Sainsbury, Mark. (2010), 'Why the World Cannot be Vague', *The Southern Journal of Philosophy* 33/1: 63–81.

Schoenfield, Miriam. (2015), 'Moral Vagueness Is Ontic Vagueness', *Ethics* 126,2,:257–282.

Sugden, Robert. (2009), 'On Modelling Vagueness – and on *Not* Modelling Incommensurability' *Aristotelian Society Supplementary Volume* 83: 96–113.

Thomas, Teru. (forthcoming), 'On Evaluative Imprecision', in Jeff McMahan, Tim Campbell, James Goodrich, and Ketan Ramakrishnan (eds.), *Ethics and Existence: The Legacy of Derek Parfit* (Oxford: Oxford University Press).

Wasserman, Ryan. (2004), 'Indeterminacy, Ignorance and the Possibility of Parity', *Philosophical Perspectives*, 18/1: 391–403.

Williams, J. Robert. (2008), 'Ontic Vagueness and Metaphysical Indeterminacy', *Philosophy Compass* 3/4: 763–788.

Williams, J. Robert. (2014),'Decision Making Under Indeterminacy' *Philosopher's Imprint* 14: 1–34.

Williams, J. Robert. (2016), 'Indeterminacy, Angst and Conflicting Values' *Ratio* 29: 412–433.

3 Parity Without Imprecise Equality

Chrisoula Andreou

1

In choosing between two options, one might, if time permits, compare the options and attempt to determine if one option is better and the other worse, or if the options are equally good (overall and relative to what matters in the situation at issue, which can include the fit between the options and the taste or preferences of the choosing agent). According to the trichotomy thesis, "the conceptual space of comparability between two [options] is spanned by the trichotomy of relations 'better than,' 'worse than,' and 'equally good': if none of those relations holds, the [options] are incomparable" (in terms of their overall value relative to what matters in the situation at issue).[1] The trichotomy thesis has been challenged based on the idea that two options that are not rankable as one better than the other or as exactly equally good might, at least intuitively, be positively related to one another as "on a par" (rather than just negatively related to one another as *not* exactly equally good and *not* one better than the other). But what is it for two items to be on a par?

A prominent and seemingly promising approach to elucidating parity is to build on the notion of rough equality – understood as involving closeness in value (more on this later) – and to cast the dispute between proponents and opponents of the trichotomy thesis as follows: according to proponents of the trichotomy thesis, all pairs of options that are roughly equally good are, if not exactly equally good, then such that one option is strictly better than the other, though not by much. According to opponents of the trichotomy thesis, this is not so; rather, in some cases, two roughly equally good options are *imprecisely equally good* – though close in value, they are not rankable in relation to one another[2] – and it is the relation between such options that the term "on a par" is best reserved for. This suggestion fits quite neatly with the familiar idea, which I will take on board, that if two options are on a par, then, even though the options are not rankable in relation to one another, they are in the same neighborhood in terms of their overall value (i.e., in terms of how

valuable they are overall) relative to what matters in the situation.[3] Given this idea, which captures the key features of the notion of parity that I am here concerned with, it is natural to conclude that options that are on a par must be close in value, and so roughly equally good.[4] But, as I will argue, this is not the case.[5]

Developing a view that I present, in part, elsewhere,[6] I will begin by focusing on the construal of parity as imprecise equality and on the case to be made for the possibility of imprecise equality. I think that the case is strong. But, building on the understanding of imprecise equality that I develop here and in some examples and ideas in some of my prior work, I explore the possibility that there can be cases of comparability beyond those acknowledged by the trichotomy thesis that are not cases of imprecise equality but are still cases of parity (wherein the options are not rankable in relation to one another but are in the same neighborhood). I then focus on the question of how two options could be in the same neighborhood in terms of their overall value (or, more precisely, in terms of how valuable they are overall relative to what matters in the situation at hand) if the options are not roughly equally good (and so are not imprecisely equally good). In responding to the question, I explain that, in evaluating a particular group of alternatives, some of which are not rankable in relation to others, one may need to use a "grading system" with wide-ranging evaluative classes. Significantly, when two options fall in the same wide-ranging evaluative class, they need not have the closeness in value necessary to qualify them as roughly equally good. Nevertheless, for options in the same class that, by hypothesis, cannot be ranked in relation to one another and cannot be "graded" using a grading system that employs narrower evaluative classes, counting the options as in the same neighborhood and, relatedly, as on a par seems perfectly appropriate even if the options cannot also be described as roughly equally good.

2

Items can be compared in all sorts of ways. Books can be compared based on, for example, their innovativeness or, alternatively, on their suitability to keep a fire going. The items of interest in this chapter are options that agents might have to choose between. Moreover, for two options to be comparable in the sense of interest here, it does not suffice that they are comparable in any old way.

> For example, two submissions for a creative writing contest are not necessarily comparable in the relevant sense just because one is better relative to *one* of the factors that matters given the context of choice, such as, say, originality. Similarly, the two submissions are not necessarily comparable in the relevant sense just because one is better in

terms of a factor that is altogether irrelevant given the choice situation, such as, say, suitability for use as a door stop.

(Andreou 2020a)

Furthermore, as suggested in section 1, the truth of a negative comparative claim such as, for example, that X and Y are *not* equally good, does not suffice to make the options comparable in the sense of interest here. For comparability in the relevant sense to obtain, there must be some *positive* relation between the options that captures how they compare in terms of their *overall value*, relative to what matters in the situation at issue, from the point of view of practical reason. Notably, whether there is a positive relation that captures how two options compare in the relevant sense does not depend on whether the two options are themselves good or bad options. In particular, two bad options can be positively related to one another as exactly equally bad.

   According to the standard view, all options (in a particular choice situation, given a particular choosing agent) are comparable (in the relevant sense); more specifically, according to the standard view, for every pair of options, one option is either better than the other or the options are exactly equally good. The standard view can be resisted in two ways. First, it might be argued that options are not invariably comparable – sometimes there is no positive relation between the options that captures how they compare in terms of their overall value. This is the *incomparability thesis*. Second (and additionally or else alternatively), it might be argued that comparable options are not always comparable as one better than the other or as exactly equally good. This way of resisting the standard view rejects the aforementioned trichotomy thesis, according to which, if two options are comparable, one option is better than the other, worse than the other, or exactly as good as the other.

   Before Ruth Chang's now prominent discussions of the trichotomy thesis, arguments that could be developed as either for the incomparability thesis or against the trichotomy thesis were sometimes automatically cast as arguments for the incomparability thesis. Consider the *small-improvement argument*, which originally played the role of the most promising argument for the incomparability thesis.[7] According to the small-improvement argument, there are options X, Y, and X+ – where X+ is a slightly improved variation of X – such that:

   (1) X is not better than Y
   (2) Y is not better than X
   (3) X+ is better than X
   (4) X+ is not better than Y

If it is granted (in accordance with the next step in the argument) that if two options are exactly equally good, then anything that is better than

one of the options will also be better than the other option, it follows from 3 and 4 that X and Y are not exactly equally good. But then, given 1 and 2, it follows that X and Y are neither one better than the other, nor exactly equally good.[8]

Filling in the small-improvement argument requires providing plausible candidates for X and Y (i.e., candidates for which 1–4 all seem clearly true). Prominent proposed cases include cases in which it seems plausible to suppose that the agent knows all there is to know about what matters in the case at hand, such as, for instance, Chang's (2002) coffee and tea case in which, by hypothesis, all that matters is whether and, if so, which drink tastes better to the agent. In Chang's case, the agent "rationally judge[s] that [a particular] cup of Sumatra Gold tastes neither better nor worse [to her] than [a particular] cup of Pearl Jasmine and that although a slightly more fragrant Jasmine would taste better than the original, the more fragrant Jasmine would not taste better than the cup of coffee" (2002: 669). If X = opting for the Pearl Jasmine, Y = opting for the Sumatra Gold, and X+ = opting for a slightly more fragrant Jasmine, we have, it seems, the sort of case we need.

But, even if this is right, we can, as Chang does, resist the incomparability thesis. For, it may be that cases that fit the small-improvement argument are best understood as cases that count against the trichotomy thesis, not against the comparability of the options under consideration. And, indeed, this seems like the best way of understanding the challenge raised by Chang's coffee and tea case. Even opponents of the standard view should, it seems, agree that Chang's coffee and tea case is not a case of incomparable options. The options seem roughly equally good. The small-improvement argument suggests that they are *not also* one better than the other or else exactly equally good, which in turn suggests that they are *imprecisely equally good*.

The preceding reasoning assumes, following Chang, that the debate between the proponents and opponents of the trichotomy thesis can be fruitfully construed as not a mere terminological dispute but as a substantive dispute about our "ordinary" or "intuitive notions of comparability and incomparability" (Chang 2016: 191–192).

In what follows, I will accept the small-improvement argument as an argument against the trichotomy thesis and consider whether all cases of parity (wherein the options are not rankable in relation to one another but are in the same neighborhood) must be understood as cases of imprecise equality. After saying more about imprecise equality, I will, building on my construal of imprecise equality and on my prior work on parity, provide considerations in favor of the conclusion that the answer is "no." As part of supporting this conclusion, I will focus on explaining how two options can be in the same neighborhood (in terms of how valuable they are overall relative to what matters in the situation) without necessarily being close in the sense implied by rough equality. Note

that, henceforth, the parenthetical remark in the preceding sentence will often be left implicit, as will my previously repeated remark regarding the notion of parity at issue in this chapter, according to which options that are on a par are not rankable in relation to one another but are in the same neighborhood.

3

As suggested previously, to say that two options are imprecisely equally good is to say that they are roughly equally good and that they are not also one better than the other or else exactly equally good. But what is it for two options, say X and Y, to be roughly equally good? One thing that seems necessary is that their overall values are close; and for this to hold it must be that were another option, W, considerably (i.e., quite a lot as opposed to just a little) better than one of the original options, it would also be better than the other. Importantly, this necessary condition leaves open the question of whether X and Y are precisely comparable to one another. Relatedly, it is compatible with there being an option that is a little better than one of the options but not better than the other; were this ruled out, rough equality would not be compatible with imprecise equality.[9]

Return to Chang's coffee and tea case. Chang's cup of Sumatra Gold and cup of Pearl Jasmine seem roughly equally good; the small-improvement argument suggests that they are not also one better than the other or else exactly equally good; as such, they seem to qualify as imprecisely equally good. Insofar as they are imprecisely equally good, there is some small improvement, *s*, such that improving one of the options with *s* does *not* make the improved option strictly better than the alternative. In particular, the small improvement of making the Pearl Jasmine slightly more fragrant does not make it strictly better than the Sumatra Gold. Still, insofar as the Pearl Jasmine and the Sumatra Gold are roughly equally good, it seems like an option that is quite a lot better than one of the options will be better than the other option too; for, if an option that is quite a lot better than one of the options is not better than the other, the original options must not have been close in value. If, for example, a glass of Dom Pérignon is quite a lot better than the cup of Pearl Jasmine but not better than the cup of Sumatra Gold, it seems false to say that the cup of Pearl Jasmine and the cup of Sumatra Gold are roughly equally good.

4

Now consider the question of whether all cases of parity must be cases of rough equality and, more precisely, cases of imprecise equality. (Note that settling this question does not necessarily settle the question of whether there are any cases of incomparability, which I here leave aside.) Given

the (partial) description of rough equality provided earlier – according to which "X is roughly equal to Y" implies that, were another option, W, considerably better than one of the options, it would be better than the other option too – there could, I contend, be cases of parity that are not cases of rough equality and so not cases of imprecise equality. In particular, there could be cases in which the options are intuitively characterizable as on a par, and yet there is another option that is quite a lot better than one of the options but not better than the other.

Consider the following case:

> Suppose that one is both compassionately and philosophically inclined, and let A [and] B… be defined as follows:
>
> A = dedicating the summer to compassionately engaging with others by volunteering at a homeless shelter
>
> B = dedicating the summer to promoting one's philosophical development by completing a book manuscript
>
> …
>
> Taking into account the context of choice, including any relevant features of the choosing agent, suppose that (from the point of view of practical reason), A is not better than B and B is not better than A.[10]

Now consider an improved version of the compassionate option, call it A*, in which one *also* provides a temporary foster home for a needy child. A* does not seem roughly as good as A, but quite a lot better. Still, like A and B, A* and B are so different, it seems like they too might be such that neither is better than the other (from the point of view of practical reason). But then we have, in A*, an option that is quite a lot better than A but not better than B. This suggests that even though A and B seem like they are on a par, they are not imprecisely equally good.

Importantly, this case is not meant to provide decisive support for the view that there can be options that are on a par but not imprecisely equally good. However, once it is supplemented with the reasoning that I will provide for the view that we can make sense of two options being in the same neighborhood without being close in the way that options that are roughly equally good are, it will, I hope, be clear that we should not assume that all cases of parity are cases of imprecise equality.[11]

Interestingly, Chang provides a characterization of options that are on a par that allows for the possibility of options that are on a par but not necessarily imprecisely equally good. In working up to this characterization, Chang (2016: 194) suggests that

> we understand value relations in terms of *evaluative differences* between items [with respect to what matters in the choice situation at issue, where] evaluative differences can be individuated along

two dimensions: (1) *bias* or *direction*, that is, whether the difference favours an option or "points to" one of them, and (2) *magnitude*, that is, whether the difference has some extent and is therefore nonzero.

We can then understand cases of parity as follows: "If A and B are on a par, then their evaluative difference does not favour one alternative over the other – it has no direction – but it nevertheless has magnitude" (Chang 2016: 195). Cases of parity thus differ from cases in which A is better than B and cases in which B is better than A, since, in cases of the latter two sorts, the evaluative difference between the options is biased toward one of the options. Cases of parity also differ from cases in which A and B are equally good since in cases of the latter sort, the evaluative difference between the two options is not only not biased, but also has zero magnitude. Moreover, there is nothing in the characterization of parity under consideration that requires cases of parity to be cases of imprecise equality. In particular, the characterization does not require that the evaluative difference between the options not be large.

I will not attempt to decisively weigh in on whether Chang's controversial characterization of parity is revealing or confounding in relation to the familiar cases of parity that Chang focuses on and that fit with the idea that options that are on a par, though not rankable in relation to one another, are in the same neighborhood in terms of their overall value relative to what matters in the situation. What must be noted, given my purposes, is that the characterization gives no answer to the question that I seek to address (regarding the notion of parity that I am concerned with in this chapter and that captures the key features of the familiar cases). That question is as follows: Allowing, provisionally, that there could be options that are on a par but not roughly equally good (and so not imprecisely equally good), in what sense would such options be in the same neighborhood in terms of their overall value relative to what matters in the situation?

To answer this question, it helps to think about grading systems and how the extent to which they can be sensibly refined depends on how simple and similar the options compared are.[12] If, for example, one is evaluating the books in one's office on the basis of their suitability to figure in some impromptu resistance training, and this is purely a function of their weight, one may be able to not only divide them up into broad classes that reflect "different grades or levels of quality" relative to what matters,[13] such as "marginal," "good," "very good," or "excellent," but also to rank them and assign them much more precise scores based on their weight. If, however, one is evaluating books from a variety of very different disciplines for a dissertation prize, it might be misguided to suppose that, with enough time and insight, one could rank each dissertation

in relation to every other dissertation.[14] Instead, it might be that properly evaluating the dissertations involves sorting them into "grade classes," such as "marginal," "good," "very good," and "excellent," and allowing that, while options that are similar enough might be rankable in relation to each other, options that are very different might not be. Options of the latter sort might only be comparable as in the same class or as on a par. Since the grade classes are, by hypothesis, inevitably quite broad, two options in the same class can be such that one of the options is considerably better than the other. It therefore cannot be assumed that options in the same grade class qualify as roughly equally good.

Suppose, for instance, that two physics dissertations fall in the class of good books. One of the physics dissertations, P*, might be considerably better than the other, P, just as the improved compassionate option in the previously described summer plan case is considerably better than the original compassionate option, though both options are good. Were all the dissertations quite similar to one another, it might be possible to subdivide the good dissertations into two distinct sets corresponding to two more refined evaluative classes (that do not overlap and together contain all and only the good dissertations), with one set containing the better of the two good physics dissertations, P*, and the other set containing the worse of the two good physics dissertations, P, and the dissertations in the former set all being better than the dissertations in the latter set. This need not, however, be possible if some of the good dissertations (which might include, in addition to the two physics dissertations, a creative writing dissertation) are very different from each other and not comparable as one better than the other or as exactly equally good. If, for example, the good creative writing dissertation, C, cannot be ranked in relation to either of the two good physics dissertations, then there is no way of subdividing the good dissertations into two distinct sets with the dissertations in the set containing P* all being better than the dissertations in the set containing P. The evaluative neighborhoods (or grade classes) that options that are on a par belong to are thus sometimes inevitably quite wide-ranging – wide-ranging enough to contain options that are not sufficiently close in value to qualify as roughly equally good, including (1) some pairs of similar options – such as P* and P in the case at hand – that are such that one of the two options is considerably better than the other, as well as (2) some pairs of options – such as P and C in the case at hand – that are not rankable in relation to each other and do not qualify as roughly equal in value because there is another option – P* in the case at hand – that is considerably better than one of the original options but not better than the other. As such, it does not follow from the fact that options that are on a par share the same evaluative neighborhood that they are roughly equally good, even if the neighborhood cannot be subdivided into two narrower evaluative classes.

Notably, one might consider weakening the idea of rough equality so that pairs of options – including pairs of similar options that are such that one of the options is considerably better than the other – can qualify as roughly equal if the options share a neighborhood that, due to some of the options in the neighborhood not being rankable in relation to each other, cannot be subdivided into two neighborhoods that figure as distinct grade classes (wherein all the options in one of the grade classes are better than all the options in the other grade class). But this just leaves us without a label for the less inclusive subset of cases we normally think of as cases of rough equality and papers over the real distinction between cases of parity that fit with the strict notion of rough equality and cases of parity that fit only with the weaker revisionary notion. Whatever labels we settle on, it is important to recognize that cases of parity can fail to qualify as cases of rough equality in the more standard, stricter sense of "rough equality."

Significantly, if one is receptive to the possibility of parity but skeptical about Chang's idea that, if A and B are on a par, then their evaluative difference has "magnitude" but no "bias," one can, in light of the preceding reasoning, refrain from taking Chang's idea on board and still have something illuminating to say about what it is for two options to be on a par (whether or not they are imprecisely equally good): if A and B are on a par, then, though not rankable in relation to one another, they share a class in a grading system that can be used to grade A and B and that is, in relation to this task, as refined as can be (assuming distinct, non-overlapping grade classes). Alternatively, those sympathetic to Chang's idea might attempt to use the distinction between cases of parity that are cases of imprecise equality and cases of parity that are not cases of imprecise equality to try to illuminate talk of magnitude and (lack of) bias in cases of parity, emphasizing that cases of parity can involve options that are either more or less close in value, but that either way, reason does not favor one option over the other. Of course, strictly speaking, the assumption that X and Y are close in value does not imply that X and Y have a difference in value with some zero or nonzero magnitude, but rather that, were some third alternative, W, considerably better than one of the original options, it would also be better than the other. Still, talk of magnitude without bias might seem more promising than when first encountered.

In the end, what I really want to emphasize is that understanding parity in accordance with the preceding discussion captures the key features of (the notion of) parity (that I am here concerned with) – namely, that options that are on a par are not rankable in relation to each other, but are in the same neighborhood in terms of their overall value relative to what matters in the situation; moreover, it allows that there are instances of parity that are not instances of imprecise equality (in the sense I describe early in the chapter). The idea that options that are on a par might not be imprecisely equally good has been previously embraced; it has not,

however, been previously explained in relation to the central idea that options that are on a par are in the same neighborhood (in terms of their overall value relative to what matters in the situation at issue). Given that options that are on a par are in the same neighborhood, it is natural to conclude that they must be close in value, and so roughly equally good. It is, I hope, illuminating to see why this need not be so.

Acknowledgments

Notes

1 The quoted description of the trichotomy thesis is from Chang (2002: 661). For reasons that will be discussed later, Chang rejects the thesis.
2 The possibility of options being "imprecisely equally good" is discussed by, for example, Parfit (2016). Earlier related discussion on "rough comparability" can be found in Parfit (1984: 431). Notably, I am *not* using "roughly equally good" and "imprecisely equally good' as interchangeable, as is sometimes done. In particular, as I am using "roughly equally good," two options that are roughly equally good can fail to be imprecisely equally good since they might be rankable in relation to one another.
3 Although the (partial) description of options that are on a par as "in the same neighborhood" is borrowed from Chang (2016), Chang does not cast all cases of parity as cases of imprecise equality. More on this later in the chapter. Note that I will put aside, as not the conception of parity at issue in this chapter, the "permissible-attitudes" conception of parity. According to the permissible-attitudes conception, two options "x and y are *on a par* if and only if it is rationally permissible to prefer x to y and also rationally permissible to prefer y to x" (Rabinowicz 2008: 30; see, relatedly, Gert 2004 and Rabinowicz 2012). Though defensible for certain purposes (at least when suitably qualified or refined), the permissible-attitudes conception of parity allows that two options can qualify as on a par even if the choosing agent rationally permissibly strongly prefers one of them. In particular, if it is rationally permissible to prefer getting a slice of chocolate cake over getting a slice of apple pie, and it is rationally permissible to prefer getting a slice of apple pie over getting a slice of chocolate cake, then this conception of parity counts "getting a slice of chocolate cake" and "getting a slice of apple pie" as on a par, even if I am the one choosing, and I rationally permissibly strongly prefer getting a slice of chocolate cake. It thus seems unsuitable with respect to an inquiry concerning the evaluation of options relative to all that matters in the case at hand, *including*, in cases involving multiple rationally permissible preferences, the fit between the options and the rationally permissible preferences of the choosing agent. In any case, my interest here is in conceptions

of parity that are tied to the familiar idea that if two options are on a par, then they are in the same neighborhood (in terms of their overall value relative to what matters in the situation); the permissible-attitudes conception of parity is not tied to this familiar idea, at least not directly enough to clearly shed light on the question that, as I will presently explain, I will be seeking to address – namely, "Could two options that are not rankable in relation to one another be in the same neighborhood (in terms of how valuable they are overall relative to what matters in the situation at hand) if they are not roughly equally good, and, if so, how could this be realized?" See, relatedly, note 5.

4 This conclusion fits neatly with Erik Carlson's suggestion that "parity should imply a certain degree of similarity, with respect to value," and "hence, if *a* is on a par with *b*, and *c* is *much* better than *a*, then *c* must be better than *b*, as well"; see Carlson (2010: 125) and keep in mind that, as indicated in section 3, for two options to be close in value, it must be that were another option, W, considerably (i.e., quite a lot rather than just a little) better than one of the original options, it would also be better than the other.

5 A version of the contrast between the permissible-attitudes approach to understanding parity, discussed in note 3, and the rough-equality approach can be found in Qizilbash (2018). I do not mean to suggest that all viable views of parity fit one or the other approach (given my description of the approaches). To the contrary, my aim is to defend a view that suggests that parity can be helpfully understood in a way that does not fit with either of these two approaches. Ruth Chang provides a characterization that also arguably occupies a different space, though it is, as will become apparent when I discuss the characterization later, one that I do not want to commit to, in part because it is not clear to me that it is defensible and in part because it does not answer the question at the end of note 3 that I will soon turn to and then attempt to answer in the rest of the chapter.

6 See Andreou, *Choosing Well* (unpublished manuscript). I thank the editors for inviting me to share this more complete presentation of my view.

7 See Chang (1997: section III.7) for a discussion of the small-improvement argument in the context of an introduction to philosophical issues associated with the idea of incomparability. See also, for example, de Sousa (1974) and Raz (1986: chapter 13) for earlier variations of the argument. (A particularly interesting complication related to de Sousa's variation is flagged in note 8 below.) As Johan Gustafsson and Nicolas Espinoza (2010: 754) note, "there are both preferential and axiological versions of the argument." Although Gustafsson and Espinoza put forward what might appear to be a decisive argument against the preferential version, a rebuttal to their challenge is provided in Carlson (2011). Both preferential and axiological versions of the argument continue to play a pivotal role in discussions concerning comparability.

8 Interestingly, in de Sousa's variation of this argument, wherein the agent is supposed to be torn between "keeping her virtue" and accepting a "tempt[ing]" monetary reward for "losing it" (1974: 544–545), the improvement of the monetary reward does not seem small ($1,500 versus $1,000). It might be suggested that, although the difference seems large, it might qualify as small if $1,000 will make a big difference in the agent's life but an extra $500 on top of that will not be very significant. This complication seems like a dispensable distraction relative to de Sousa's reasoning, which only requires an option that is clearly better than $1,000, but I find the complication quite suggestive. As will become apparent, considering more than small improvements in small-improvement type arguments is, I think, a good move to make

in terms of helpfully complicating debate about parity, rough equality, and incomparability. According to the view I will develop, even given moderate-to-large-improvement scenarios that do not enable a ranking of the options, the options may still be comparable as on a par.

9 See, relatedly, Qizilbash (2018) regarding the "'mark of parity' on the rough equality view."

10 This case, with the addition of an improved option, A*, that I will get to shortly, is drawn from my "Incomparability and the Huge Improvement Arguments" (2021), in which the case (which I first presented at the Tanner Humanities Center in 2016) figures in an assessment of a variation of the small-improvement argument for incomparability – namely, the huge-improvement argument for incomparability. My tentative suggestion there is that the case raises problems for the huge-improvement argument for incomparability. See, relatedly, Martijn Boot's discussion of the "large improvement phenomenon" (2017), wherein options A and B are such that A is not better than B, B is not better than A, and yet A and B cannot qualify as equally good, not even roughly equally good, because A*, which is a considerably improved version of A, is better than A but not better than B. Boot (2017) counts large-improvement cases as cases involving incomparable options (though he uses the term "incompletely comparable" based, in part, on the idea that, although the options are "overall incomparable" – i.e., incomparable with respect to all the "relevant values taken together" – they are "usually comparable with respect to the [relevant] values [taken] separately" (316)). A notable complication that I will flag but not delve into here is that, for Boot (2017), all the evaluations in play must be "detached from…subjective preference" (317); as such, they may fail to capture the overall evaluation of the options relative to *all* that matters in the case at hand from the point of view of practical reason, which can include the fit between the options and the taste or preferences of the choosing agent (Andreou, 2021).

11 Notably, my reasoning later in the chapter suggests that, although, in certain cases, parity can exhibit a persistence in the face of considerable improvements that speaks against counting the options as imprecisely equally good, the persistence need not be so radical that the cases should be understood as cases of incomparability rather than cases of parity. For some interesting discussion regarding persistent unrankability, see Handfield and Rabinowicz (2018) and Herlitz (2020).

12 In this paragraph and the next, I build on my suggestion, in Andreou (2015), that "as we move away from toy cases and introduce more subtleties, the categories used in the comparison of options may involve thinner concepts with broad application. Where the most refined categories in play are extremely broad, there is a great deal of room for parity to figure as highly pervasive" (21). Notably, and as I explain in Andreou (2015), like Nien-hê Hsieh (2005), "I associate parity with the possibility of options 'clumping' together…, but my view concerning this possibility and its implications differs substantially from Hsieh's…. According to Hsieh's understanding of parity and clumpiness, parity amounts to equality" (Andreou 2015: 15, note 15).

13 The quoted phrase is borrowed from Andreou (2020b).

14 A variation of this example and of my discussion of it in the remainder of this paragraph and the next two paragraphs appears in my *Choosing Well* (unpublished manuscript).

References

Andreou, C. (2015), 'Parity, Comparability, and Choice', *Journal of Philosophy* CXII: 5–22.

Andreou, C. (2020a), 'Empowering Rationality: Rethinking the Limits of Rational Choice', *American Philosophical Quarterly* 57: 105–116.

Andreou, C. (2020b), 'In a Different League: Intransitivity, Betterness, and League-Based Satisficing', in A. Sauchelli (ed.), *Derek Parfit's Reasons and Persons: An Introduction and Critical Inquiry* (Routledge).

Andreou, C. (2021), 'Incomparability and the Huge Improvement Arguments', *American Philosophical Quarterly* 58/4: 307–318.

Boot, M. (2017), 'Problems of Incommensurability', *Social Theory and Practice* 43: 313–342.

Carlson, E. (2010), 'Parity Demystified', *Theoria* 76: 119–128.

Carlson, E. (2011). 'The Small-Improvement Argument Rescued', *Philosophical Quarterly* 61: 171–174.

Chang, R. (1997), 'Introduction', in R. Chang (ed.), *Incommensurability, Incomparability, and Practical Reason* (Harvard University Press).

Chang, R. (2002), 'The Possibility of Parity', *Ethics* CXII/4: 659–688.

Chang, R. (2016), 'Parity, Imprecise Comparability, and the Repugnant Conclusion', *Theoria* 82: 182–214.

de Sousa, R. (1974), 'The Good and the True', *Mind* 84: 534–551.

Gert, J. (2004), 'Value and Parity', *Ethics* 114: 492–510.

Gustafsson, J., and Espinoza, N. (2010), 'Conflicting Reasons in the Small-Improvement Argument', *Philosophical Quarterly* 60: 754–763.

Handfield, T., and Rabinowicz, N. (2018), 'Incommensurability and Vagueness in Spectrum Arguments', *Philosophical Studies* 175: 2373–2387.

Herlitz, A. (2020), 'Spectrum Arguments, Parity and Persistency', *Theoria* 86: 463–481.

Hsieh, N. (2005), 'Equality, Clumpiness and Incomparability', *Utilitas* xvii/2: 180–204.

Parfit, D. (1984), *Reasons and Persons* (Clarendon Press).

Parfit, D. (2016), 'Can We Avoid the Repugnant Conclusion?', *Theoria* 82: 110–127.

Qizilbash, M. (2018), 'On Parity and the Intuition of Neutrality', *Economics and Philosophy* 34: 87–108.

Rabinowicz, W. (2008), 'Value Relations', *Theoria* 74: 18–49.

Rabinowicz, W. (2012), 'Value Relations Revisited', *Economics and Philosophy* 28: 133–164.

Raz, J. (1986), *The Morality of Freedom* (Clarendon Press).

Part II

Incommensurability and Ethical Theory

4 On "Incommensurability," "Discontinuity," and the Repugnant Conclusion

"Imprecise Equality" or Vagueness?

Mozaffar Qizilbash

1 Preliminaries: Two Kinds of "Incommensurability"[1]

In his celebrated discussion of "incommensurability" in *Well-Being: Its Meaning, Measurement and Moral Importance*, James Griffin (1986: chapter 5)[2] distinguishes a number of distinct senses of "incommensurability."[3] Of these, he suggests that the strongest is "incomparability." Griffin (1986: 83) argues that – at least in the realm of prudential value – there is no "incomparability." "Incomparability" (of the relevant sort) exists if, as between two options (or realizations of prudential value), one is not better than the other and they are not of equal value either. Griffin argues that some candidate cases of "incomparability" may instead be cases of "rough equality" and since "rough equality" is a form of equality, these are not cases of "incomparability." Griffin's notion of "rough equality" is a close relation to Derek Parfit's idea of "rough comparability" in *Reasons and Persons* (Parfit 1984: 431) and his subsequent discussion of "imprecise equality" (Parfit 2016). At the same time, Griffin discusses "discontinuity," which he treats as a "weaker" form of "incommensurability." In cases of "discontinuity," the claim is that, as between two prudential values, *A* and *B*, either enough of value *A* outranks any amount of value *B*; or once one has a certain amount of *B*, any amount of *A* outranks any further amount of *B*. Griffin (1986: 85) accepts that there are cases of this sort but he argues that since they involve ranking, there is no incomparability in these either. Furthermore, in an endnote, where he discusses Parfit's views, Griffin (1986: 338–340) conjectures that "it would seem likely that this incommensurability in prudential cases would get transferred to interpersonal calculation" of the sort that is relevant to population ethics.

The common feature of the candidate examples of "rough equality" and "discontinuity" (in prudential value) is that there is typically a difference of kind or type (e.g. between poetry and prose, or between different categories of prose), or some form of qualitative difference (e.g. in the pleasure or pain) involved in the items or options being compared. While

there are various definitions of "incommensurability" (see Griffin 1986; Raz 1986a, b; and Chang 1997), in this chapter, I take "incommensurability" to cover distinct cases with this common feature. As a result, I follow Griffin (1986: 77) and use "incommensurability" in a "way that allows different senses" and to cover cases – such as cases of "discontinuity" and "rough equality" – where items are comparable in terms of value (see Griffin, 1986: 76 and 97).[4]

There have been two dominant ways of modeling "rough equality" in the subsequent literature. The first focuses on Griffin's suggestion that "rough equality" is a form of vagueness or "indistinctness" and on the possibility of a *vague ordering* (see Griffin 1986: 96). This possibility has been most fully explored and developed by John Broome (1997, 2000, 2004, and 2012; see also Griffin 2000).[5] The second reading follows Parfit and Griffin in supposing that there are two kinds of "rough equality" or "rough comparability." The first sort involves an inability to make precise discriminations or ignorance. Both Parfit and Griffin argue that there are other cases where the problem of comparability is not of this sort, but rather in the nature of the world or, put another way, in the nature of the *relation* between objects of value. It is this second form of "rough equality" that links this literature to the discussion of a distinct relation aside from the standard "trichotomy" of "better than," "worse than," and "exactly equal in value." The relation of "parity" as it has been explained by Ruth Chang (1997, 2002a, 2002b, 2005, 2016, and forthcoming) and the relation "in the same league as" that Parfit (1984: 431) mentions are also distinct relations and close relatives of "rough equality" of this sort. In this chapter, I follow earlier work (see Qizilbash 2002, 2005a, 2007a, b; and 2018) and use "parity" as a term of art to refer to "rough equality" as a distinct relation. This use of "parity" within a nonstandard set of value relations extended to allow for "rough equality" may, in some respects, be inconsistent with Chang's account of "parity" because Chang (2016; and forthcoming) distinguishes her account from the views of both Parfit and Griffin.

The form of "incommensurability" that Griffin himself invokes in the context of population ethics is "discontinuity" rather than "rough equality." Furthermore, for Griffin's remarks about "discontinuity" to be plausible, they should arguably allow for vagueness (see Qizilbash 2005a). By contrast, in his later work, Parfit defends a form of "discontinuity" (or "strong lexical view" in his terms[6]) and invokes "imprecise equality" rather than conceptual vagueness in his discussion of the "Repugnant Conclusion." This may be another case where there is a question about whether the best way to address these problems involves invoking a determinate but nonstandard value relation such as parity *or* vagueness about which value relation holds (see Broome 1997; Rabinowicz 2009a; and Qizilbash 2012 *inter alia*). If we set aside other potential ways of avoiding the Repugnant Conclusion, and there is a choice between

"parity" – interpreted in terms of rough or imprecise equality – and vagueness, which should we favor?

The chapter is organized as follows: in Section 2, I summarize the account of value relations; in Section 3, I discuss examples of "discontinuity" and suggest that vagueness is relevant to these; in Section 4, I extend the discussion to population ethics and "discontinuity" combined with either parity as "imprecise equality" or vagueness in the context of the Repugnant Conclusion; and Section 5 concludes.

2 "Rough Equality," "Rough Comparability," and "Parity"

I begin with an account of value relations (set out more fully in Qizilbash 2005a) which allows for "rough equality" as a distinct relation. I write "G" for a primitive relation which means "better than or exactly as good as." This is the "conventional" use of "at least as good as." "Better than" (B) and "exactly as good as" (E) are then defined as follows:

Definition 1: xBy if and only if xGy and not (yGx)
Definition 2: xEy if and only if xGy and yGx

I assume that G is a quasi-ordering so that it is transitive and reflexive but not necessarily complete. That is, I assume that for all x in the set of alternatives X, xGx (reflexivity), and for all x, y, and z in X: if (xGy and yGz), then xGz (transitivity of G). Following standard results about quasi-orderings (see Sen 1979: 10), transitivity of G implies transitivity of B: for all x, y, z in X, if (xBy and yBz), then xBz. It is also easy to show that B is irreflexive – so that for all x in X, not (xBx) and asymmetric – so that for all x, y in X, if xBy then not (yBx) (see Qizilbash 2005a: theorem 1). Completeness of G requires that for all nonidentical x and y in X, xGy or yGx.[7] If it holds, we can say that all alternatives are *commensurate* so that either one is better than the other, or they are exactly as good. If instead, like Griffin, one believes that (in the prudential realm) all alternatives are comparable – even if they are not all commensurate in the sense defined here – one would need to allow for an alternative comparative relation between options because they may be comparable without being commensurate. In Griffin's terms, alternatives that are comparable but not commensurate are "roughly equal" in value. Because "rough equality" of the relevant sort involves a determinate (fourth) value relation between alternatives rather than vagueness, and because "rough equality" can also be understood as a form of vagueness I use "on a par with" (P) as a term of art for this relation. I also write "C" for a primitive and nonstandard relation meaning "comparable with." While this relation is primitive, it is important that comparability is, in a sense, broader than commensurateness as it is defined here. In particular, if all items are commensurate, they are all comparable, but if they are all

comparable it does not necessarily follow that they are all commensurate. Furthermore, the primitive notion of "comparability" allows for the possibility that there are distinct cases where we might say that (while one is not better than the other), items are in some sense "equal in value." They may be "roughly equal" in value or "in the same league" – even if they are not exactly as good. To allow for these cases, we can define "parity" as follows:

Definition 3: xPy if and only if xCy and not (xGy or yGx).

Finally, there is a non-conventional sense of "at least as good as" (A) which can be defined as follows:

Definition 4: xAy if and only if xCy and not (yBx).

This relation is arguably invoked in Derek Parfit's discussion of "rough comparability." Parfit (1984: 431) discusses three candidates for a literary prize: a novelist and two poets. He writes,

> We might claim, of the Novelist and the First Poet, that neither is worse than the other. This would not be claiming that these two cannot be compared. It would be asserting rough comparability. There are many poets who would be worse candidates than this Novelist, and many Novelists who would be worse candidates than the First Poet. We are claiming, of these two, that something important can be said about their respective merits. Neither is worse than the other. They are in the same league.
>
> (Parfit 1984: 431)

It is in this sort of case that Parfit thinks that "rough comparability" is not merely a matter of ignorance. As he puts it, "The rough comparability is here *intrinsic*, not the result of ignorance" (Parfit 1984: 431). Because in the example he discusses here, neither is worse than the other, but they are not exactly as good either, in the system I have outlined, they must be on a par. Indeed, to this degree, Parfit's uses of "in the same league" and "roughly comparable with" track the relation "on a par with" as I have defined it.

If we allow for "parity" understood as a distinct relation of "rough equality" or "rough comparability" we allow for incompleteness of G because in cases of parity neither xGy nor yGx. These cases involve alternatives that are not commensurate and are, in this sense, *incommensurate*[8] even though – in terms of the system of value relations outlined here – they do not involve any form of incomparability. Furthermore, while in this relation system E is transitive, P is not (see Qizilbash 2002: 144–147). Consider again Parfit's example of two poets and a novelist. On this relation system, this is an example of parity. While both poets are on a par with the novelist – since while they are not exactly equal in

value and one is not worse than the other, yet they are comparable and – the second poet is better than the first. Nonetheless, while the second poet is on a par with the novelist and the novelist is on a par with the first poet, it is not true that the second poet is on a par with the first. P is non-transitive. And because the unconventional "at least as good as" relation can hold when alternatives are on a par, A is also non-transitive (see also Qizilbash, 2005a and 2007a). Here again, this relation system tracks Parfit's discussion. In the terms he himself uses, Parfit (1984: 431) believes that "[w]hen there is only rough comparability, *not worse than* is not a transitive relation." Parfit's use of "not worse than" here arguably tracks relation A.

This account of value relations treats "comparable with" as a primitive relation. How might one distinguish cases of incomparability and parity if one accepts this account? If one treats the nonstandard relation of parity as a form of "rough" or "imprecise" equality, then to say that items are comparable requires that some form of equality must hold if one is not better than another. If they are not precisely equal, they must be roughly equal in value. If items are roughly but not precisely equal, then while slight changes in the value of options may not make one better than the other, any significant change in the value of one will tip the balance. While this "mark of parity" is not part of, or required by, the formal system of relations outlined here, it follows from interpreting parity in terms of "rough equality" (see Qizilbash, 2002, 2005a; and 2018).[9] This mark also distinguishes parity from incomparability: in cases of incomparability, while neither of two items is better than the other, even some significant change in the value of one does not make it better or worse than the other. On accounts where "on a par with" is interpreted in terms of "in the same neighborhood as" or "in the same league as" – this mark may fail to hold because there are arguably cases where items that are not close and, to this degree, not "roughly equal" in value while being in the same neighborhood or league (see Andreou 2015; and 2021a). These accounts might, nonetheless, be squared with the mark of parity as it is defined here, if, to qualify as a "significant" change in value, it must be large enough to "carry" one item from one "neighborhood" of value to another. Yet if "neighborhoods" are very large, the intuition that parity is "rough equality" loses force.[10]

While the "mark of parity" is intuitively compelling if items are roughly equal in value, if one adopts the previous definitions but does not accept that parity is a form of "rough equality," one might adopt an alternative way of distinguishing "parity" and "incomparability." On this weaker version of the "mark of parity" when options are on a par, while some slight change in value may not tilt the balance in favor of one of them, either some significant increase in the value of one of the items will make it better *or* some significant reduction in the value of one of the items will make it worse. Then cases of "incomparability" can be defined so that while two items are incommensurate, no change in the value of one makes it better or worse than the other (Qizilbash 2018: 91–92).

This alternative way of characterizing the distinction between "parity" and "incomparability" might fit better with an account of parity understood as belonging to the same league or neighborhood, which may allow for large value neighborhoods or leagues.[11]

3 "Discontinuity" and Vagueness

Now consider "discontinuity." Here Griffin's claim is that, as between two (prudential) values, *A* and *B*, either (i) enough of value *A* outranks any amount of value *B*, or (ii) once one has a certain amount of *B*, any amount of *A* outranks any further amount of *B*. Griffin (1986: 85) accepts that there are cases of this sort, especially if one focuses on specific realizations of value – "value-tokens" – rather than values in the abstract – "value-types." For example, he suggests that

> it is more plausible that, say, fifty years of life at a very high level of well-being – say, the level which makes possible satisfying personal relations, some understanding of what makes life worth while, appreciation of great beauty, the chance to accomplish something with one's life – outranks any number of years at the level just barely worth living.

But he adds that "we should not expect to find sharp discontinuities at extreme ends of the scale." He goes on:

> Fifty years at a very high level – say, with the very best Rembrandts, Vermeers, and de Hoochs – might be outranked by fifty-five years at a slightly lower level – no Rembrandts, Vermeers, and de Hoochs, but the rest of the Dutch school. And then fifty-five years at that level might be outranked by sixty years at a slightly lower level – no Dutch School but a lot more of the nineteenth-century revival of the Dutch School. And so on, step-by-step, until, it seems, we must eventually reach the point where the original fifty years would be outranked by a sufficiently large number of years of life just barely worth living.
> (Griffin 1986: 86)

This "beauty-kitsch example" is a form of what is now known as "spectrum" – or, somewhat misleadingly as "continuum" – examples. In each step along the spectrum, there is a small reduction in the quality of enjoyment, which is traded off against some quantity – in this case, some number of years of enjoyment of beauty. There is, in Griffin's view, no problem of comparability in this example. He adds,

> This step-by-step approach seems irresistible. Yet it presents us with two embarrassments: a Sorites Paradox and a slippery slope. If we

take enough pebbles away from a heap, it ceases to be a heap. But then, since one pebble more or less could never make the difference between its being a heap or not, if we remove them singly it can never cease being a heap. Similarly, one might argue, with "appreciation of beauty" or "deep personal relations"; and slow, step-by-step changes we can never lose the appreciation of beauty or deep loving relations.

But we obviously can – I want to set the Sorites problem to the side. We can indeed reach a point, with the series of subtractions, that we just saw (first Rembrandt, Vermeer, and de Hooch, then the whole seventeenth century Dutch School …), at which we lost the appreciation of beauty. We may have the kicks of kitsch but they are different.

(Griffin 1986: 87)

In this example, at a very high level, a slight reduction of quality does not lead to a change in the *kind* of life lived. Yet if enough is lost in terms of quality so that there is a change in the kind of appreciation or enjoyment, greater amounts of quantity cannot compensate for the loss of quality.

As we reach the bottom of the spectrum, Griffin worries, nonetheless, that enough of the kicks of kitsch must compensate for the loss of quantities of the other prudential values. At the bottom, when "we also lose deep personal relations, accomplishment, and all the rest of the substantial values in life, then by parity of reasoning, since what is left is, though slight, not nothing, enough life with this residue must be more valuable" (Griffin 1986: 87). To resist this judgment Griffin invokes the idea of a "basic preference" and his informed desire view of welfare – on which well-being consists in the fulfillment of fully informed desires (see Griffin 1986: part 1):

We do seem, when informed, to rank a certain amount of life at a very high level above any amount of life at a very low level. And if we do, it is likely to be a basic preference; that is, it could not be based upon other judgements about the amount of value in the objects being ranked. Nor need there be any extreme perfectionism at work in this preference; those who held this preference could also be willing to sacrifice the best Rembrandt, Vermeer, or de Hooch, for all the various, though lesser, achievements of the Dutch school.

(Griffin 1986: 87)

Griffin thinks that informed preference favors *some* amount of the enjoyment of Rembrandt, Vermeer, and so on over the enjoyment of *any* amount of the kicks of kitsch. This is the "discontinuity" claim. And because Griffin suggests that in this sort of example there is a Sorites at work, one would expect to find an imprecise borderline as one moves

down the spectrum – in this case a rough threshold where enough genuine beauty is lost and what remains are the kicks of kitsch.

Griffin does not develop the claim that there is a Sorites in this example. And there is ongoing debate about whether cases of this sort involve a Sorites (see Temkin 1996 and 2012; Qizilbash 2005b; and Pummer 2018 and forthcoming). If it can be shown that there is a Sorites or Sorites-like structure in such cases that would support the idea that they involve vagueness. Yet, there are other "symptoms" of vagueness, notably whether or not there is a predicate that admits borderline cases or an imprecise borderline. To this degree, we need not settle debates about whether or how spectrum examples involve a Sorites or Sorites-like structure to believe that there is some form of vagueness involved. As long as there is an imprecise borderline or if there are borderline cases – for example, in the "beauty-kitsch" example, borderline cases of "beauty" and an imprecise borderline between cases of "beauty" and "kitsch" – theories of vagueness come into play (see Qizilbash, 2005b; and Knapp, 2007). Similarly, in Griffin's discussion of this example, we start with a life at a "very high" level, and it can certainly be argued that, like "tall" and "bald," "high" (or "very high") is a vague predicate since there are borderline cases of "high" (or "very high") and an imprecise borderline between cases that are and are not high (or "very high").

The beauty-kitsch example is a close relative of what Parfit calls the "single-life analogue of the Repugnant Conclusion" (Parfit 1986: 160). He offers the following example: "Suppose that I can choose between two futures. I could live for another 100 years, all of an extremely high quality. Call this *the Century of Ecstasy*. I could instead live for ever, with a life that would always be barely worth living. Though there would be nothing bad in this life, the only good things would be muzak and potatoes. Call this *the Drab Eternity*" (Parfit 1986: 160).

Commenting on these futures, Parfit confesses, "I believe that, of these two, the Century of Ecstasy would give me a better future. And this is the future that I would prefer. Many people would have the same belief, and preference" (Parfit 1986: 160). This belief and preference are consistent with Griffin's suggestion that – for someone whose preferences are (fully) informed – enough years enjoying the best of the Dutch School are preferred to any amount of enjoyment of the kicks of kitsch. It is, in Griffin's terms, an example of "discontinuity." But in the discussion immediately preceding this example, Parfit discusses population paradoxes and, in a note, he rules out the possibility that in cases of this sort there is a Sorites Paradox. He writes,

> It may be objected that my argument is like what are called *Sorites Arguments*, which are known to lead to false conclusions. Suppose we assume that removing any single grain from a heap cannot turn a heap of sand into something that is not a heap. It can then be argued

that even if we remove every single grain, we must still have a heap.
… A Sorites Argument appeals to a series of steps, each of which is
assumed to *make no difference*.

(Parfit 1986: 160)

Parfit makes this argument in the context of the "Mere Addition
Paradox," but I shall explain it in terms of the beauty-kitsch example.
Parfit suggests that "[a] Sorites Argument appeals to a series of steps,
each of which is assumed to *make no difference*." By contrast, the step-
by-step sequence, which involves slight reductions in the quality of
enjoyment with some increase in the time spent in enjoyment, involves
pairwise comparisons between options where in each step one is *better*.
He concludes that "[t]he objections to Sorites Arguments are therefore
irrelevant" (Parfit 1986: 160).

   This line of argument has been made in the context of a variety of spec-
trum examples (see, for example, Temkin 1996; and 2012), and I shall
not evaluate it here. Nonetheless, it is worth noting that in the beauty-
kitsch example, Griffin is also claiming that at the top of the spectrum,
55 years without the very best of the Dutch School (but with enjoyment
of slightly lesser works) is *better than* 50 years of the best of the Dutch
School. It is somewhat further down the spectrum that, on Griffin's read-
ing, some loss of quality may no longer result in the same *kind* of appre-
ciation. And in Parfit's example of the Drab Eternity and the Century of
Ecstasy, the equivalent to Griffin's logic would be some point further in
the slide down from ecstasy to "muzak and potatoes." In this context,
Parfit invokes J. S. Mill's discussion of "higher" and "lower" pleasures:

   I claim that, though each day of the Drab Eternity would be worth
   living, the Century of Ecstasy would give me a better life. This is like
   Mill's claim about the "difference in quality" between human and
   pig-like pleasures. It is often said that Mill's "higher pleasures" are
   merely *greater* pleasures: pleasures with more value. … This would
   be so if the value of the pleasures lay on the same scale. But this is
   what I have just denied. Though each day of the Drab Eternity would
   have some value for me, *no* amount of this value could be as good for
   me as the Century of Ecstasy.

(Parfit 1986: 161)

In Griffin's terms, this is a case of "discontinuity."[12] Furthermore, there
is arguably a vague borderline between the "higher" and "lower" plea-
sures. And some pleasures might classify as "borderline higher." Indeed, if
vagueness is relevant in the beauty-kitsch example, it is equally relevant
in this one. Parfit's own example thus suggests that – while he and Griffin
appear to take different views on the Sorites – vagueness is also relevant
to Parfit's views to the degree that he invokes "discontinuity."

4 Avoiding the Repugnant Conclusion: "Discontinuity" with "Imprecise Equality" or Vagueness?

I now turn to "discontinuity" in the context of population ethics. I begin with a formalization of these examples (following Broome 2004: 55–56; and Qizilbash 2005b: 127). In what follows, we are concerned with populations of various sizes. We begin with a population of ten billion people, living at a very high quality of life. We then reduce the quality of life of those alive somewhat and increase the number of people who are alive at the new quality of life. If we believe that the new population is better than the first, and repeatedly increase the size of the population while slightly reducing the quality of life of those who are alive in this way we might be led to:

> *The Repugnant Conclusion*: for any possible population of at least ten billion people, all with a very high quality of life, there must be some much larger imaginable population whose existence, if other things are equal, would be better even though its members have lives that are barely worth living.
>
> (Parfit 1984: 388)

Parfit believes that moral theories should avoid this conclusion. In what follows, I shall write w for a level of (lifetime) well-being, write w_1 as a very high quality of life, w_2 as a less high quality of life, w_3 as a quality of life less high than w_2, and so on. I assume that there is a finite range of levels of well-being between w_1 and w_k *and* that w_k is a level that is barely worth living. I write numbers n_1, n_2, n_3, and so on, where n_2 (n_3) is smaller than n_3 (n_4) and so on. Finally, I write a world with n_i people living at welfare level w_i as $W(n_i, w_i)$. The step-by-step or spectrum property that can lead to this conclusion is:

> *Property S*: there is a sequence of numbers 10 billion, $n_2, n_3, ..., n_k$ starting with 10 billion, with the property that a population of n_i people living at level of well-being w_i is better than a population of $n_{(i-1)}$ living at $w_{(i-1)}$, for any integer i such that k≥i>10 billion.

We can also define:

> *Property D*: a population of 10 billion people living at well-being level w_1 is better than any number n_k of people living at level of well-being w_k.

Returning to the earlier account of value relations, which followed a reading of Griffin's and Parfit's works, B is irreflexive, asymmetric, and transitive. Then we know (following theorem 2 in Qizilbash 2005b: 127)

that if a sequence $w_1,\ldots, w_k$ has Property S, then it does not have Property D.[13] In the light of this result to avoid the Repugnant Conclusion we must reject Property S, or we must reject asymmetry or transitivity of B. Since on my reading Parfit does not reject asymmetry or transitivity of B, he must reject Property S.

It is worth noting here that there is another route to rejecting the Repugnant Conclusion. If we drop the assumption that the final level of well-being (w_k) in the range is (i.e., falls to the level of) a life barely worth living we can, arguably, accept the intuition in spectrum examples, without accepting the Repugnant Conclusion. In terms of the formalization above, we can then accept Property S and reject Property D while not accepting the Repugnant Conclusion. Ken Binmore and Alex Voorhoeve (2003) have developed this line of argument in the context of some spectrum examples.[14] In the current context, this argument appears to provide a neat way of avoiding the Repugnant Conclusion. However, the argument typically hinges on an assumption about the formal representation of value taking a particular form (involving "asymptotes"; see, for example, Binmore and Voorhoeve 2003: 277).[15] And the ways of avoiding the Repugnant Conclusion that Griffin and Parfit advance do not involve making any such assumption.

Suppose then that we do not adopt any assumption about the representation of value of the sort that Binmore and Voorhoeve make and that there is no sequence of the sort required by Property S, which culminates in n_k people living at a level of well-being w_k, which is a life barely worth living. If so, there is some first level of well-being in the sequence such that we cannot reduce the level of well-being somewhat and find a larger number of people living at the lower level of well-being, which makes the resulting world better than the world with the smaller population. It seems hard to believe that there is a first and precise level of well-being in the spectrum where this is the case.[16] This is, arguably, where vagueness is relevant if one starts from the supposition that this sort of example involves "discontinuity." Again, I take "high" to be the relevant vague predicate. Starting from a very high quality of life, when one reduces the quality of life in a step-by-step way, there will be some level of well-being that is no longer "high." We might, following Griffin's suggestion, then define this as a level where enough of what makes life good has been lost so that it is no longer true that lowering the quality of life to this level and increasing the number of people alive makes the world better (Griffin 1986: 340).[17] There is an imprecise borderline between those levels of well-being that are "high" and those that are not. There will be a zone of levels of well-being that are definitely "high," those which are definitely "not high" and a vague zone in between where the level of well-being is neither definitely high nor definitely not high (for a fuller discussion, see Qizilbash 2005b: 118–126). As long as a life barely worth living is definitely not "high," in this sense, we can avoid the Repugnant Conclusion.[18]

This response to spectrum examples and avoidance of the Repugnant Conclusion does not invoke the relation of "parity" or "rough equality." Because Parfit invoked "rough comparability" and "in the same league as" in his discussion of the "mere addition paradox," there is now literature on "mere addition," incompleteness, and "parity" defined in various ways that attempt to formalize some of Parfit's discussion of this paradox (Blackorby, Bossert and Donaldson 1996; 1997; and 2005; Qizilbash, 2005a, 2007a, 2007b; and 2018; Rabinowicz, 2009a; and 2009b; and Gustafsson 2019), and there is also some critical literature relating to the relevant proposals (Arrhenius 2000, 2016 and forthcoming; and Nebel 2015 *inter alia*). In the more recent literature, notably in Parfit's final paper on this topic, spectrum (or "continuum") cases, the Repugnant Conclusion, and "mere addition" are discussed together. In the context of "mere addition," the suggestion is that the addition of people who have lives worth living does not affect existing people, and their addition does not involve any social injustice (Parfit 1984: 420) at a range of levels of well-being roughly comparable with the status quo. Parfit invokes "rough comparability" in cases of mere addition because we cannot plausibly claim that such addition makes the world better or worse or that the two worlds are exactly as good either. If, like Parfit (1984: 430–432), we believe that these worlds are, nonetheless, comparable (and accept the account of value relations advanced earlier), we must believe that they are roughly comparable in value or on a par.[19] Yet if we restrict our attention to spectrum examples of the sort which I discuss in this chapter, where there is no mere addition, then any potential roughness in comparison involved in such addition cannot be invoked without some further explanation.

If mere addition is not the source of "incommensurability," where does it enter? In his later work, Parfit tries to "defend a strong form of lexical superiority," which is akin to what Griffin calls "discontinuity." Parfit writes,

> When we say that things of kind P are *lexically better* than things of kind Q, we can mean that, though the existence of more Qs would always be non-diminishingly better, the existence of some sufficient number of Ps would be better than any number of Qs. There is a similar sense of *lexically worse than*.

In defense of this form of lexical superiority, Parfit invokes what he calls "evaluative imprecision." Such imprecision is relevant only if one item is not better than the other. If so, Parfit claims that either they are precisely or imprecisely equally good. In the first case, the relation is transitive, and in the latter case, it is not (Parfit 2016: 114). And as a result of non-transitivity of imprecise equality, "not worse than" is non-transitive (Parfit 2016: 115). Parfit leaves no room for incomparability. These remarks

about value relations and transitivity map directly onto the relation system described in Section 2: E is transitive; P is non-transitive, and A maps onto (Parfit's sense of) "not worse than" and is non-transitive. Parfit thus invokes a further form of "incommensurability" – imprecise or rough equality which in the system of relations outlined here would be defined as "parity" – in defense of the lexical view (or "discontinuity"). Yet, since Parfit has already invoked one form of "incommensurability" ("discontinuity"), the further invoking of "imprecise equality" is hard to explain. If there is a difference in the *kind* of lives lived at some point in the spectrum, and "discontinuity" (or a "lexical view") is invoked to account for it, it is not obvious why one should also invoke "imprecise equality" or parity. Inasmuch as he does so, Parfit *appears to* double count "incommensurability." [20]

Why does Parfit invoke "imprecise equality" in spectrum examples? The *imprecise lexical view*, for Parfit, addresses one of the chief weaknesses of standard lexical views. His line of argument runs as follows:

> Some … writers claim that lexical views cannot be applied to actual or possible things that are in the same category and could together form some continuum, with each thing being only slightly different from its neighbours. We cannot defensibly claim that everything above some point on such a continuum is lexically better than everything below this point, since there cannot be a great difference in value of similar things that are just above or just below this point.
>
> (Parfit 2016: 116)

Parfit believes that this objection fails when there is imprecise comparability. When there is imprecise equality, instead of appealing to some single point

> in this continuum which separates greater and lesser goods, we can appeal to various zones in which things would be imprecisely equally good, bounded by zones in which it would be indeterminate whether one of two things would be better, or things would be imprecisely equally good.
>
> (Parfit 2016: 116)

This view *appears to be* a close relation of views that invoke "rough equality" or "parity" in the context of mere addition since those views also allow for a zone where parity holds, which is surrounded by vague zones (see Qizilbash 2005a, 2007a and b; and 2018; and Rabinowicz 2009a and b; and 2012). Indeed, Ruth Chang (2016 and forthcoming) defends a similar view that invokes a zone of parity in "continuum" cases.[21] However, Parfit's motivation – as with the account that invokes vagueness – is to allow for *imprecision*. While semantic vagueness is ruled

out, except at the borderlines of the zone of imprecise equality, in Parfit's view imprecise equality does much the same work that vagueness does on the "discontinuity" view that starts from Griffin. Parfit suggests that the relevant "continuum arguments" assume that "[c]ompared with the existence of many people who would all have some quality of life, there are some much larger numbers of people whose existence would be better, though these people's quality of life would all be slightly lower" (Parfit 2016: 120). He claims that we could reject this assumption by claiming that "such larger worlds would not be better, but would at most be imprecisely equally good" (Parfit 2016: 121). And he believes that this way of rejecting the assumption is less implausible than the Repugnant Conclusion and that the balance of plausibility favors imprecise equality.

Parfit's view does nonetheless run into difficulties. A formal analysis of Parfit's imprecise equality view – presented by Wlodek Rabinowicz (forthcoming) – shows that it has implications that are implausible. I here briefly explain Rabinowicz's version of this argument (which itself follows the analysis of Toby Handfield and Rabinowicz, 2018) in the terminology used here. Consider some point in the finite sequence of population levels. Suppose that we start at some level of well-being w_i and population n_i in the sequence, and we find some somewhat lower w_j and some number, say, 200 billion, which is greater than n_i. I will write this world as $W(w_j, 200\text{ billion})$. Suppose next that $W(w_j, 200\text{ billion})$ is on a par with (and so not better than) $W(w_i, n_i)$. Is that enough to violate Property S? If we think of parity in terms of "rough" or "imprecise" equality, it seems that this will not be enough. If worlds that are on a par are roughly equal in value, then we can invoke the "mark of parity" to distinguish parity from incomparability. Consider again $W(w_j, 200\text{ billion})$, which is on a par with $W(w_i, n_i)$. Parfit (2016, 110) holds the Simple View according to which "[a]nyone's existence is in itself good, and makes the world in one way better, if this person's life is good to live, or worth living." Since at well-being level w_j life is worth living, we can make the world better by increasing the number of people who live at this level and find some $W(w_j, n_j)$, which is much, and thus significantly, better than $W(w_j, 200\text{ billion})$.[22] If so, then by the mark of parity, $W(w_j, n_j)$ is better than $W(w_i, n_i)$, and there is no violation of Property S. Indeed, if this were not so, then if parity is rough equality, $W(w_i, n_i)$ and $W(w_j, 200\text{ billion})$ would be incomparable rather than on a par (see also Herlitz, 2020: 473–478).[23] If Parfit has in mind a form of "rough equality" when he talks of "imprecise equality," his view does not violate Property S and given the other assumptions I have made based on my reading of Parfit, he must reject Property D and cannot avoid the Repugnant Conclusion.

Parfit's view must be amended in some way to avoid that conclusion. One way to do this is to drop the requirement that parity is a form of "rough equality" and allow for the weaker version of the "mark of parity" discussed earlier. On this alternative, as we saw, when parity holds

between items, while some slight change in value may not tilt the balance in favor of one of the items, either some significant increase in value will make it better than the other or some significant decrease will make it worse. Now suppose once more that two worlds $W(w_i, n_i)$ and $W(w_j,$ 200 billion) in the sequence are on a par. Then on this weakening of the mark of parity, while some significant worsening of $W(w_j,$ 200 billion) may make it worse, even some significant improvement which involves increasing the number of people who are alive at level w_j may not make the world better than $W(w_i, n_i)$. If so, at w_j, there is no n_j such that $W(w_j, n_j)$ is better than $W(w_i, n_i)$. In this way, the sequence can violate Property S, and given the other assumptions adopted in the system of value relations defined earlier, we can accept Property D and avoid the Repugnant Conclusion. But we must drop the view that parity is "rough equality," and that involves departing from Parfit's view as I have interpreted it here.[24] As I argued earlier, this way of understanding parity may nonetheless be consistent with interpreting "on a par with" as "in the same neighborhood or league as."[25] It may also be consistent with another ("fitting-attitudes") view of parity that does not require that parity is "rough equality" (see Rabinowicz, 2009a and forthcoming).

In contrast to Parfit's view, the view that began from Griffin's account of "discontinuity" only invokes one form of "incommensurability." To this degree, it is arguably more parsimonious and does not double count "incommensurability" – and this point may count in its favor. Vagueness is invoked for much the same reason that Parfit invokes "evaluative imprecision": to allow for a zone rather than a first and single point in the sequence where "discontinuity" sets in. One might speculate about why Parfit did not favor a vagueness view. It may be that he favored a zone of imprecise equality (with rough borders) over vagueness *simpliciter* because he believed that there was no Sorites or Sorites-like structure in the relevant spectrum examples. Yet as I argued earlier, even if we set aside the question of whether there is a Sorites or Sorites-like structure at work in these examples, in the "single-life analogue of the Repugnant Conclusion" when Parfit invokes "discontinuity," vagueness is arguably relevant. Alternatively, Parfit may not have favored the vagueness view because of various objections to it. It is thus worth asking: what objections are there to this view? And, in particular, do these objections lead one to favor Parfit's view over the vagueness view?

One objection is that while introducing vagueness makes the threshold less sharp, there is still indeterminacy about a first and unique point in the spectrum below which *any* increase in quantity does not compensate for some fall in quality. The objection is that there *is* such a point, irrespective of whether there is indeterminacy about it. As Handfield and Rabinowicz (2018: 2385) put it, "This is counter-intuitive, whether or not it is indeterminate where this point occurs" (see also Pummer, forthcoming). My reading of this objection is that it presses the point that accepting the

implications of "discontinuity" conflicts with our intuitions since (given the various assumptions made) it implies a first and unique point in the sequence where some reduction in welfare cannot be compensated by an increase in population. Since both Griffin and Parfit invoke "discontinuity," this objection does not favor Parfit's view over the position which began from Griffin's remarks. [26]

Another objection to views that invoke vagueness in the context of spectrum examples is that while in the case of many vague predicates the indeterminacy – say about whether or not someone is "bald," "tall," or "poor" – can be defused by arbitrary stipulation, it would be odd to attempt to defuse controversies in the context of spectrum examples in value theory in this way. It is (arguably) odd because these controversies involve substantive value judgments (see Chang, forthcoming; and 2002a: 137–139). Yet, it is recognized that there are debates in ethics and public policy in which substantive value judgments are in play where arbitrary stipulations might be invoked to address vagueness or imprecise thresholds.[27] As John Rawls (2001, 35) notes, "To some degree all our concepts, and not only our moral and political concepts, are vague and subject to hard cases." It would be surprising if population ethics did not involve vague concepts or imprecise thresholds. Let us suppose, nonetheless, that this response to the objection proves unconvincing. Does this objection necessarily favor "discontinuity" views that invoke rough equality or parity, where – as in Parfit's view – the zone of parity or rough equality is bordered by vague zones over those that invoke only vagueness. To the degree that Parfit's view invokes vague zones, the same objection might be leveled at it. For example, in the zone where it is indeterminate whether one world is better than or imprecisely equal in value to the other, it might be argued that what is at stake is a substantive value judgment rather than a vague borderline, and that arbitrary stipulation cannot be used to decide what value relation holds as it might be in cases of vagueness.

5 Conclusions

Two distinct approaches to avoiding the Repugnant Conclusion and addressing spectrum examples emerge from James Griffin's *Well-Being* and Derek Parfit's later work. In one, only one form of "incommensurability" – "discontinuity" – is invoked in the slide from a high quality of life to a life barely worth living, while vagueness is invoked to address the presence of borderline cases and imprecise borders. In Parfit's case, imprecise equality is invoked to allow for imprecision in valuation, while vagueness is invoked only to allow for the absence of sharp borderlines of the zone of imprecise equality. In this chapter, an account of value relations has been presented that is consistent with a reading of Griffin's and Parfit's views and that allows for a distinct relation of parity understood

as rough equality. It is argued on the basis of this reading and account that Parfit's invocation of imprecise equality cannot plausibly address some of the implications of allowing for "discontinuity" or a lexical view of the sort Parfit defends if it is interpreted in terms of parity understood as rough equality. The implied way of distinguishing cases of parity and incomparability if parity is rough equality – the "mark of parity" – must be weakened to make this view plausible. Parfit's view also appears to double count "incommensurability" by invoking both "discontinuity" (i.e. a lexical view) and imprecise equality in the context of these examples. When the respective merits of the two views – the one invoking only vagueness and the other invoking parity understood as "rough equality" in the context of "discontinuity" in avoiding the Repugnant Conclusion – are compared, this point arguably favors the view that does not invoke imprecise equality.

Notes

1 For helpful comments and exchanges on a very early version, I thank Roger Crisp as well as Chrisoula Andreou, Gustaf Arrhenius, John Broome, Krister Bykvist, Ruth Chang, Anders Herlitz, Wlodek Rabinowicz and other participants at the Workshop on "Incommensurability: Vagueness, Parity and Other Non-Conventional Value Relations" at the Institute for Future Studies, Stockholm in December 2019. I am extremely grateful to Henrik Andersson, Anders Herlitz, and Wlodek Rabinowicz for very helpful exchanges and insightful comments on an earlier version. I acknowledge a debt of gratitude to Derek Parfit. To the degree that it contains any insight, this chapter is dedicated to the memory of James Griffin. Any error is mine.

2 Griffin's first influential discussion of the subject predates this: see Griffin (1977).

3 Because there are a variety of senses of "incommensurability" in this literature, I place the word in inverted commas throughout the chapter.

4 In his discussion, Griffin (1986: 77) distinguishes various senses of "incommensurability" – where "incomparability" is the "strongest" sense (since it bars all comparisons in terms of "better" "worse" and "equal in value"), whereas "weaker" versions include "discontinuity," which allows for such comparisons.

5 To the degree that "rough equality" is a form of "incommensurability" (as Griffin implies, see Griffin, 1986: 97), and "rough equality" can be understood in terms of a vague ordering, then "incommensurability" can be a form of "vagueness" in Griffin's terms. For this reason, while Griffin (2000: 287) notes that he himself never says that "incommensurability is vagueness," Broome (2000: 30) takes him to make this claim. Nonetheless, the discussion of "incommensurability and vagueness" (which starts from Broome, 1997, 2000; and 2004; see also Rabinowicz, 2009b and Qizilbash, 2012) typically treats these as mutually exclusive, so that the implicit notion of "incommensurability" is what (in this chapter) I define as "incommensurateness." Because, as will become clear, "rough equality" understood as a distinct relation is not a form of "vagueness," it makes sense to ask "rough equality or vagueness"?

6 See also Arrhenius and Rabinowicz (2005).

7 I am using "or" in its inclusive sense in outlining the account of value relations.

8 Sometimes "incommensurable" is used in this sense. See, for example, Rabinowicz (forthcoming).

9 This "mark" is distinct from what Joseph Raz (1986b: 325–326) calls the "mark of incommensurability." Raz uses the "mark of incommensurability" to determine whether or not options are, in the terms I am using, *incommensurate*. If, of two options, "neither is better than the other," but "it is possible for one of them to be improved without thereby becoming better than the other, or if there can be another option which is better than the one but not better than the other, then the two original options are incommensurate." While Raz's mark distinguishes cases of exact equality from incommensurateness, it does not distinguish cases of parity or rough equality from incomparability since in both cases items are incommensurate.

10 On the related matter of whether a zone of parity (understood as "rough equality" or otherwise) can be "wide" see Qizilbash (2018: 96–99).

11 See, in particular, Andreou (2021b: 19), who writes that on her view, "[p]erhaps every pair of options can be located in a shared league with either an upper bound *or* a lower bound."

12 Indeed, this is the natural way to read Mill's discussion of "higher" and "lower" pleasures. See Mill (1962: 259). Roger Crisp (1997: 30–31) provides a reading along these lines.

13 The proof runs as follows. Suppose to the contrary that a sequence of levels of well-being, $w_1, w_2, …, w_k$ has Property S, as well as Property D. Then it follows that there is a sequence of numbers, 10 billion, $n_2, …, n_k$ such that $W(n_2, w_2)$ B $W(10 \text{ billion}, w_1)$; $W(n_3, w_3)$ B $W(n_2, w_2)$; …; $W(n_k, w_k)$ B $W(n_{k-1}, w_{k-1})$. It then follows from transitivity of B that $W(n_k, w_k)$ B $(10 \text{ billion}, w_1)$. But the sequence has Property D so that $W(10 \text{ billion}, w_i)$ B $W(n_k, w_k)$. Asymmetry of B then implies that not $W(n_k, w_k)$ B $W(10 \text{ billion}, w_1)$. This is a contradiction.

14 I thank John Broome for pressing this point. Nonetheless, Erik Carlson (2005) provides a formulation of spectrum examples that suggests that Binmore and Voorhoeve's examples may not capture the intuitions in these examples. In the population context, the equivalent formulation would ensure that there is a sequence of discrete steps that ends with a life barely worth living.

15 John Broome (2010) uses the same line of argument and a "badness function" involving "asymptotes" against an argument for "a certain sort of discontinuity."

16 This is not a claim about whether there is some pair of worlds *outside the spectrum* where some number of people living at some level of well-being is better than any number of people living at some somewhat lower level of welfare which might, for example, fall below a life barely worth living. There is nothing unintuitive about that claim. I thank John Broome for raising this possibility.

17 It is worth noting that in the endnote in which Griffin (1986: 338–340) makes these remarks, he discusses Parfit's "mere addition paradox." His remarks are, nonetheless, relevant to spectrum cases in population ethics, with or without "mere addition."

18 For critical discussions of Griffin's "discontinuity" view when it is used in the context of population ethics, see Arrhenius (2000: 92–100) and Ryberg (1996: 210–212).

19 John Broome has argued that invoking worlds that are incommensurate (as they are when they are on a par or roughly equal) in the context of mere addition is *ad hoc*. On this, see Broome (2004; and 2007), as well as Qizilbash (2007a, 2007b, and 2018) and Rabinowicz (2009a).

20 One might argue in his defense that Parfit invokes one form of "incommensurability" – "imprecise equality" – to address an objection that is addressed to another – to "discontinuity" setting in at some point in the sequence. On this argument, Parfit *does not* double count "incommensurability." I am grateful to Wlodek Rabinowicz for this point. Yet if evaluative imprecision is invoked to rebut the relevant objection, one must inevitably ask, Why invoke "imprecise equality" – or incommensurateness more generally – rather than vagueness? Invoking "imprecise equality" surely requires more in the way of explanation, and, to avoid the concern about "double counting," the explanation should not be about qualitative differences in the of lives lived in the spectrum since those differences are invoked to explain the presence of "discontinuity."

21 Unlike Parfit, Chang (2016: 210; and forthcoming) drops the lexical view while invoking a zone of parity in addressing spectrum cases.

22 This would be so even if one allows for vagueness at the borderline of the zone of parity.

23 Handfield and Rabinowicz (2018) and Rabinowicz (forthcoming) discuss this sort of case in terms of "radical incommensurability" and its persistence. Anders Herlitz (2020) provides an illuminating discussion that uses and refines these terms and relates them to the literature on parity.

24 For a formal analysis and critical discussion of Parfit's imprecise equality view, see also Jensen (2020). Because Parfit's view invokes "discontinuity," some standard criticisms leveled at Griffin's views also apply to Parfit's.

25 These contrasting ways of distinguishing parity and incomparability in the context of the Repugnant Conclusion are closely related to the helpful distinction between "strong" and "weak non-persistency" advanced by Herlitz (2020: 475–478).

26 My reading of this objection may, of course, be contested. In particular, it can be argued that while on the vagueness view there is a single point where "discontinuity" sets in, this is not so on Parfit's view. I am grateful to Wlodek Rabinowicz for this point. Nonetheless, Handfield and Rabinowicz (2018: 2385) also suggest that if we want to avoid "discontinuity" (or "radical inferiority" in their terms) setting in at some determinate point in the spectrum, and still insist that there is "discontinuity" between the first and last element, then to preserve transitivity of "better than" we must posit either an implausible (or "radical") form of incommensurateness or an implausible (or "radical") form of indeterminacy or vagueness. Here again, the implications of accepting "discontinuity" appear to make *both* accounts – of "imprecise equality" in as much as it is a form of incommensurateness and of vagueness – implausible. In the absence of further analysis, the objection does not favor one account over the other. Henrik Andersson (2021) provides a more extensive and helpful discussion of some of Handfield and Rabinowicz's objections.

27 Perhaps the best-known example is abortion (see, for example, Williams 1995: 214), which has also been discussed in terms of a spectrum example (see Temkin 2000).

References

Andersson, H. (2021), 'Spectrum Arguments, Indeterminacy, and Value Superiority' in H. Andersson and A. Herlitz (eds.), *Value Incommensurability: Ethics, Risk and Decision-Making* (Routledge).

Andreou, C. (2015), 'Parity, Comparability and Choice', *Journal of Philosophy* CXII: 5–22.

Andreou, C. (2021a), 'Parity Without Imprecise Equality', in H. Andersson and A. Herlitz (eds.), *Value Incommensurability: Ethics, Risk and Decision-Making* (Routledge).

Andreou, C. (2021b), 'Incomparability and the Huge Improvement Argument', *American Philosophical Quarterly* 58/4: 307–318.

Arrhenius, G. (2000), *Future Generations: A Challenge for Moral Theory* (Uppsala: University Printers, Ekonomikum).

Arrhenius, G. (2016), 'Population Ethics and Different-Number-Based Imprecision', *Theoria* 82: 166–181.

Arrhenius, G. (forthcoming), *Population Ethics: The Challenge of Future Generations*. (Oxford University Press).

Arrhenius, G. and Rabinowicz, W. (2005), 'Millian Superiorities', *Utilitas* 17: 127–146.

Binmore, K and Voorhoeve, A. (2003), 'Defending Transitivity from Zeno's Paradox', *Philosophy and Public Affairs* 31: 272–279.

Blackorby, C., Bossert, W. and Donaldson, D. (1996), 'Quasi-Orderings and Population Ethics', *Social Choice and Welfare* 13: 129–151.

Blackorby, C., Bossert W., and Donaldson, D. (1997), 'Critical-Level Utilitarianism and the Population-Ethics Dilemma', *Economics and Philosophy* 13: 197–230.

Blackorby, C., Bossert, W. and Donaldson, D. (2005), *Population Issues in Social Choice Theory, Welfare Economics and Ethics* (Cambridge University Press).

Broome, J. (1997), 'Is Incommensurability Vagueness?' in R. Chang (ed.), *Incommensurability, Incomparability, and Practical Reason* (Harvard University Press).

Broome, J. (2000), 'Incommensurable Values', in R. Crisp and B. Hooker (eds.), *Well-Being and Morality: Essays in Honour of James Griffin* (Oxford University Press).

Broome, J. (2004), *Weighing Lives* (Oxford University Press).

Broome, J. (2007), 'Reply to Qizilbash', *Philosophy and Phenomenological Research* LXXV: 152–157.

Broome, J. (2010), 'No Argument Against the Continuity of Value: Reply to Dorsey', *Utilitas* 22: 494–496.

Broome, J. (2012), *Climate Matters: Ethics in a Warming World* (Norton & Company).

Carlson, E. (2005), 'Intransitivity without Zeno's paradox', in T. Rønnow-Rasmussen and M. Zimmerman (eds.), *Recent Work on Intrinsic Value* (Dordrecht: Springer).

Chang, R. (1997), 'Introduction', in R. Chang (ed.), *Incommensurability, Incomparability, and Practical Reason* (Harvard University Press).

Chang, R. (2002a), *Making Comparisons Count* (Routledge).

Chang, R. (2002b), 'The Possibility of Parity', *Ethics* 112: 659–688.

Chang, R. (2005), 'Parity, Interval Value and Choice', *Ethics* 115: 331–350.

Chang, R. (2016), 'Parity, Imprecise Comparability and the Repugnant Conclusion', *Theoria* 82: 182–214.

Chang, R. (forthcoming), 'How to Avoid the Repugnant Conclusion', in J. McMahan, T. Campbell and J. Goodrich (eds.), *Ethics and Existence. The Legacy of Derek Parfit* (Oxford University Press).

Crisp, R. (1997), *Mill on Utilitarianism* (Routledge).

Griffin, J. (1977), 'Are There Incommensurable Values?' *Philosophy and Public Affairs* 7: 39–59.

Griffin, J. (1986), *Well-Being: Its Meaning, Measurement and Moral Importance* (Clarendon Press).

Griffin, J. (2000), 'Replies', in R. Crisp and B. Hooker (eds.), *Well-Being and Morality: Essays in Honour of James Griffin* (Oxford University Press).

Gustafsson, J. (2019), 'Population Axiology and the Possibility of a Fourth Category of Absolute Value', (with a Corrigendum). *Economics and Philosophy* 36: 81–111.

Handfield, T. and Rabinowicz, W. (2018), 'Incommensurability and Vagueness in Spectrum Arguments: Options for Saving Transitivity of Betterness', *Philosophical Studies* 175: 2373–2387.

Herlitz, A. (2020), 'Spectrum Arguments, Parity and Persistency', *Theoria* 86: 463–481.

Jensen, K. (2020), 'Weak Superiority, Imprecise Equality and the Repugnant Conclusion', *Utilitas* 32: 294–315.

Knapp, C. (2007), 'Trading Quality for Quantity', *Journal of Philosophical Research* 32: 211–233.

Mill, J.S. (1962), *Utilitarianism* (William Collins and Co. Ltd).

Nebel, J. (2015), 'Incommensurability in Population Ethics' BPhil Dissertation. (Oxford University).

Parfit, D. (1984), *Reasons and Persons* (Oxford University Press).

Parfit, D. (1986), 'Overpopulation and the Quality of Life', in P. Singer (ed.), *Applied Ethics* (Oxford University Press).

Parfit, D. (2016), 'Can We Avoid the Repugnant Conclusion?', *Theoria* 82: 110–127.

Pummer, T. (2018), 'Spectrum Arguments and Hypersensitivity', *Philosophical Studies* 175: 1729–1744.

Pummer, T. (forthcoming), 'Sorites On What Matters', in J. McMahan, T. Campbell and K. Ramakrishnan (eds.), *Essays in Honour of Derek Parfit. Vol. 2* (Oxford University Press).

Qizilbash, M. (2002), 'Rationality, Comparability and Maximization', *Economics and Philosophy* 18: 141–156.

Qizilbash, M. (2005a), 'The Mere Addition Paradox, Parity and Critical-Level Utilitarianism', *Social Choice and Welfare* 24: 413–431 (and Erratum Note).

Qizilbash, M. (2005b), 'Transitivity and Vagueness', *Economics and Philosophy* 21: 109–131.

Qizilbash, M. (2007a), 'The Mere Addition Paradox, Parity and Vagueness', *Philosophy and Phenomenological Research* LXXV: 129–151.

Qizilbash, M. (2007b), 'The Parity View and Intuitions of Neutrality', *Economics and Philosophy* 23: 107–114.

Qizilbash, M. (2012), 'Incommensurability or Vagueness? A Comment on Rabinowicz and Sugden', *Proceedings of the Aristotelian Society* CXII: 331–336.

Qizilbash, M. (2018), 'On Parity and the Intuition of Neutrality', *Economics and Philosophy* 34: 87–108.

Rabinowicz, W. (2009a), 'Broome and the Intuition of Neutrality', *Philosophical Issues* 19: 389–411.

Rabinowicz, W. (2009b), 'Incommensurability and Vagueness', *Proceedings of the Aristotelian Society* LXXXIII: 71–94.

Rabinowicz, W. (2012), 'Value Relations Revisited', *Economics and Philosophy* 27: 133–164.

Rabinowicz, W. (forthcoming), 'Can Parfit's Appeal to Incommensurabilities in Value Block the Continuum Argument for the Repugnant Conclusion?', in J. McMahan, T. Campbell and J. Goodrich (eds.), *Ethics and Existence. The Legacy of Derek Parfit* (Oxford University Press).

Rawls, J. (2001), *Justice as Fairness. A Restatement*. E. Kelly (ed.) (The Belknap Press of Harvard University Press).

Raz, J. (1986a), 'Value Incommensurability: Some Preliminaries', *Proceedings of the Aristotelian Society* 86: 117–134.

Raz, J. (1986b), *The Morality of Freedom*. (Clarendon Press).

Ryberg, J. (1996), 'Parfit's Repugnant Conclusion', *The Philosophical Quarterly* 46: 202–213.

Sen, A. (1979), *Collective Choice and Social Welfare* (North Holland).

Temkin, L. (1996), 'A Continuum Argument for Intransitivity', *Philosophy and Public Affairs* 25: 175–210.

Temkin, L. (2000), 'An Abortion Argument and the Threat of Intransitivity' in R. Crisp and B. Hooker (eds.), *Well-Being and Morality. Essays in Honour of James Griffin* (Oxford University Press).

Temkin, L. (2012), *Rethinking the Good: Moral Ideals and the Nature of Practical Reasoning* (Oxford University Press).

Williams, B. (1995), 'Which Slopes Are Slippery?' in *Making Sense of Humanity and Other Philosophical Papers* (Cambridge University Press).

5 Spectrum Arguments, Indeterminacy, and Value Superiority

Henrik Andersson

1 Introduction[1]

The assumption that the value comparative "better than" is transitive is central to most normative reasoning.[2] If it is not a transitive relation then there may be alternatives, such as *A*, *B*, and *C* for which *A* is better than *B*, *B* is better than *C*, and yet *C* is better than *A*. This possibility would make it impossible to pick a best alternative since it allows for a cycle. However, in his book *Rethinking the Good: Moral Ideals and the Nature of Practical Reasoning*, Larry Temkin (2012) forcefully questions the assumption that "better than" is a transitive relation.[3] Roughly, his explanation as to why our value relations might not be transitive is that, depending on what I compare a specific object with, different considerations will be more or less weighty. Or, as Temkin (2012) puts it, the assessment of relative goodness is sometimes "essentially comparative." When comparing *A* with *B*, some specific consideration is central, and *A* is better than *B* with respect to this consideration. The same consideration may be central when comparing *B* with *C*, and *B* is better than *C* with respect to this consideration. However, when comparing *C* with *A*, some other consideration may be more central, making *C* better than *A*. This can explain the potential cyclicity.

Temkin's arguments in favor of this possibility are very detailed and sophisticated; it is impossible to do justice to all of them here. However, his spectrum arguments are central to his claim. Temkin uses spectrum arguments to argue that the belief that the "better than" relation is transitive is incompatible with other more intuitive beliefs, so we should conclude that "better than" is not transitive. While there are several possible formulations of spectrum arguments, I will focus on the formulation of the argument that Temkin deems "the simplest, most powerful counterexample to the transitivity of the betterness relation" (Temkin 2012: 135).[4] If it can be shown that Temkin's most convincing formulation of the argument fails, then this casts doubt on the endeavor in general.

I will argue that Temkin's argument does not threaten the transitivity of betterness. First, I will provide an informal account of why his

argument fails. Roughly, I will argue that if we acknowledge that there can be a zone of vagueness in Temkin's spectrum argument, then we need not accept his conclusion. Second, I will show how the formal features of our value relations provide a neat explanation as to why we ought not to reject the transitivity of the betterness relation and that we instead should reject a belief expressed by Temkin about evaluative trade-offs. It is important to note, in the context of this book, that my rebuttal of Temkin's argument is conservative in that it does not refer to non-conventional value relations such as parity, imprecision, or rough equality. As such, my objection could be understood to be part of the larger effort to provide an account of the normative landscape without referring to non-conventional value relations; the "at least as good" relation and vagueness suffice.[5]

2 The Spectrum Argument

Temkin's preferred formulation of the spectrum argument starts by noting that the following views seem intuitively appealing:

> *View One*: For any unpleasant or "negative" experience, no matter what the intensity and duration of that experience, it would be better to have the experience than one that was only a "little" less intense but twice (or three or five times) as long.
>
> *View Two*: There is, or could be, a spectrum of unpleasant or "negative" experiences ranging in intensity, for example, from extreme forms of torture to the mild discomfort of a mosquito bite, such that one could move from the harsh end of the spectrum to the mild end in a finite series of steps, where each step would involve the transformation from one negative experience to another that was only a "little" less intense than the previous one.
>
> *View Three*: The mild discomfort of a mosquito bite would be better than two years of excruciating torture, no matter how long one lived and no matter how long the discomfort of a mosquito bite persisted.
>
> *View Four*: "All-things-considered better than" is a transitive relation. So, for any three outcomes, A, B, and C, which involve unpleasant experiences of varying intensities and durations, if, all things considered, A is better than B and B is better than C, then A is better than C.
>
> (Temkin 2012: 135)

Now consider a series of lives from P_1 to P_n. All these lives are very lengthy and all include 15 mosquito bites per month. P_1 and P_2 are similar in all relevant respects, except that P_1 includes two years of excruciating torture, while P_2 includes four years of almost as intense torture. P_3

stands in the same relation to P_2 as P_2 does to P_1, i.e., P_3 includes eight years of almost as intense torture as does P_2. We would say, in accordance with View One, that P_1 is better than P_2, and that P_2 is better than P_3, and so on. However, in accordance with View Two, we will eventually reach P_n, which is such that it does not include excruciating torture, but very minor discomfort for a very long time. It may very well be that this minor discomfort is just one extra mosquito bite per month. Given View Four, P_1 is better than P_n, but according to View Three, P_n is better than P_1. It has thus been shown that Views One, Two, Three, and Four are incompatible. Something has to go, but all four views seem correct. Moreover, rejecting any of the four views seems to lead to both practical and theoretical problems.[6] Temkin argues that it is View Four that must be rejected, as rejecting any of the other views would come at too high a price.

3 The Spectrum Argument and Indeterminacy

There are, of course, ways in which one can object to spectrum arguments of this type, and Temkin discusses several such objections. There is, however, one type of objection I wish to discuss further. This objection is that there must be a point in the spectrum where View One does not apply between two adjacent outcomes, breaking the chain of "better than" relations.[7] One explanation as to why that might be the case is that differences in degree may give rise to a difference in kind, and since there clearly is a difference in degree in the spectrum argument, there could also be a difference in kind. That is, not only do the pains differ in intensity, but the experiences also differ in a more fundamental manner that cannot be fully accounted for in terms of intensity – i.e., they have different phenomenological characteristics. This difference in kind becomes clear when one considers the difference between excruciating torture and a mosquito bite. Such a difference is not merely a difference in degree but also a difference in kind; they belong to two different categories of experiences. If there is a difference in kind between experiences at the opposite ends of the spectrum, then we should expect this difference in kind to arise somewhere in the spectrum. That is, a difference in kind can also, albeit less obviously, be found between two adjacent experiences in the spectrum. This means that even though two adjacent experiences are similar, there could be a difference in kind between them such that when we compare them the factors that are relevant and significant for comparing other adjacent experiences in the spectrum no longer apply (Temkin 2012: 266). The simple trade-off expressed in View One does not hold for these experiences since the pains are of a sufficiently different kind.

Temkin admits that there are such differences in kind in the spectrum; the reason that View Three applies to experiences at the opposite ends

of the spectrum is that there is a difference in kind between these experiences, excruciating torture differs in kind from a mosquito bite. Temkin even takes this to explain why transitivity does not hold. Due to the difference in kind between experiences at the opposite ends of the spectrum, factors that are relevant and significant when comparing these experiences differ from the factors that are relevant and significant for comparisons of experiences that differ merely in degree.

The objection, however, states that this difference in kind not only holds between experiences at the opposite ends of the spectrum; rather, the difference arises somewhere between adjacent experiences in the spectrum. For these experiences, due to their difference in kind, View One is not applicable, so the transitivity of "better than" is not threatened. That is, somewhere in the spectrum, two adjacent pains are of such different characters that the difference in time is not enough to outweigh the difference in kind and consequently we should reject View One.

I am inclined toward the gist of such an objection. However, Temkin is not convinced by it. His reply takes the following form: P_1 clearly differs in kind from P_n, and the objection states that there is a break somewhere in the spectrum from P_1 to P_n where this change takes place. Let us call the two adjacent experiences for which View One does not hold P_{k-1} and P_k. Temkin then asks whether View Three applies to P_{k-1} and P_k:

> For our opponent's argument to work, he needs the difference between P_k and P_{k-1}'s pain to be *akin to the difference between the pain of intense torture and the pain of a mosquito bite*. If there *were* such a difference, then, indeed, we would agree that View Three applied to the two "nearby" pains P_k and P_{k-1}; correspondingly, we would reject View One, and the threat to the transitivity of "better than" would evaporate.
>
> (Temkin 2012: 272)

This response is misdirected, as Temkin focuses on View Three rather than View One. We should not assume that these views are mutually exclusive. The real issue is whether View One applies or not. The relevant question is thus whether the difference in kind is enough for View One not to apply. Temkin's view of this is that the difference is not enough. According to him, even though P_{k-1} and P_k may differ in kind, they are clearly very similar. For example, P_{k-1} is more similar to P_k than to P_1, and P_k is more similar to P_{k-1} than to P_n. Given that they are so very similar, we should expect View One to hold between them, according to Temkin.

Surprisingly, with this response, Temkin acknowledges that the experiences may indeed differ in kind. According to Temkin, however, the focus on different kinds is a red herring; instead, we should focus on the similarity of the adjacent experiences. He writes,

In sum, though it may have helped illuminate how it could be the case that different factors could be relevant and significant for comparing different alternatives, the issue of different *kinds* is a red herring, and I ought not to have put some of my earlier discussions in those terms. The point is simply that when the difference between two pains is "sufficiently" great, View Three is appropriate for comparing them, whereas when the difference between two pains is "sufficiently" small, View One is appropriate.

(Temkin 2012: 274)

This, however, does not amount to a rebuttal of the claim that differences in kind constitute a counterexample to View One. Rather, Temkin seems to dig in his heels and once again claim that since P_{k-1} and P_k are next to each other in the spectrum, View One must be applicable, thus disregarding the claim that if P_{k-1} differs in kind from P_k, then View One might not be applicable. The fact that P_{k-1} and P_k are adjacent does not rule out the possibility that they differ in kind. It is well-known that small changes can trigger larger changes, such as the proverbial straw that broke the camel's back. Small nonevaluative changes that trigger big changes in value are no less familiar. Adding a pinch of cilantro can, according to some, ruin an otherwise tasty dinner.

What one might find implausible in my suggestion, however, is that there is an abrupt change in the spectrum. One might find it implausible that along a fine-grained pain spectrum there exists an abrupt change in kind between two points. It may seem odd that there is such a *sharp* break in the spectrum (see Temkin 2012: 275).

Let us grant that a sharp break of this kind seems implausible. It could, however, still be that there is a break but that this break is not sharp. This is a familiar idea. A single grain of sand does not constitute a heap, but if we keep on adding grains of sand, we will eventually end up with a heap. There is, however, no exact lower limit on the number of grains of sand that is needed to rightly call the collection of grains "a heap." For some number of grains, it is indeterminate whether it is a heap or not. Similarly, there could be a zone in which it is indeterminate whether View One applies to the adjacent outcomes.[8] It may, for example, be indeterminate whether the difference in kind is to be found between P_k and P_{k-1}, P_{k+1} and P_k, or P_{k+2} and P_{k+1}. This means that although we cannot determinately judge for which outcomes in the spectrum View One does not apply, it will not be true for all adjacent experiences that View One will apply, since for P_{k-1} to P_{k+2} it will be indeterminate. This should make the claim more plausible.

Temkin (2012: 534–538) discusses a similar idea presented by Christopher Knapp (2007). Knapp argues that trade-offs between intensity and time seem to be desirable only when we are dealing with quantitative differences and not when it comes to qualitative differences. If

we have a spectrum ranging from very intense pain (600) to very mild pain (1), there is a difference in quality between the pains at the different ends of the spectrum, while two pains at the same end of the spectrum will only differ in quantity. For pains in the middle of the spectrum (300 and 301, for example), it is indeterminate whether they are very intense or very mild.[9] And since it is indeterminate for two adjacent points in the middle of the spectrum whether there is a difference in quality, then it is indeterminate whether one of these outcomes that lasts for a certain time is better than its adjacent outcome that lasts much longer. Consequently, it is neither true nor false that View One holds and thus we have a break in the chain.[10] However, Temkin is not much impressed by this objection:

> We must clearly reject its key premise that a pain's being a borderline case of a qualitative distinction guarantees that it would be indeterminate whether or not View One or View Three applied for any comparisons involving that pain. This is simply not so. After all, even if one grants that pain 301 is both indeterminately very intense and indeterminately very mild, and so a borderline case of qualitative distinction, it *doesn't* follow that a trade-off between *one* year of pain 301 and *five* years of pain 301 would involve a borderline case *where a qualitative difference* was at stake! Since there is *no* difference between the intensity of pain in the two cases, only a difference in duration, it cannot be indeterminate whether a qualitative difference is at stake.
>
> (2012: 536)

Admittedly, I find this reply somewhat unclear. It may be that he has shown that Knapp's "key premise" must be rejected, but this does not counter the more general reply that I have presented here.

However, Temkin continues his response to the line of reasoning I have presented here and provides a more general answer. According to him, it might very well be that it is indeterminate whether pains 300 and 301 are very intense or very mild, but it is determinate that the difference between them is small, and View One is appropriate to apply when we compare pains with a small difference between them in intensity.

Once again, Temkin seems to dig his heels in and claim that the two experiences are sufficiently similar, and by doing so he is neglecting the possible consequence of the difference in kind that holds between the experiences. His reply consequently seems to misconstrue the original objection. It was argued that – for the same reason that View One does not apply to experiences at the opposite ends of the spectrum – if adjacent experiences differ in kind, then View One is not applicable to these adjacent experiences. That is, if there is a difference in kind, then different criteria may be relevant to the comparison, blocking the chain of betterness relations. It may be indeterminate where this difference in kind lies

on the spectrum, but it does, however, exist. For this reason, View One should be rejected. View One is supposed to be applicable to any two pains such that one is slightly less intense than the other. We now know that this is false since for some such pains there is a difference in kind, and for these pains, View One is not applicable.

By rejecting View One, Temkin's "most powerful" counterexample to the transitivity of "better than" fails, so there is no need to reject the view that "better than" is not a transitive relation.

4 The Spectrum Argument and Value Superiority

Let me try to make the same point but in a somewhat different manner. I take the issue at stake to be whether we should reject View One or View Four. Temkin has shown that if we accept the very reasonable Views Two and Three, Views One and Four will be incompatible. If we are to reject one of these views, we must have an explanation as to why the view might not hold. Temkin has given explanations as to why View Four might not always hold by arguing that the assessment of relative goodness is sometimes essentially comparative. However, accepting this as an explanation requires a revision of our conception of the nature of value: it would require that we "rethink the good." Moreover, View Four is fundamental to our conception of value, so if an explanation can be given as to why View One does not hold for all adjacent experiences in the spectrum, then this casts doubt upon Temkin's conclusion. I gave such an explanation above, so now we have reasons to doubt Temkin's conclusion.

Here I will give a somewhat more technical explanation as to why, given Views Two, Three, and Four, View One fails to apply. I will show that for two adjacent experiences in the spectrum, one will be *weakly inferior* to the other, which supports the claim that they differ in kind, and consequently, View One cannot hold for these two experiences. Thus, the chain of experiences allegedly connected in the manner described by View One is broken. So, by assuming transitivity, an explanation as to why View One fails to hold can be given. That is, from the knowledge we have of the structure of value, an explanation can be given as to why we should reject View One.

Interestingly, the argument I will present is similar to the view that Mozaffar Quizilbash, in this volume, attributes to James Griffin. Griffin refers to "discontinuity" and vagueness to show how the spectrum argument can be dismantled. I will not, however, use the term "discontinuity," but prefer the terms "superiority" and "inferiority." My argument will be based on research by Gustaf Arrhenius and Wlodek Rabinowicz (2015) regarding superiority in value.[11] According to them, there are two kinds of superiority:

> *Strongly superior*: An object *e* is *strongly superior* to an object *e'* if and only if *e* is better than any number of *e'*-objects.

> *Weakly superior*: An object *e* is *weakly superior* to an object *e'* if and only if for some number *m*, *m e*-objects are better than any number of *e'*-objects.[12]
>
> <div align="right">(Arrhenius and Rabinowicz 2015: 232)</div>

In their writings on the topic, Arrhenius and Rabinowicz (2015) have made the following interesting discovery:

> *Observation 5*: Suppose that "is at least as good as" is a complete and transitive relation on the domain. Then, in any finite sequence of objects in which the first element is weakly superior to the last element, there exists at least one element that is weakly superior to its immediate successor.[13]
>
> <div align="right">(Arrhenius and Rabinowicz 2015: 237)</div>

Arrhenius and Rabinowicz's discussions are framed in terms of superiority; however, for the current discussion, it is more fitting to talk about strong and weak *inferiority*:[14]

> *Strongly inferior*: An object *e* is strongly inferior to an object *e'* if and only if *e* is worse than any number of *e'*-objects.

> *Weakly inferior*: An object *e* is weakly inferior to an object *e'* if and only if for some number *m*, *m e*-objects are worse than any number of *e'*-objects.

We can now see that it is plausible that excruciating torture may be weakly inferior to the mild discomfort of a mosquito bite: two years of torture would always be worse than the discomfort of a mosquito bite, no matter how long the mosquito bite persisted.

I also take it that the results arrived at by Arrhenius and Rabinowicz are true for weak inferiority. The mirror image of the proof for Observation 5 should provide us with the proof for Observation 5'.

> *Observation 5'*: Suppose that "is at least as good as" is a complete and transitive relation on the domain. Then, in any finite sequence of objects in which the first element is weakly inferior to the last element, there exists at least one element that is weakly inferior to its immediate successor.

If we accept that experiences at the opposite ends of the spectrum are related by weak inferiority, which seems to fit well with Temkin's characterization of their relation, then it follows that somewhere along the spectrum, one experience will be weakly inferior to the next. This surprising result can be accounted for by the fact that there is a difference in kind between two adjacent items in the spectrum.[15] Furthermore, it can be indeterminate where this change in kind takes place.[16]

Arrhenius and Rabinowicz's findings on superiority thus strengthen the support for the claim I made in the first part, i.e., that there is a

difference in kind between two adjacent experiences. It has been shown that two adjacent experiences in the spectrum are related by a weak inferiority relation but it has not been shown that the chain of "better than" relations is broken. Even if e_k is weakly inferior to e_{k+1}, it does not follow that e_k is worse than e_{k+1}. The existence of weak inferiority in the spectrum should make us doubt View One, but the view has yet not been ruled out on formal grounds.

If Temkin's argument is to work, View One must apply to all adjacent experiences in the spectrum. View One claims that for any pain e_i, a slightly less intense experience that is twice (or three or five times) as long is worse. From this, it follows that for any two items in the spectrum, e_k and e_{k+1}, e_k is better than e_{k+1}. According to the definition of weak inferiority, for some number m, $m\ e_k$ objects are worse than any number of e_{k+1} objects; e_k may be better than e_{k+1}, and yet $m\ e_k$ may be worse than any number of e_{k+1}.

However, with one very reasonable assumption, we reach a contradiction. According to Temkin, e_k is very similar to e_{k+1}, and if that is the case, it seems reasonable that if we prolong both experiences so that they become, for example, twice as long, then clearly e_k would still be better than e_{k+1}. If 10 years of intense headache is better than 20 years of a slightly less intense headache, then clearly 20 years of intense headache must be better than 40 years of slightly less intense headache. Or, more generally, if e_k is better than e_{k+1}, then $k\ e_k$ must be better than $k\ e_{k+1}$. This proportionality assumption entails that $m\ e_k$ is better than $m\ e_{k+1}$, which contradicts the result arrived at through the definition of weak inferiority: $m\ e_k$ objects are worse than any number of e_{k+1} objects. This contradiction can be avoided by rejecting View One.[17]

The generality of the proportionality assumption can be questioned. It is doubtful if it is true for all contexts, and it may have boundaries such that it is not true for all values of k. For example, if e_k and e_{k+1} are two pleasurable experiences, it is possible that one has a diminishing marginal value while the other does not or that both have a diminishing marginal value but to different degrees. If that is the case, it is possible that e_k is better than e_{k+1} yet $k\ e_k$ is worse than $k\ e_{k+1}$.[18] The validity of the proportionality assumption can also be questioned when applied to extremely short experiences. One second of extreme pain may be better than a day of moderate pain, but 100 seconds of extreme pain may not be better than 100 days of moderate pain.[19]

However, for this context, i.e., for spectra of the kind discussed by Temkin, the proportionality assumption is plausible. We prefer the more intense experience to the less intense just because the latter is a longer experience – it is a trade-off we want to make. In this context, when k is larger than 1, the less intense experience becomes even proportionally longer than the more intense experience. This seems to be a trade-off that is in our favor. In fact, the consideration that speaks in favor of

the proportionality assumption seems similar to what speaks in favor of View One. According to View One, a negative experience will be better than a slightly less intense but much longer experience, so proportionally extending the experience should lead to the same result.

In other words, if one sets out to argue that the proportionality assumption is mistaken in this specific context, one will also struggle to defend View One. The intuitions that support the proportionality assumption in the context of Temkin's spectrum argument are hard to rebut since these concern views of trade-offs that seem central to the spectrum argument.

5 Why We Should Accept Indeterminacy

I have argued that Temkin's most convincing formulation of a spectrum argument does not lead to the conclusion that we should reject the transitive property of "better than." First, I argued that it is possible that somewhere in the spectrum there is a difference in kind between two adjacent experiences, and thus View One is not applicable between these experiences. Temkin's reply to this line of reasoning is not convincing, especially if one takes into consideration that it can be indeterminate where in the spectrum this change takes place.

Second, I explained why we should reject View One. It is not only that it is incompatible with the other three views; it was also shown that our knowledge of the structure of value relations provides a neat explanation as to how spectrum arguments can be blocked. Central to this explanation was the concept of weak inferiority. It could, of course, be objected that the possibility that one experience is weakly inferior to its adjacent experience is just as counterintuitive as the possibility that "better than" is not transitive. The explanation I have given is consequently not to be preferred to the explanation proposed by Temkin. Furthermore, the counterintuitive result may serve as another reason to reject View Four, since it is only if we accept transitivity that we can arrive at the result that one experience is weakly inferior to its adjacent experience. However, one must remember that there is no clear-cut boundary between experiences that differ in kind. It may be indeterminate where this change takes place since the borderline can be vague. Acknowledging the role of indeterminacy should thus make this strategy more appealing. Furthermore, the result is only counterintuitive if we believe that weak inferiority implies that the two experiences must differ greatly in value. This assumption might be premature, however, as Arrhenius and Rabinowicz (2015: 536) note: "contrary to appearances, *e* might still be weakly superior to *e′*, even though it is better only by a small margin." That is, even if they do not differ greatly in value, there can be a difference in kind between the two adjacent experiences. Given such a difference, one can doubt whether View One holds for these two experiences: if one experience can

be weakly inferior to its adjacent experience, then it might be that these experiences are so different that View One does not apply.

However, an interesting result from Toby Handfield and Wlodek Rabinowicz (2018) may make my suggestion seem less plausible. They have argued that the indeterminacy needs to be of a certain kind to block the result of spectrum arguments, i.e., it needs to be radical. This is how they characterize radical indeterminacy that blocks a spectrum argument involving harms:

> Two harm types admit *radical indeterminacy* if and only if, for at least one of them, H, for all quantities K′ of H at least as large as some threshold K, it is indeterminate for every sufficiently large quantity of the other type, H′, whether K′H is worse than this quantity of H′.
> (Handfield and Rabinowicz 2018: 2384)

While weak inferiority and indeterminacy are sufficient to explain why a specific spectrum argument will fail, radical indeterminacy is needed to block the possibility of constructing a new spectrum argument. That is, even if there is a break in the spectrum and it is indeterminate where this takes place, it could be possible to construct a new spectrum by multiplying the number of items in the spectrum so that the indeterminacy disappears. With a sufficient number of items, it should become determinate what value relation obtains between the items, so a new spectrum argument can be made. The indeterminacy must consequently be radical to avoid this possibility. What makes their findings relevant here is that they argue that radical indeterminacy is hard to accept, making the strategy I have proposed less appealing. Handfield and Rabinowicz (2018) present two reasons meant to support the claim that it is hard to accept radical indeterminacy. It is, however, not clear that these considerations put in doubt the existence of radical indeterminacy. First, they discuss whether "indeterminacy that arises from a vague threshold in one relevant dimension must eventually be overwhelmed by a large enough difference in a second relevant dimension" (Handfield and Rabinowicz 2018: 2384). If that were the case, then appealing to indeterminacy to block spectrum arguments would be a dead end. However, they admit that no research suggests that indeterminacy must behave in this manner and that they lack robust intuitions on this matter. This reason should thus not be attributed too much weight. The second reason as to why radical indeterminacy is expressed as follows:

> [R]adical indeterminacy arguably does not do very much to directly address the paradoxical features of spectra that exhibit radical inferiority between the first harm type and the last. As we have seen, if we assume that the betterness ranking across such a spectrum is *complete*, then there are adjacent, qualitatively very similar harm types in

the spectrum such that one is radically inferior to the other. This is still counterintuitive, whether or not it is indeterminate where this point occurs. If it indeed is indeterminate, then – as we have seen – radical indeterminacy will have to obtain between some adjacent types in the spectrum. Alternatively, if there is no radical inferiority between any two adjacent types, then we know that the only way to avoid violating transitivity is if at some point in the spectrum there is radical incommensurability between two adjacent types. Again, this remains surprising, even if it is indeterminate where this point occurs.

(Handfield and Rabinowicz 2018: 2384)

This is indeed true, but it is not a reason to reject radical indeterminacy; rather, it is a statement concerning the difficulties of making sense of spectrum arguments. It is true that spectrum arguments force us to accept some counterintuitive claims. In this chapter, I have argued that if we find intransitivity too hard to accept, we should, in the terminology of Handfield and Rabinowicz, accept radical indeterminacy.

Furthermore, Rabinowicz (2021) has recently argued that radicality of this kind may not, after all, be too difficult to explain. If we accept his fitting-attitudes analysis of value relations it is, for example, possible to account for some forms of such radicality.[20]

6 Spectrum Arguments and Non-conventional Value Relations

Throughout this chapter, I have assumed that the "at least as good" relation fully exhausts the domain of value relations. As evident in this volume, this assumption can and has been questioned. Several arguments have been presented to show that there might be additional sui generis value relations (e.g., Chang 2002; Rabinowicz 2008). Notions such as "parity," "incomparability," "imprecision," and "rough comparability" have all been invoked to attempt to dismantle the tension in spectrum arguments. For example, it has been argued that there is no need to assume that two adjacent experiences either are related by an at least as good as relation or are incomparable, as they might just as well be on a par. This relation between two adjacent experiences would block the spectrum argument (see Qizilbash 2007; Chang 2016), though it is not a promising strategy. First, much doubt has been expressed as to whether such a relation is plausible. Second, as Anders Herlitz (2020b) has shown, Handfield and Rabinowicz's arguments applied explicitly to parity allow us to conclude that radical parity must be confused since it is in tension with how parity is usually characterized.[21]

The approach taken here could be understood as part of a larger endeavor. My argument could be understood as part of the attempt to fully account for all value relations by referring to the concept of "at

least as good." Several arguments put pressure on the assumption that "at least as good" fully exhausts the domain of value relations. These arguments, however, face the challenge of explaining why phenomena such as "parity," "incomparability," "imprecision," and "rough comparability" cannot be accounted for in terms of the vagueness of "at least as good." I and others have argued that this more conservative approach is just as successful in correctly depicting the axiological landscape (e.g., Broome 1997; Andersson 2017; Elson 2017). It has now been shown that this more conservative approach can also help explain why "better than" is a transitive relation, so there is no need to introduce any further value relations to arrive at this result.

To give an example, my objection to Temkin's argument resembles Derek Parfit's view of a similar matter. When discussing the related topic of the Repugnant Conclusion, Parfit (1984: 110) argues that the notion of imprecision allows us to avoid the conclusion that

> [c]ompared with the existence of many people who would all have some very high quality of life, there is some much larger number of people whose existence would be better, even though these people would all have lives that were barely worth living.

While he refers to something similar to weak inferiority to avoid this conclusion, he also needs to introduce the notion of "imprecision." This approach could perhaps successfully show why we need not accept Temkin's conclusion that "better than" is not transitive. However, my account has the advantage of only appealing to the familiar notions of vagueness and indeterminacy; there is no need to introduce a further mysterious notion.

7 Conclusion

This chapter's result is important because it clarifies how Temkin's four views are incompatible. For those who find it hard to reject Views Two, Three, and Four, an explanation has now been given as to why View One ought to be rejected: There is a difference in kind between experiences at the opposite ends of the spectrum, and this difference arises between two adjacent experiences. Since there is such a difference in kind, View One is not applicable and the chain of betterness relations does not hold throughout the spectrum. The findings regarding the structure of value arrived at by Arrhenius and Rabinowicz neatly support the claim that if we accept Views Two, Three, and Four, we must reject View One. One experience will be weakly inferior to its adjacent experience, so View One is not applicable. This weak inferiority can be accounted for by the fact that the experiences are of a different kind. Furthermore, it can be indeterminate where this change takes place.

When we face incompatible claims, we need an explanation as to which claim we should reject. Temkin has explained why the transitivity of "better than" might not always hold. Now we also have an explanation as to why View One might not always hold.

Notes

1 I owe much gratitude to many for helping me develop the ideas presented here. A first version of this chapter that included the main argument was written in 2017, and several versions have since been circulated and presented on various occasions. The feedback I received from Erik Carlson during my PhD defense inspired me to write the chapter in the first place. Since then I have received helpful comments from Mattias Gunnemyr, Anders Herlitz, Jakob Werkmäster, Wlodek Rabinowicz, the audience at the Hard Cases and Rational Choice workshop in Bern in November 2017, and the audience at the Higher Seminar in Practical Philosophy in Lund at which this chapter was presented.

2 Throughout most of this chapter, I will assume that there are only three value comparatives: "better than," "worse than," and "equally as good as." If there are other value comparatives beyond these three, then it is, of course, possible that these additional comparatives are not transitive. I will return to the importance of this assumption.

3 He discusses the very specific value relation the "all things considered better than in a wide reason-implying sense." Temkin (2012: 13) describes the relation in the following way: "Roughly, on this use, outcome *A* is better than outcome *B*, all things considered, if one would have more reason to prefer *A* to be realized than *B*, from an impartial perspective."

4 An early formulation of this can be found in Temkin (1996); the argument is, however, inspired by an argument by Stuart Rachels (1998).

5 John Broome (1997) makes perhaps the most important contribution to the attempt to provide an account of the normative landscape without referring to non-conventional value relations; see also Constantinescu (2012), Andersson (2017), Elson (2017), and Broome (2021).

6 All four views are grounded in intuitions that are common in axiological theorizing. If we are to reject one of them, then this will have consequences for theories that encompass them. The practical problems of rejecting any of the views is because they are relevant when discussing trade-offs. As Temkin (2012: 23) acknowledges, practical decisions often involve trade-offs. Trade-offs are therefore central for many of our everyday decisions, as well as for life-changing decisions. The philosophical literature often connects these discussions to population axiology and the distribution of benefits, i.e., discussions that clearly involve trade-offs, but they are also relevant to more mundane situations since these also often involve trade-offs. The cost of rejecting View Four can also be accounted for by referring to rational choice theory. Value relations are often believed to guide us to rational choice among alternatives. However, if the betterness relation can be cyclical, then we will not be guided toward the best alternative. For a more nuanced view of the practical implications of rejecting View Four, see Herlitz (2020a).

7 There could be several such breaks along the spectrum but, for the objection to hold, one suffices.

8 In fact, there could be several such zones, but in line with the discussion in note 7, one zone is all that is needed for the objection to hold.

9 I take this to be a simplified version of Temkin's description of the spectrum. It is more plausible that the spectrum consists of several categories and that these categories have indeterminate boundaries. Consequently, the pains in the middle are neither very intense nor very mild. However, the important point is correct even in the simplified description: it can be indeterminate where the difference in kind is found.

10 Not only is it indeterminate whether View One applies and an experience is better than its adjacent much longer experience, but it is also indeterminate whether the same experience is worse than the much longer experience.

11 The idea of superiority in value be traced back to at least John Stuart Mill (1863).

12 James Griffin (1986) had a similar distinction in mind.

13 Very roughly, the proof of the observation starts by proving the following: "*Lemma 1*: Suppose that 'is at least as good as' is a weak order, that is, a complete and transitive relation on the domain. For any objects e, e', and e'', if e is weakly superior to e'', e is weakly superior to e' or e' is weakly superior to e''" (Arrhenius and Rabinowicz 2015: 241). Observation 5, of course, entails Lemma 1 and that Lemma 1 entails Observation 5 can also be shown. Lemma 1 is then proven by assuming that (i) e is weakly superior to e'', but (ii) e is not weakly superior to e'. To establish Lemma 1, we need to show that e' is weakly superior to e''. From (i) we know that there must exist some number m for which (1) m e-objects are better than any number of e''-objects. From (ii) we can infer that there exists some number m' for which "(2) m e-objects are not better than m' e'-objects. But then, given that 'is at least as-good as' is a complete relation, (2) implies that (3) m' e'-objects are at least as good as m e-objects. By the transitivity of 'is at least as good as', if one object is at least as good as another, which is better than some third object, then the first object is better than the third. Consequently, (3) and (1) imply that (4) m' e'-objects are better than any number of e''-objects. (4) implies that e' is weakly superior to e''" (Arrhenius and Rabinowicz 2015: 242).

14 While the idea of superiority easily translates to inferiority, it is also possible to consider similar structures for "equally as good" and non-conventional value relations, such as "parity." For more on this matter, see Henrik Andersson and Andrés Garcia (n.d.).

15 This view seems to be endorsed by James Griffin (1986: 338–340). For a helpful discussion of Griffin's view, see Mozaffar Qizilbash (2005, 2021). This view also shares some features with Derek Parfit's (1984) attempt to avoid the Repugnant Conclusion. For more on the similarities and differences, see the following under "Spectrum Arguments and Non-Conventional Value Relations" and Qizilbash (2021).

16 As Qizilbash has argued, if one finds the suggestion of one experience being inferior to its adjacent experience to be implausible, one should take into account that the boundary between these experiences may be vague. Qizilbash (2005) does not use the same terms but has framed his discussion in terms of Griffin's "discontinuity."

17 It could, of course, also be avoided by rejecting the claim that the first experience is weakly inferior to the last. However, Temkin seems to endorse this inferiority claim, which, after all, is almost exactly what his View Three states.

18 This possibility was suggested to me by Wlodek Rabinowicz.

19 I owe this example to Anders Herlitz.

20 Rabinowicz (2021) provides an explanation of radical parity and not radical indeterminacy. Roughly, radical parity can be accounted for in terms of a class of permissible preference orderings that differ from one another as to where

the break in the spectrum is located. The fact that Rabinowicz's account shares structural features with the so-called supervaluationistic account of vagueness makes the possibility of explaining the occurrence of radical indeterminacy within this framework promising. For more on the concept of "parity", see the following.

21 More specifically, it is in tension with the so-called chaining argument that is an essential part of the argument for a fourth possible value relation. See also Rabinowicz (2021) for an attempt to explain radical parity.

References

Andersson, H. (2017), *How It All Relates: Exploring the Space of Value Comparisons* (Media-Tryck).

Andersson, H., and Garcia, A. G. (n.d.), 'Millian Value Relations Generalised and Interpreted'.

Arrhenius, G., and Rabinowicz, W. (2015), 'Value Superiority', in I. Hirose and J. Olson (eds.), *The Oxford Handbook of Value Theory* (Oxford University Press).

Broome, J. (1997), 'Is Incommensurability Vagueness?', in R. Chang (ed.), *Incommepnsurability, Incomparability, and Practical Reason* (Harvard University Press), 67–89.

Broome, J. (2021), 'Incommensurateness Is Vagueness', in H. Andersson and A. Herlitz (eds.), *Value Incommensurability: Ethics, Risk, and Decision-Making* (Routledge).

Chang, R. (2002), 'The Possibility of Parity', *Ethics* 112/4: 659–688.

Chang, R. (2016), 'Parity, Imprecise Comparability and the Repugnant Conclusion', *Theoria* 82/2: 182–214.

Constantinescu, C. (2012), 'Value Incomparability and Indeterminacy', *Ethical Theory and Moral Practice* 15/1: 57–70.

Elson, L. (2017), 'Incommensurability as Vagueness: A Burden-Shifting Argument', *Theoria* 83/4: 341–363.

Griffin, J. (1986), *Well-Being: Its Meaning, Measurement and Moral Importance* (Clarendon Press).

Handfield, T., and Rabinowicz, W. (2018), 'Incommensurability and Vagueness in Spectrum Arguments: Options for Saving Transitivity of Betterness', *Philosophical Studies* 175: 2373–2387.

Herlitz, A. (2020a), 'Non-Transitive Better Than Relations and Rational Choice', *Philosophia* 48/1: 179–189.

Herlitz, A. (2020b), 'Spectrum Arguments, Parity and Persistency', *Theoria* 86/4: 463–481.

Knapp, C. (2007), 'Trading Quality for Quantity', *Journal of Philosophical Research* 32/1: 211–233.

Mill, J. S. (1863), *Utilitarianism*.

Qizilbash, M. (2005), 'Transitivity and Vagueness', *Economics and Philosophy* 21/1: 109–131.

Qizilbash, M. (2007), 'The Mere Addition Paradox, Parity and Vagueness', *Philosophy and Phenomenological Research*, 75: 129–151.

Qizilbash, M. (2021), 'On "Incommensurability," "Discontinuity" and the Repugnant Conclusion: "Imprecise Equality" or Vagueness?', in H. Andersson

and A. Herlitz (eds.), *Value Incommensurability: Ethics, Risk, and Decision-Making* (Routledge).

Parfit, D. (1984), *Reasons and Persons* (Oxford University Press).

Parfit, D. (2016), 'Can We Avoid the Repugnant Conclusion?', *Theoria* 82/2: 110–127.

Rabinowicz, W. (2008), 'Value Relations', *Theoria* 74/1: 18–49.

Rabinowicz, W. (2021), 'Can Parfit's Appeal to Incommensurabilities in Value Block the Continuum Argument for the Repugnant Conclusion?', in J. McMahan, T. Campbell, J. Goodrich, and K. Ramakrishnan (eds.), *Ethics and Existence: The Legacy of Derek Parfit* (Oxford University Press).

Rachels, S. (1998), 'Counterexamples to the Transitivity of 'Better Than'', *Australasian Journal of Philosophy* 76/1: 71–83.

Temkin, L. (1996), 'A Continuum Argument for Intransitivity', *Philosophy & Public Affairs* 25/3: 175–210.

Temkin, L. (2012), *Rethinking the Good: Moral Ideals and the Nature of Practical Reasoning* (Oxford University Press).

6 Incommensurability and Vagueness in Population Axiology

Gustaf Arrhenius

1 Introduction

Population axiology concerns how to evaluate populations in regard to their goodness, that is, how to order populations by the relations "is better than" and "is as good as." The main problem in population axiology has been to find an adequate theory about the value of populations where the number of people, the quality of their lives, and their identities may vary. This field has been riddled with impossibility results, which seem to show that our considered beliefs are inconsistent in cases where the number of people and their welfare varies.[1] There have been many creative but unfortunately failed suggestions for how to eschew these impossibility results.[2] Here I shall consider two suggestions to the effect that incommensurability or vagueness could help.[3] We shall start, however, by discussing incommensurability and Derek Parfit's famous "Repugnant Conclusion."

2 The Repugnant Conclusion and Incommensurability

The Repugnant Conclusion can be stated as follows:

> *The Repugnant Conclusion*: For any population consisting of people with very high positive welfare, there is a better population in which everyone has a very low positive welfare, other things being equal.[4]
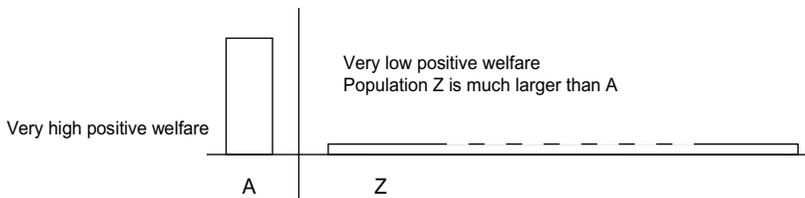
Figure 6.1 The Repugnant Conclusion.

In Figure 6.1, the width of each block represents the number of people, whereas the height represents their lifetime welfare. Dashes indicate that the block in question should be much wider than shown; that is, the population size is much larger than shown.

A population could consist of all the past, present, and future lives (in a possible world), or all the present and future lives, or all the lives during some shorter time span in the future such as the next generation, or all the lives that are causally affected by, or consequences of a certain action or series of actions, and so forth.[5]

All the lives in Figure 6.1 have positive welfare, or, as we also could put it, all the people have lives worth living. The A-people have very high welfare, whereas the Z-people have very low positive welfare.[6] The reason for this could be that in the Z-lives there are, to paraphrase Parfit, only enough ecstasies to just outweigh the agonies, or that the good things in those lives are of uniformly poor quality, e.g., eating potatoes and listening to Muzak.[7] However, since there are many more people in Z, the total sum of welfare in Z is greater than in A. Hence, a theory like Total Utilitarianism, according to which we should maximize the welfare in the world, ranks Z as better than A – an instance of the Repugnant Conclusion.[8]

A theory could avoid the Repugnant Conclusion by introducing incommensurability. It could imply that at least one large population enjoying very high welfare is incommensurable with all populations with very low positive welfare (that is, the former population is neither at least as good as, nor worse than, the latter populations) and that no population with very high welfare is at least as good as all populations with very low welfare.[9]

This is quite an unsatisfactory way to avoid the Repugnant Conclusion and arguably wasn't intended by Parfit. Those who find the conclusion repugnant are also likely to think, I surmise, that it is counterintuitive that a large population enjoying very high welfare is incommensurable with all populations with very low positive welfare. Perhaps, although I doubt it, it can be reasonably believed that some populations with very high welfare are incommensurable with some populations with very low positive welfare, but that some populations with very high welfare are incommensurable with *all* larger populations with very low positive welfare seems, given that other things are equal, clearly counterintuitive.

Some incommensurability among populations is pretty plausible, I think, if there are other considerations apart from welfarist ones that are relevant for the evaluation of populations. If some kind of pluralism is true and there are other values than welfare, then it wouldn't be remarkable if some populations turn out to be incommensurable. For example, it might be that both liberty (of some kind) and welfare should count but that there is no way of weighing gains in welfare against losses in liberty and vice versa. If one population is better than another population in respect to welfare but the other is better in respect to liberty, then these two populations would be incommensurable if the aforementioned pluralism were true.

It is important to remember, however, that we are discussing cases where other things are equal: roughly, the populations that we are comparing only differ in respect to the welfare levels of their constituent lives and size. In general, for an appeal to incommensurability to have any credibility as an argument against the adequacy condition we are discussing here, one must produce a good *welfarist* reason for incommensurability. With respect to the Repugnant Conclusion, I don't see any such reason presenting itself (more on this later).

Let's formulate an adequacy condition that avoids this unsatisfactory way of dealing with the Repugnant Conclusion:

> *Quality*: There is a perfectly equal population with very high welfare which is at least as good as any population with very low positive welfare, other things being equal.

Avoidance of the Repugnant Conclusion implies that there is at least one population with very high welfare that is at least as good as or incommensurable with all larger populations with very low welfare. Quality is in one sense logically stronger than avoidance of the Repugnant Conclusion since it rules out axiologies that imply that at least one population with very high welfare is incommensurable with all populations with very low positive welfare but none is at least as good as all such populations.[10]

It is otherwise a very weak condition. A theory that implies that most but not all large populations with very high welfare are worse than some populations with very low welfare doesn't violate Quality. Likewise, neither a theory that yields that only one perfectly equal population with very high welfare is better than all populations with very low positive welfare, nor a theory that deems all such pairs of populations to be equally good, violates Quality (nor do these theories imply the Repugnant Conclusion).

If one holds that the Repugnant Conclusion is unacceptable, then it would be odd, one might argue, to accept such theories. Arguably, the axiological intuition most people have about the relation between populations with very high and very low positive welfare is much stronger than what is captured by Quality. Perhaps we believe that if the high welfare population is sufficiently large, then such a population and any larger high welfare population is better than any very low welfare population. As true as this might be, one should remember that Quality is only a necessary and not a sufficient condition for an acceptable axiology.

3 Critical-Level Utilitarianism

Charles Blackorby, Walter Bossert, and David Donaldson have in a number of papers developed a theory, or rather a family of theories, called Critical-Level Utilitarianism.[11] John Broome has also defended a version

of this theory.[12] We are here interested in a version of this theory that incorporates incommensurability, but to get there, we need to first present the simplest version of the theory and its properties. As such, Critical-Level Utilitarianism, or CLU for short, is a modified version of Total Utilitarianism. The contributive value of a person's life is her or his welfare minus a positive critical level. The value of a population is calculated by summing these differences for all individuals in the population. CLU could thus be expressed in the following form:

$$CLU(X) = \begin{cases} \sum_{i=1}^{n} (u_i - k) & n > 0 \\ 0 & n = 0 \end{cases}$$

In the formula, n is the population size of X, and u_i is the numerical representation of the welfare of the i:th life in population X, and k is the critical level.

The critical level k is supposed to be the level at which it is axiologically neutral whether a life is created or not, what Broome calls "the neutral level for existence." Blackorby et al. and Broome don't equate this level with the welfare level of a life that is neutral *for* a person – that is, neutral welfare, an option that the classical Total Utilitarian would use. As Broome writes, "[T]he neutral level for existence [the critical level] is positive, once the zero of lifetime wellbeing is normalized at the level of a constantly neutral life."[13] Hence, since the critical level is positive, the contributive value of lives with positive welfare below the critical level is negative.

Consequently, assuming that the critical level is higher than very low positive welfare, the Repugnant Conclusion is avoided and Quality is satisfied since the value of a huge population with positive but very low welfare will be negative.

CLU violates a number of other intuitively attractive conditions, however. Here we will limit the discussion to a few that will be especially relevant when we proceed to the discussion of the incommensurability version of CLU. First, CLU violates the following attractive condition:

> *Non-Sadism*: An addition of any number of people with positive welfare is at least as good as an addition of any number of people with negative welfare, other things being equal.

It is easy to see that CLU violates Non-Sadism and that its violation of this condition becomes especially disturbing if the critical level is set high:
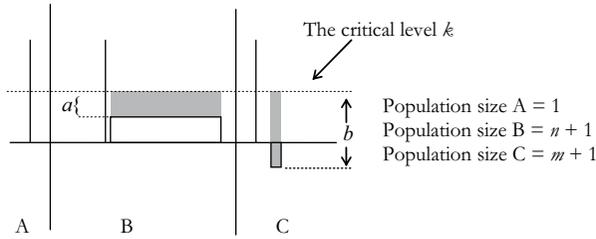
Figure 6.2 CLU and Non-Sadism.

Population A in Figure 6.2 consists of one person with welfare well above the critical level. In outcome B, we have added n people with positive welfare x. Their welfare is a units below the critical level k, as indicated in the diagram. The negative value of this addition is thus $n(x - k) = -na$, which is represented by the gray area in outcome B. In C, m people with negative welfare y have been added, $m < n$. Their welfare is b units below the critical level, $b > a$, as indicated in the diagram. The negative value of this addition is $m(-y - k) = -mb$, which is represented by the gray area in outcome C. For any values of a and b, we can find values of n and m such that $mb < na$, as in our previous case (the gray area in outcome C is smaller than the gray area in outcome B). In such cases, it is better to add the people with negative welfare rather than the people with positive welfare, a clear violation of Non-Sadism.

CLU implies especially troublesome violations of Non-Sadism, such as

> *The Very Sadistic Conclusion*: For any population with negative welfare, there is a population with positive welfare that is *worse*, other things being equal.

There is always a population with sufficiently many people with positive welfare slightly below the critical level such that the total negative value of these people is greater than that of a given population made up of people with negative welfare. This holds irrespective of how much people suffer and of how many they are. Thus, CLU implies the Very Sadistic Conclusion and violates

> *Weak Non-Sadism*: There is a negative welfare level and a number of lives at this level such that an addition of any number of people with positive welfare is at least as good as an addition of the lives with negative welfare, other things being equal.

Finally, to avoid the worst versions of the Repugnant Conclusion where the Z-lives are just barely worth living, the critical level has to be set fairly high. If the critical level is set at the level of a fairly good life, however, we get the Very Sadistic Conclusion that a population with horribly tormented lives may be better than a population with fairly good lives. In

this sense, CLU is caught between the sadistic Scylla and the repugnant Charybdis.[14]

4 Incomplete CLU

Blackorby et al. acknowledge the aforementioned and related problems, and they have suggested an interesting solution involving incommensurability.[15] Here's the idea. Instead of using one critical level, they propose an interval of critical levels when comparing populations of different size. The interval of critical levels is assumed to be between zero and a positive welfare level α. The idea is that a population A is better than another population B if and only if A is better than B for all critical levels in the interval. If A is better than B for only some critical levels in the interval, and B is better than A for some other critical levels, then A and B are incommensurable; that is, A is neither at least as good as B nor worse than B. They call this principle Incomplete Critical-Level Utilitarianism (Incomplete CLU for short).[16] We shall formulate this principle as follows:

> *Incomplete CLU*: Population A is at least as good as B if and only if $CLU(A) \geq CLU(B)$ for all k, $0 \leq k \leq \alpha$, where α is the upper bound of the critical interval.

Given the standard properties of "at least as good as" it follows from the above that if A is better than B for only some critical levels, and B is better than A for some other critical levels, then A and B are incommensurable (since then it is neither true that A is at least as good as B, nor that B is at least as good as A).[17] As Blackorby et al. point out, Incomplete CLU avoids the Repugnant Conclusion and the Sadistic Conclusion.[18]

It does this in a questionable manner, however, since it does this by rendering all the populations involved incommensurable. For example, let's say that A is a large population with very high welfare and total welfare x and that B, C, D, and so forth, are populations with very low welfare and with total welfare greater than x. Assume that very low welfare is below the maximal critical level α. If $k = 0$, then CLU is equivalent to Total Utilitarianism and, consequently, $CLU(A) < CLU(B)$, $CLU(A) < CLU(C)$, and so forth. If $k = \alpha$, on the other hand, then the value of populations B, C, D, and so forth are going to be negative whereas the value of A is going to be positive. Thus, Incomplete CLU renders all populations B, C, D, and so forth, with very low welfare and with total welfare greater than x as incommensurable with A.

There are at least two problems with this result. First, it is a rather extreme result. Along the lines of the discussion of the Repugnant Conclusion and the Quality Condition, even if (and that is a big "if") it can be reasonably believed that some populations with very high welfare are incommensurable with some populations with very low positive

welfare, it is counterintuitive that *all* of the high welfare populations are incommensurable with *all* larger populations with very low positive welfare but higher total welfare, given that other things are equal.

Consider for example a low welfare B-population with total welfare just slightly higher than A. If one rejects the Repugnant Conclusion, then one will of course consider A better than B since one would also think so if the total welfare was much higher in B than in A. If one accepts the Repugnant Conclusion, one might be a bit unsure about this case (if one isn't a hard-nosed Total Utilitarian who of course will rank B as better than A), perhaps even think that there is some kind of incommensurability here, but somewhere along a spectrum of low welfare populations with successively higher total welfare, one will judge the low welfare population as better than A. Hence, there seems to be no intuitive space for a wholesale classification of all the involved populations as incommensurable. It is a kind of "greedy" incommensurability (more on this below).

The second problem concerns the source of the incommensurability. As we pointed out previously, for an appeal to incommensurability to have any credibility as an argument against the adequacy conditions involved in the impossibility theorems, and in particular for welfarists such as Blackorby et al., one must produce a good *welfarist* reason for incommensurability. There are, I think, three plausible sources of incommensurability among populations that are relevant in respect to the adequacy conditions involved in the impossibility theorems in population ethics.

The first apparent source of incommensurability from a welfarist perspective has to do with comparisons of different people's welfare: one can reject interpersonal comparability of welfare. This move certainly yields extensive incommensurability among populations, but it would be, I surmise, too extensive to be plausible, and rejecting interpersonal comparability of welfare leads to Arrowian impossibility theorems.[19] At any rate, Blackorby et al. are obviously not denying interpersonal comparisons of welfare since their theories presuppose the meaningfulness of such comparisons.

The second welfarist source of incommensurability can be found in the orderings of lives. It seems possible that there are pairs of lives such that we cannot say whether one is better than the other, nor can we say whether they are equally good. In real life, such cases are of course numerous because of epistemological problems. But it also seems possible that there are lives whose welfare is incommensurable in principle. An example might be a life that has somewhat more joy and true friendships but somewhat less important achievements than another. This kind of incommensurability would carry over to population axiology. Let's say that we have two populations of the same size consisting of lives whose welfare is incommensurable; that is, we cannot determine whether the lives in one of the populations have at least as high welfare as the lives in the other populations, and *vice versa*. Other things being equal, these populations are incommensurable.

If there are lives whose welfare is incommensurable, then the relation "has at least as high welfare" is not complete over the set of all possible lives and we will only have a quasi-ordering of possible lives. The adequacy conditions discussed here only presuppose a quasi-ordering of lives, however. Blackorby et al., on the other hand, presuppose completeness, since they assume that welfare can be measured on a ratio-scale and measurement on this scale, in turn, presupposes the completeness of the relation "has at least as high welfare as" over the set of lives whose welfare is measured. In other words, incompleteness in the ordering of lives in regard to welfare is not available for Blackorby et al. as a source of incommensurability among populations.

More importantly, incompleteness in the ordering of lives would hardly yield the kind of incommensurability among populations that Incomplete CLU implies. It would be bizarre to claim that lives with very high welfare are incomparable in regard to welfare with lives enjoying very low positive welfare, or, for that matter, that hellish lives are incomparable with flourishing lives. Hence, the plausible incommensurability among lives that may exist cannot support the kind of incommensurability to which Blackorby et al. resort and can hardly be wielded as an argument against the adequacy conditions in population ethics that we have discussed here.

The third somewhat plausible source of incommensurability from a welfarist point of view is the following: we might find it impossible to weigh a greater number of small gains in welfare against a smaller number of great losses. And one might think that the Repugnant Conclusion involves such a weighing since we are weighing lives with very high positive welfare against lives with very low positive welfare and that a move from A to B would involve a great loss for the A-people and just a slight gain for the B-people.

There are several problems with this reasoning, however. Let me bring up some of the more important ones. First, talk about "gains" and "losses" might be quite misleading in the present context – it sounds like we are "taking" welfare from some well-off people and "giving" it to some worse-off people whom we are considering "moving" from an existing population to another population. This need not be the case, however. For example, the compared populations might be two future populations consisting of different people. So no one is gaining or losing, at least as long as we don't think one can gain from being created and lose from not being created. Especially the latter claim is a tall order to defend.[20]

Second, and more interestingly, this won't generate the kind of incommensurability generated by Incomplete CLU and points to a general problem for an appeal to incommensurability in population ethics. With this source, it would be counterintuitive if the zone of incommensurability wasn't limited but "greedy."[21] Even if we could get incommensurability when we weigh a greater number of small gains in welfare against a

smaller number of great losses, it seems likely that we should get out of the zone of incommensurability at some point if we increase the number of people that gain. Applied to the Repugnant Conclusion, we would then get incommensurability in some zone of the sequence of populations with very low positive welfare but increasing total welfare. At some point, however, when the number of people gaining is so much greater than the number of people losing, one might not find it a hard call anymore to judge that the gains outweigh the losses. So this source will generate a limited incommensurability among populations and not the wholesale incommensurability implied by Incomplete CLU.

We should also consider the incommensurability implied by Incomplete CLU in cases involving lives with negative welfare. Here it yields even more counterintuitive results, I'm afraid. Consider the Very Sadistic Conclusion. For any number n of hellish lives, there is a number $m > n$ of lives with positive welfare just below the highest critical level, such that a population consisting of the hellish lives is incommensurable with the population consisting of the lives with positive welfare. Thus, Incomplete CLU avoids the Very Sadistic Conclusion but, again, in a disputable manner since the population with hellish lives is not incommensurable with the population consisting of the lives with positive welfare but clearly worse.

Lastly, although Incomplete CLU can avoid the Repugnant Conclusion and the Very Sadistic Conclusion, it cannot avoid violating Quality and Weak Non-Sadism. According to the former condition, there is at least one perfectly equal population with very high welfare that is at least as good as all populations with very low welfare, other things being equal. Incomplete CLU implies that for any population with very high welfare, there is a population with very low positive welfare that is incommensurable with or better than the former.

Weak Non-Sadism yields that there is a negative welfare level and a number of lives at this level such that an addition of lives with positive welfare is at least as good as an addition of the lives with negative welfare, other things being equal. According to Incomplete CLU, for any addition of lives with negative welfare, irrespective of how much people suffer and of how many they are, there is an addition of lives with positive welfare that renders the compared populations incommensurable.

5 Vague CLU

Broome suggests a structurally similar solution to Blackorby et al.'s, but instead of incommensurability, he introduces vagueness in the ordering of populations. The idea is that the critical level is vague.[22] Broome adopts a supervaluationist account of vagueness. According to this account, we can say that a population A is better than B if and only if it is better under every sharpening (or interpretation) of the critical level; otherwise, it is indeterminate whether it is better.

We can, with Broome, assume that the sharpenings of the critical level lie in an interval between zero and a positive welfare level α.[23] Hence, we can say that population A is better than B if and only if *CLU*(A) > *CLU*(B) for all sharpenings k, $0 \leq k \leq \alpha$; otherwise, it is indeterminate whether A is better than B. The structural similarity with Blackorby et al.'s proposal should now be evident. Let's call this theory Vague Critical-Level Utilitarianism or Vague CLU for short.

Does it help substituting vagueness for incommensurability? Well, the implications will be similar to those of Incomplete CLU pointed out earlier, although instead of the populations being rendered incommensurable, it is indeterminate whether they are better or not. For example, according to Vague CLU, for any addition of lives with negative welfare, irrespective of how much people suffer and of how many they are, there is an addition of lives with positive welfare such that it is indeterminate whether is better to add the lives with positive welfare rather than the hellish lives. Of course, that is exactly what we think we can say with confidence, and the reason why Weak Non-Sadism is so compelling. What we believe is that the addition of lives with positive welfare is determinately better than the addition of the lives with very negative welfare. All the other troublesome implications of Incomplete CLU discussed previously can be reproduced for Vague CLU, including the greediness problem. An appeal to vagueness doesn't help here.[24]

Notes

1 The informal Mere Addition Paradox in Parfit (1984: 419ff) is the *locus classicus*. For an informal proof of a similar result with stronger assumptions, see Ng (1989: 240). A formal proof with slightly stronger assumptions than Ng's can be found in Blackorby and Donaldson (1991). For theorems with much weaker assumptions, see my (1999), (2000a), and especially (2000b), (2001), (2011), (forthcoming).

2 See Arrhenius (2000b), (forthcoming).

3 For some other suggestions (most of them presented just as a solutions to a specific simple impossibility result) and criticism of these, see e.g., Arrhenius (2016), (2021); Hájek and Rabinowicz (2021); Handfield and Rabinowicz (2018); Parfit (2014), (2016); Qizilbash (2000), (2021); Rabinowicz (2021).

4 Here's how Parfit (1984: 388) formulates the conclusion: "For any possible population of at least ten billion people, all with a very high quality of life, there must be some much larger imaginable population whose existence, if other things are equal, would be better, even though its members have lives that are barely worth living." Hence, our formulation is more general than his. The *ceteris paribus* clause in the formulation is meant to imply that the compared populations are roughly equal in all other putatively axiologically relevant aspects apart from individual welfare levels.

5 More exactly, a population is a finite set of lives in a possible world. A, B, C,… $A_1, A_2,…, A_n$, A∪B, and so on, denote populations of finite size. We shall adopt the convention that populations represented by different letters, or the same letter but different indexes, are pairwise disjoint. For example, A∩B = $A_1 \cap A_2$ = ∅. We shall assume that for any natural number n and any welfare level X,

there is a possible population of *n* people with welfare X (for a discussion of this *No-Limit Assumption*, see Arrhenius (2000b: ch. 3), (forthcoming)).

6 We shall say that a life has *neutral welfare* if and only if it is equally as good for the person living it as a neutral welfare component and that a life has *positive* (*negative*) welfare if and only if it has higher (lower) welfare than a life with neutral welfare. A welfare component is neutral relative to a certain life *x* if and only if *x* with this component has the same welfare as *x* without this component. There are a number of alternative definitions of a neutral life in the literature, many of which would also work fine in the present context. For a discussion, see Arrhenius (2000b), (forthcoming: chs. 2 and 9), Broome (1999), (2004), Bykvist (2007: 101), and Parfit (1984: 357–358 and appendix G). Notice also that we actually don't need an analysis of a neutral welfare in the present context but rather just a criterion, and the criterion can vary with different theories of welfare.

7 See Parfit (1984: 388) and Parfit (1986: 148). For a discussion of different interpretations of the Repugnant Conclusion see Arrhenius (2000b), (forthcoming) and Parfit (1984), (2014), (2016).

8 Throughout this chapter "better" means "better, all things considered," if not otherwise indicated.

9 We could extend this definition for the purpose of this chapter to include Ruth Chang's (2002), (2005) (2021) proposal that there is a forth value-relation: "on a par" – that is, define incommensurability as "neither at least as good as, nor on a par, nor worse than." Likewise for Parfit's (2014), (2016) idea of imprecise equality. For a discussion of the latter, see Arrhenius (2016), (2021).

10 Quality doesn't imply avoidance of the Repugnant Conclusion, and vice versa, but given full comparability among populations and satisfaction of a weak dominance condition, avoidance of the Repugnant Conclusion implies satisfaction of Quality. For a proof, see Arrhenius (2000b), (forthcoming: appendix B).

11 See Blackorby, Bossert, and Donaldson (1995), (1997), (2005) and Blackorby and Donaldson (1984). These authors also propose a more refined version of CLU where the contributive value of people's welfare is dampened by a strictly concave function. This modification has no relevance for the arguments made here. For a precursor to Blackorby et al.'s theory, see Kavka (1982).

12 Broome (2004). Broome develops his theory from some foundational ideas in axiology, drawing on some of the results from his earlier book *Weighing Goods* (1991), and carefully working toward a comprehensive population axiology. Unfortunately, space limitations prevent me from doing full justice to Broome's rich book, and I shall here focus on the implications of the population axiology that he formulates in the final chapters.

13 Broome (2004: 259). See also the discussion of Broome's definition of a life with neutral welfare in Arrhenius (forthcoming: section 2.2.3). Blackorby et al.'s zero represents the welfare of a life with neutral welfare where neutral welfare is the limit that the welfare of all kinds of lives approaches as they get shorter and approach zero length (2005: 25). Cf. Arrhenius (forthcoming: section 2.2.3, fn. 27).

14 Broome (2004: 213–214) acknowledges this dilemma: "To ease the discomfort of the … repugnant conclusion, I suggested that the neutral level might be a reasonably good level of life. If this is so, the … [Sadistic Conclusion] is more poignant. A life just below the neutral level will also be reasonably good. It may contain no suffering, so the [distribution with such lives] … may contain no suffering. Yet according to the [Sadistic Conclusion], this distribution

is supposed to be worse than a distribution that contains a very great deal of suffering. – The implausibility arises principally from the interaction between the positive [repugnant conclusion] and [the Sadistic Conclusion]. To make ourselves comfortable with the positive one, we need a high neutral level, but this makes us uncomfortable with the [Sadistic Conclusion]... – The best we can hope for is a compromise that reduces the strain on our intuitions to a tolerable level. I see no guarantee that this is possible."

15 See Blackorby et al. (1997: 216–219), (2005: ch. 7).
16 See Blackorby et al. (1997: 216–219, 226). That the critical levels consist of all numbers between zero and a positive welfare level is not part of Blackorby et al.'s definition of Incomplete CLU, but they assume this in their discussion of it. See also Blackorby et al. (2005: 219–221, 248–252).
17 It also follows that if A is at least as good as B but B isn't at least as good as A, then A is better than B; and if they are both at least as good as the other one, then they are equally good.
18 See Blackorby et al. (1997: 218–219, 226) and Blackorby et al. (2005: 221).
19 For a discussion, see Arrhenius (2000b), (2000a), (forthcoming).
20 For a discussion of whether it can be better for a person to exist than not to exist, see, e.g., Arrhenius (2015); Arrhenius and Rabinowicz (2010), (2015).
21 Broome (2004: ch. 12) uses the term "greedy" to describe incomparability that spreads way beyond its intuitive limitations. See also Arrhenius (forthcoming); Handfield and Rabinowicz (2018); Herlitz (2020).
22 Broome (2004: 180). Cf. Broome (2021).
23 Broome (2004: 180, 259).
24 I would like to thank Anders Herlitz and Henrik Andersson for helpful discussions. Financial support from Riksbankens Jubileumsfond (grant M17-0372:1) is gratefully acknowledged.

References

Arrhenius, G. (1999). 'An Impossibility Theorem in Population Axiology with Weak Ordering Assumptions', in R. Sliwinski (Ed.), *Philosophical Crumbs: Essays Dedicated to Ann-Mari Henschen-Dahlquist on the Occasion of her Seventy-Fifth Birthday* (pp. 11–21). Uppsala: Department of Philosophy, Uppsala University: Uppsala Philosophical Studies.

Arrhenius, G. (2000a). 'An Impossibility Theorem for Welfarist Axiologies', *Economics and Philosophy*, 16(02), 247–266.

Arrhenius, G. (2000b). *Future Generations: A Challenge for Moral Theory*. Retrieved from http://www.diva-portal.org/smash/record.jsf?pid=diva2:170236

Arrhenius, G. (2001). 'What Österberg's Population Theory Has in Common With Plato's', in *Omnium-gatherum. Philosophical Essays Dedicated to Jan Österberg on the Occastion of his Sixtieth Birthday* (Vol. 50, pp. 29–44). Uppsala: Department of Philosophy, Uppsala University: Uppsala Philosophical Studies.

Arrhenius, G. (2011). 'The Impossibility of a Satisfactory Population Ethics', in H. Colonius and E. N. Dzhafarov (Eds.), *Descriptive and Normative Approaches to Human Behavior, Advanced Series on Mathematical Psychology* (pp. 1–26). London: World Scientific Publishing Company.

Arrhenius, G. (2015). 'The Affirmative Answer to the Existential Question and the Person Affecting Restriction', in I. Hirose and A. Reisner (Eds.), *Weighing and Reasoning: Themes from the Philosophy of John Broome* (pp. 110–125). Oxford; New York, NY: OUP Oxford.

Arrhenius, G. (2016). 'Population Ethics and Different-Number-Based Imprecision', *Theoria*, *82*(2), 166–181. https://doi.org/10.1111/theo.12094

Arrhenius, G. (2021). 'Population Ethics and Conflict-of-Value Imprecision', in J. McMahan, T. Campbell, and J. Goodrich (Eds.), *Ethics and Existence: The Legacy of Derek Parfit*. Oxford: Oxford University Press.

Arrhenius, G. (forthcoming). *Population Ethics: The Challenge of Future Generations*. Oxford: Oxford University Press.

Arrhenius, G., and Rabinowicz, W. (2010). 'Better to Be than not to Be?', in H. Joas and B. Klein (Eds.), *The Benefit of Broad Horizons: Intellectual and Institutional Preconditions for a Global Social Science* (pp. 399–414). Leiden, Boston: BRILL.

Arrhenius, G., and Rabinowicz, W. (2015). 'The Value of Existence', in I. Hirose and J. Olson (Eds.), *The Oxford Handbook of Value Theory* (pp. 424–444). Oxford/New York: Oxford University Press.

Blackorby, C., Bossert, W., and Donaldson, D. (1995). 'Intertemporal Population Ethics: Critical-Level Utilitarian Principles', *Econometrica*, *63*(6), 1303–1320. https://doi.org/10.2307/2171771

Blackorby, C., Bossert, W., and Donaldson, D. (1997). 'Critical-Level Utilitarianism and the Population-Ethics Dilemma', *Economics and Philosophy*, *13*(02), 197–230. https://doi.org/10.1017/S026626710000448X

Blackorby, C., Bossert, W., and Donaldson, D. J. (2005). *Population Issues in Social Choice Theory, Welfare Economics, and Ethics*. New York: Cambridge University Press.

Blackorby, C., and Donaldson, D. (1984). 'Social Criteria for Evaluating Population Change', *Journal of Public Economics*, *25*(1–2), 13–33. https://doi.org/10.1016/0047-2727(84)90042-2

Blackorby, C., and Donaldson, D. (1991). 'Normative Population Theory', *Social Choice and Welfare*, *8*(3), 261–267.

Broome, J. (1991). *Weighing Goods: Equality, Uncertainty and Time*. Oxford: Blackwell.

Broome, J. (1999). *Ethics out of Economics*. Cambridge: Cambridge University Press.

Broome, J. (2004). *Weighing Lives*. Oxford: Oxford University Press.

Broome, J. (2021). 'Incommensurateness is Vagueness', in H. Andersson and A. Herlitz (eds.), Value Incommensurability: Ethics, Risk, and Decision- Making (Routledge).

Bykvist, K. (2007). 'The Good, The Bad, and the Ethically Neutral', *Economics and Philosophy*, *23*(01), 97–105. https://doi.org/10.1017/S0266267107001253

Chang, R. (2002). 'The Possibility of Parity', *Ethics*, *112*(4), 659–688.

Chang, R. (2005). 'Parity, Interval Value, and Choice*', *Ethics*, *115*(2), 331–350.

Chang, R.,   (2021). 'Are Hard Cases Vague Cases?', in H. Andersson and A. Herlitz (eds.), Value Incommensurability: Ethics, Risk, and Decision-Making (Routledge).

Hájek, A., and Rabinowicz, W. (2021). Degrees of Commensurability and the Repugnant Conclusion. *Noûs, n/a(n/a)*. https://doi.org/10.1111/nous.12388

Handfield, T., and Rabinowicz, W. (2018). 'Incommensurability and Vagueness in Spectrum Arguments: Options for Saving Transitivity of Betterness', *Philosophical Studies*, *175*(9), 2373–2387. https://doi.org/10.1007/s11098-017-0963-9

Herlitz, A. (2020). 'Spectrum Arguments, Parity and Persistency', *Theoria*, *86*(4), 463–481. https://doi.org/10.1111/theo.12249

Kavka, G. S. (1982). 'The Paradox of Future Individuals', *Philosophy & Public Affairs*, *11*(2), 93–112.

Ng, Y.-K. (1989). 'What Should We Do About Future Generations?', *Economics and Philosophy*, *5*(02), 235–253.

Parfit, D. (1984). *Reasons and Persons* (1991st ed.). Oxford: Clarendon.

Parfit, D. (1986). 'Overpopulation and the Quality of Life', in P. Singer (Ed.), *Applied Ethics* (1 edition, pp. 145–164). Oxford/New York: Oxford University Press.

Parfit, D. (2014). *How We Can Avoid the Repugnant Conclusion*. Mimeo, University of Oxford, Faculty of Philosophy.

Parfit, D. (2016). 'Can We Avoid the Repugnant Conclusion?', *Theoria*, *82*(2), 110–127. https://doi.org/10.1111/theo.12097

Qizilbash, M. (2000). 'Comparability of Values, Rough Equality, and Vagueness: Griffin and Broome on Incommensurability', *Utilitas*, *12*(2), 223–240. https://doi.org/10.1017/S0953820800002806

Qizilbash, M. (2021). 'On "Incommensurability", "Discontinuity" and the Repugnant Conclusion: "Imprecise Equality" or Vagueness?', in A. Herlitz and H. Andersson (Eds.), *Value Incommensurability: Ethic, Risk and Decision-Makings*. New York: Routledge.

Rabinowicz, W. (2021). 'Can Parfit's Appeal to Incommensurabilities in Value Block the Continuum Argument for the Repugnant Conclusion?', in T. Campbell, J. McMahan, and J. Goodrich (Eds.), *Ethics and Existence. The Legacy of Derek Parfit*. Oxford: Oxford University Press.

Part III

Incommensurability and Decision Theory

7 Nondeterminacy and Reasonable Choice

Anders Herlitz

Many have argued that some practical reasons admit of what could generally be called *nondeterminacy*: practical reasons sometimes fail to fully determine that for two comparable options, one is determinately at least as good as the other.[1] The by now most common explanations of such nondeterminacy are that it is either due to vagueness (e.g., Broome 1997, 2004, 2021; Elson 2017) or because some non-conventional comparative relation such as parity obtains between two alternatives (e.g., Chang 2002; Rabinowicz 2008). Although the question of how to explain nondeterminacy has received significant attention, little has been done to identify *which* practical reasons admit of nondeterminacy and *how* they do this, i.e., which practical reasons can entail that an agent who attempts to follow and act on these reasons will face nondeterminacy problems and how these reasons entail this. This chapter hypothesizes that clues as to which practical reasons admit of nondeterminacy and how they do this can be found by studying the challenges that nondeterminacy poses for choice theory and theories of how to act rationally when practical reasons fail to fully determine what one ought to do. It argues that when otherwise acceptable decision strategies violate basic requirements of rationality because of nondeterminacy, reasons cannot admit of it, and outlines three possible explanations of how to account for this.

The chapter is structured as follows. The first section presents, in detail, nondeterminacy and a standard revision of conventional decision criteria in light of nondeterminacy. The second, third, and fourth sections introduce some challenges that nondeterminacy poses for choice theory: dynamic choice problems, underdetermination, cyclical evaluations, and violations of basic contraction consistency. The fifth section discusses some constraints that can be used with decision strategies so that these problems are avoided. In the sixth section, four views of how to understand such constraints are discussed. There is a brief concluding section.

1 Nondeterminacy and Determinate Maximality

There are good reasons to expect that practical reasons sometimes fail to fully determine what one ought to do. Here are three arguments to that effect. First, sometimes alternatives are so different from each other that it seems implausible to expect practical reasons to guide action. Are the reasons to cater to the needs of a sick relative stronger than the reasons to defend one's country against foreign occupiers? Are the reasons to finish that last work assignment on time stronger than the reasons to help one's child with their homework? Are the reasons to order the healthy salad for lunch stronger than the reasons to order the tacos that one knows will bring much more pleasure?

Second, consider the small improvement argument (see de Sousa 1974; Raz 1986; Sinnott-Armstrong 1988; Chang 2002). Consider a situation in which someone, Cleo, must choose between a career in the diplomatic corps and a career as a civil rights lawyer. Cleo recognizes many good reasons to join the diplomatic corps: it would give her a chance to promote human rights globally, serve her country, meet many interesting people, and see the world. She also recognizes many good reasons to become a civil rights lawyer: it would give her a chance to promote the rights of minorities in her community, serve her community, keep living in her hometown where she has her friends and family, and have her own family and raise children in the same loving environment that she grew up in. To Cleo, the reasons that apply to the choice do not determine that either option is better than the other. Would anything change if one of the options was improved? Would one option be clearly better than the other if it was slightly improved? Would Cleo have more reason to join the diplomatic corps than to become a civil rights lawyer if she learned that the diplomatic salary was slightly higher than she previously thought? Cleo values money and thinks that a higher salary speaks in favor of a certain career option, but would a relatively small amount of money change how the options relate to each other? To many, this seems implausible. Small improvements of one option do not tip the scale when one compares two very dissimilar options, neither of which is superior to the other. Yet, if a small improvement does not tip the scale, the initial two options cannot have been equally good. The two options must be related to each other in some other way. We can expect to find this phenomenon, what Joseph Raz (1986) once called the "mark of incommensurability," in many areas.

Third, consider a more theoretically motivated argument. It is widely recognized in population ethics that it is impossible to reconcile widely held intuitions regarding which of two populations is better; for example: "regardless of its size, a population in which everyone leads lives barely worth living is worse than a population of one billion people leading blissful lives," and "compared to a population in which everyone leads

lives at positive welfare level *W*, a population that is twice as big in which everyone leads lives that are almost as good, at positive welfare level *W*–, is better" (see, e.g., Parfit 1984; Arrhenius 2000, forthcoming). Similarly, when one compares large benefits to few people with small benefits to many people, it is strikingly difficult not to end up with deeply counterintuitive conclusions, for example, that it is better to cure billions of mild headaches than to save a life (Temkin 1996, 2012; Norcross 1997; Rachels 1998). One promising way to solve these problems is by accepting the fact that reasons sometimes fail to fully determine that any option is best (Qizilbash 2007; Parfit 2016; Herlitz 2019a; Rabinowicz 2019).

Nevertheless, the aim of this chapter is *not* to establish that situations in which practical reasons fail to fully determine what one ought to do are possible, relevant, or common. Furthermore, I will not engage with the matter of what the best explanation for this is, whether it is sometimes indeterminate which option one has most reason to favor due to vagueness (e.g., Broome 1997), whether some alternatives are sometimes incomparable (Raz 1986), whether some alternatives are sometimes "on a par" (e.g., Chang 2002), whether options are sometimes imprecisely equally as good (Parfit 2016), or whether there might be some other non-conventional comparative relation (Rabinowicz 2008). Rather, I wish to explore the following question: if these situations are relevant, how are they relevant, and under what conditions can we expect them to arise?

While the aforementioned explanations differ in some respects, they have one thing in common that is captured by the concept of nondeterminacy:

> *Nondeterminacy: x* and *y* are *nondeterminate in their ranking* with respect to *p* (i.e., that with respect to which the ranking is made) if it is not determinately true that *x* is more *p* than *y*, not determinately true that *x* is less *p* than *y*, and not determinately true that *x* and *y* are equally as *p*.

Nondeterminacy is *irreflexive* (i.e., no item is nondeterminate in its ranking with itself), *symmetric* (i.e., if *x* is nondeterminate in its ranking with *y*, then *y* is nondeterminate in its ranking with *x*), and *non-transitive* (i.e., it is not the case that if *x* and *y* are nondeterminate in their ranking and *y* and *z* are nondeterminate in their ranking, then *x* and *z* are necessarily nondeterminate in their ranking; see Carlson 2010). Proponents of the vagueness explanation, incomparability, parity, imprecise equality, and other non-conventional comparative relations can all agree that it is sometimes the case that some options are nondeterminate in their ranking (although they of course disagree as to why).

Nondeterminacy can actualize serious challenges to rational choice. However, it is not true, as one might initially believe, that nondeterminacy

undermines the very possibility of rational choice assessments in the sense that practical reasons that admit of nondeterminacy fail to give any action guidance. To see this, consider the possibility that practical reasons provide us with the following ranking of x, y, and $x+$ (let "≈" denote "nondeterminate in their ranking" and "<" denote "determinately worse than"):

$$x \approx y; \; x+ \approx y; \; x < x+$$

It is obvious that this ranking, which incorporates nondeterminacy between x and y and between $x+$ and y, does not undermine the possibility of rational choice assessments, as x is not a rational choice and should not be chosen.

It is also easy enough to revise conventional decision criteria so that they can say precisely the foregoing. According to Amartya Sen, proponents of incomparability, parity, and other non-conventional comparative relations can say that only *maximal* options are rational, i.e., only options that are *not worse than* any alternative are rational (see, e.g., Sen 1970, 1997, 2017). x is worse than $x+$, and thus not maximal. Proponents of the vagueness explanation of nondeterminacy can say that only options that are optimal (i.e., at least as good as every alternative) on some admissible precisification are rational choices (Fine 1975; Broome 2009). If p is vague and therefore fails to fully determine a conventional ranking of x, y, and $x+$, one can precisify (or specify) p in different ways so that a conventional ranking results. For instance, one might precisify p so that it can be determined that x is less p than y and y less p than $x+$. However, if p, which is vague, manages to fully determine that x is less p than $x+$, it is not admissible to precisify p so that x is at least as p as $x+$. That is an inadmissible precisification. There is no admissible precisification on which x is optimal with respect to p. Proponents of Wlodek Rabinowicz's version of the fitting-attitudes framework for value, which explains value relations in terms of permissible preference orderings, can say that options that are dispreferred to some alternative on *all* permissible preference orderings are irrational (Rabinowicz 2008). Within this fitting-attitudes framework for value, it is permissible to prefer y to x and to prefer $x+$ to y, but if it can be determined that x is worse than $x+$, then x is dispreferred to $x+$ on all permissible preference orderings.

To stick to the concept of nondeterminacy, I will use a more general criterion that reflects the standard ways of revising rational choice criteria within the different explanatory frameworks:

> *Determinate maximality (DM)*: An option, x, is *determinately maximal* with respect to p if and only if (a) x is not worse than any alternative with respect to p, and (b) it is not the case that on all admissible precisifications of p, there is an alternative that is better than x.[2]

Regardless of one's view of how to explain nondeterminacy (e.g., vagueness, incomparability, parity, imprecise equality, or other non-conventional comparative relations), DM can be used to make rational choice assessments of options in choice sets in which some alternatives are nondeterminate in their ranking.

2 More on Determinate Maximality

One might now think that nondeterminacy problems do not pose a challenge to practical reasoning at all or that the challenge is a minor one (see, e.g., Hsieh 2007). Indeed, conventional rational choice theory cannot be used to guide action in choice situations characterized by nondeterminacy because there is no *optimal* alternative, but in all choice situations characterized by nondeterminacy, there is some determinately maximal alternative. So, why not just revise the decision criterion conventionally used in rational choice theory and hold that it is rational to choose any determinately maximal alternative? This section presents three arguments that challenge the plausibility of this view: it can lead to the formation of sequences of choices that are suboptimal; it neglects the fact that there might be resolutional remainder; it fails to recognize that the justificatory power grounded in comparative relations differs between situations in which some alternative is optimal and in which some alternative is determinately maximal but not optimal.

First, accepting nondeterminacy and revising one's decision criteria so that an option that is determinately maximal is rational leads to dynamic choice problems (see, e.g., McClennen 1990; Andreou 2016; Gustafsson 2016). If one adopts the view that all determinately maximal options are rational, one accepts a view that it can be permissible to form sequences of choices that are not determinately maximal (and even suboptimal), and a decision-maker who follows the strategy is exposed to a weak kind of money pump.

Consider an illustration. There are again three options: x, $x+$, and y: x is nondeterminate in its ranking with y; $x+$ is an improved version of x, so x is worse than $x+$, but $x+$ is also nondeterminate in its ranking with y. In other words, we again have

$$x \approx y; \ x+ \approx y; \ x < x +$$

Now, it is obvious that in a choice situation in which an agent chooses between x, $x+$, and y, $x+$ and y are the only determinately maximal alternatives. According to the view that it is rational to choose any determinately maximal alternative, one of them ought to be chosen. This seems sound.

However, the approach seems less sound when one considers choice sequences. Consider the possibility that an agent will make two

sequential choices, first (at t_1) a choice between $x+$ and y, and then (at t_2) a choice between sticking to the first choice or switching to x:

t1   t2   x+
                                      x+
                          x+
                                      x
                  x+
                                          x
                              x
                  y
                          y   y
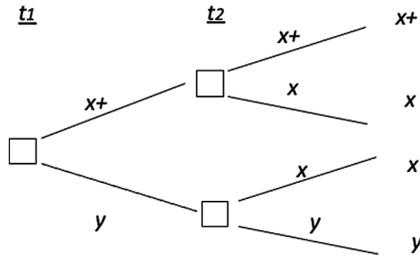
According to the view that it is rational to choose any option that is determinately maximal, it is rational to choose y at t_1 and x at t_2. However, this means that the agent ends up with a choice that is not determinately maximal (x is worse than $x+$, which could be chosen if the agent had chosen $x+$ at both t_1 and t_2), which, according to the approach that was followed, is an irrational choice. Holding all determinately maximal alternatives to be rational choices thus seems to lead to (or, more precisely, permit) irrational choices in some situations.

This phenomenon also reveals that agents who use DM as the only criterion when making choices between options that are nondeterminate in their ranking are exposed to a weak kind of money pump (Gustafsson 2016). To see this, imagine that the aforementioned sequence consisted of three sequential choices, that the agent at t_3 had the opportunity to either stick to their choice at t_2 or (if that was not their choice at this point) choose $x+$. An agent who chooses y followed by x would not only be rationally obliged to switch to $x+$ at t_3 but also rationally obliged to pay some money (amounting to the difference between x and $x+$) to change their choice. If the choice sequence could be iterated, the agent would be rationally permitted to financially ruin themselves for no reason or gain whatsoever.

Adopting the view that one can make rational choices when facing options that are nondeterminate in their ranking by choosing any determinately maximal alternative means that one accepts that it can be rational to form a choice sequence that leads to an irrational choice. This is not the place to explore how often this will be a problem or the conditions under which it poses a problem. But it is obvious that it is a possibility, and this possibility alone gives us reason to question whether DM is a necessary and sufficient criterion for rational choice when some options are nondeterminate in their ranking. In the literature on dynamic inconsistency and money pumps, it has been suggested that these problems can be avoided by adopting either the view that rational agents are *resolute* and do not change their minds (McClennen 1990) or the view that rational agents are sophisticated choosers who use backward induction to foresee problems when they make sequential choices (Rabinowicz 1995).

In what follows, I will suggest that dynamic choice problems should instead be solved by the introduction of some new decision criteria.

Second, it has been argued that contrary to situations in which two alternatives are optimal, when two alternatives are determinately maximal but not optimal, there is resolutional remainder:

> [In choice situations characterized by nondeterminacy], there is resolutional remainder; given a list of admissible ways in which the perplexity might be resolved, there is still a further question as to how the perplexity is to be resolved, for that resolution is not simply given by arbitrarily opting for one admissible resolution over another.
>
> (Chang 2002: 684–685)[3]

In some choices between options that are nondeterminate in their ranking, it seems as though something is unresolved if one merely picks an option that is determinately maximal. To take this intuition seriously, this something ought to be addressed by practical reasoning in some way.

Third, there is a difference between what might be called the justificatory power of an option being optimal and an option being determinately maximal but not optimal. When two options are (determinately) equally good, this provides a sufficient justification for an agent to choose either one of them. When neither of two options is (determinately) at least as good as the other, this might provide some justification for an agent to choose either one of them, but this justification seems weaker. In other words, it is not obvious that the justification of a choice that is grounded in how alternatives relate to each other is sufficiently strong to meet the requirements of rationality when an option is merely determinately maximal but not optimal, as opposed to when an option is optimal (i.e., at least as good as all alternatives).

3 Introducing New Criteria

A different way of responding to nondeterminacy problems is to say that when two options are nondeterminate in their ranking with respect to some decision criterion or reason, some other criterion or reason should be used to rank them so that one can establish a conventional ranking of all pairs of options. This approach clearly deals with potential resolutional remainder and, insofar as it can determine an option that is at least as good as all alternatives, it leads to a strong justification.

For instance, someone may propose the idea that when some set P of practical reasons that apply to the choice admits of nondeterminacy, one ought to make pairwise comparisons of the options that are nondeterminate in their ranking by ranking them with respect to some other reason, Q, that does not admit of nondeterminacy. In other words:

> *Imposing conventional pairwise rankings: x* is worse than *y* if and only if *x* is worse than *y* with respect to the set of primary practical

reasons P, or x and y are nondeterminate in their ranking with respect to P and x is worse than y with respect to secondary reason Q.

At a general level, and insofar as one does not qualify what Q might be and how it might relate to P, this view is problematic. To see this, consider the following possibility. Assume, again, that with respect to P, x, $x+$, and y relate to each other in the following way:

$$x \approx y; \; x+ \approx y; \; x < x+$$

If one does not qualify Q in any way, it is perfectly possible that with respect to Q, x, $x+$, and y relate to each other in the following way:

$$x+ < y; \; y < x; \; x+ < x$$

Imposing conventional pairwise rankings then can generate the following cyclical ranking of the options:

$$x+ < y; \; y < x; \; x < x+$$

Thus, $x+$ is worse than y because $x+$ and y are nondeterminate in their ranking with respect to P, and $x+$ is worse than y with respect to Q; y is worse than x because y and x are nondeterminate in their ranking with respect to P, and y is worse than x with respect to Q, and x is worse than $x+$ because x is worse than $x+$ with respect to P. According to this approach it is then irrelevant how x and $x+$ relate to each other with respect to Q.

This is not just a theoretical possibility. To see this, consider an illustration. Imagine that x, $x+$, and y are different wines and that P reflects the set of reasons that underlie how much one will enjoy them. Let Q reflect the reason to favor more environmentally friendly wines, something an agent could plausibly use as a criterion to establish pairwise rankings of alternative wines that are nondeterminate in their ranking with respect to how much she will enjoy them. There is no obvious and necessary relation between how much one will enjoy a wine and how environmentally friendly the production of the wine is. Therefore, it is perfectly possible that insofar as how much one will enjoy a wine admits of nondeterminacy, imposing conventional pairwise rankings of wines by looking at how environmentally friendly the production is can lead to cyclical evaluations. Assume that x is a white wine (say a Pinot Grigio) that is nondeterminate in its ranking with red wine y (say a Bordeaux) with respect to how enjoyable it is to consume and that $x+$ is a slightly better white wine (say a Chablis) that is also nondeterminate in its ranking with red wine y. As it happens, the production process of x is the most environmentally friendly, followed by the production process of y, with the production process of $x+$ being the least environmentally friendly. An agent who

makes pairwise rankings of the wines by determining that a wine is worse than another if and only if it is less enjoyable, or if it is nondeterminate in its ranking with the other wine but its production process is less environmentally friendly, will generate a cyclical ranking of the wines.

It is sometimes suggested that one can avoid accepting unpalatable rankings such as these by individuating alternatives in different ways depending on the comparative context (see Broome 1991; Arrhenius 2009; Voorhoeve 2014; Herlitz 2019b, 2020a). If one, for instance, individuated y in different ways depending on whether one compares y with x (say y^*) or with $x+$ (say y^{**}), the ranking would no longer be cyclical. The following is not a cyclical ranking of the options:

$$x+ < y^{**};\ y^* < x;\ x < x+$$

Although in principle a theoretically possible way to refute the argument, this solution is not appealing. As John Broome (1991) has pointed out, individuating alternatives in different ways is always possible, but because of this, one must have a good reason to change the individuation. In general, alternatives ought to be individuated in such a way that all features that are pertinent in light of what matters to the comparison are included in the individuation (Broome 1991; Voorhoeve 2014; Herlitz 2019b, 2020a). If I wish to evaluate two wines in light of how enjoyable they are, I ought to individuate them so that their enjoyable properties are included. In the previous example, it is obvious that there is a way of individuating the alternatives so that both the features pertinent in light of P and the features pertinent in light of Q are included and generate a cyclical ranking. For instance, one can individuate the three wines in a way that includes both their enjoyable features and facts about how environmentally friendly their production processes are and end up with a cyclical ranking of the wines. This suffices for the argument that imposing conventional pairwise rankings can lead to cyclical rankings.

What started as a nondeterminacy problem – a problem relating to the fact that reasons fail to determine an option that is at least as good as all alternatives – has now turned into a cyclical ranking of the options, with all options being *worse* than some alternative (see Herlitz 2020b).[4] Although acyclicity is not necessary for rational choice (see, e.g., Schwartz 1986; Duggan 2013; Herlitz 2020c), cyclical orderings are generally seen as a bigger problem than nondeterminacy (see, e.g., Broome 2004).

Furthermore, it is questionable whether the methods that can be used to identify rational options when one deals with cyclical rankings generate plausible results in this case. Conventional ways of dealing with cycles (e.g., referring to Schwartz sets or uncovered options) would say that all options are equally rational choices in the previous example. This seems questionable. If x, $x+$, and y are three wines, one is primarily interested in how much one will enjoy the wines, and x is determinately less enjoyable than $x+$, it is questionable whether it is rational to choose x (see Herlitz 2020b).

4 Two-step Models

Rather than introducing a new reason in order to determine how options relate to each other in pairwise comparisons, one might be attracted by the idea of introducing a new reason in order to rank the options that remain after DM has been applied. After all, it seems implausible to consider x a viable option at all if it is determinately true that x is worse than $x+$ with respect to what matters to the choice. Also, this approach clearly deals with potential resolutional remainder, and since it determines an option that is at least as good as all alternatives, it leads to as strong a justification as the one provided in cases in which there is no nondeterminacy.

In other words, one might propose the following idea:

> *Two-step model*: Choose an option that is determinately maximal with respect to P. If more than one option is determinately maximal with respect to P, choose an option in this set that is at least as good as all alternatives in the set with respect to Q.

This has been a somewhat popular move among those who recognize nondeterminacy problems and take them seriously. Chang (2013) has famously argued that when what she calls "given reasons" fail to fully determine what one ought to do, one ought to introduce a created reason – grounded in the agent's volition – and do what is best with respect to the created reason. Norman Daniels (2008) has argued that substantive principles used to rank different healthcare policies admit of nondeterminacy and fail to fully determine what policymakers ought to do; in those situations, he argues that one ought to resort to a deliberative process and rank alternatives with respect to the criteria that transpire from such a process. Fröding and Peterson (2012) have suggested that a virtuous agent might face options that are nondeterminate in their ranking with respect to moral reasons and that the virtuous agent will compare these options in terms of how good they are with respect to relevant non-moral reasons. I myself have advocated such two-step models in the context of allocating scarce health-related resources, in social choice situations, and in the context of clinical choice situations, person-centered care, and shared decision-making (Herlitz 2017, 2018, forthcoming; Herlitz & Sadek 2021).

However, this method violates what many (see, e.g., Chernoff 1954; Sen 1970; Fleurbaey, Tungodden and Vallentyne 2009) take to be a basic requirement of rationality, i.e., the so-called basic contraction consistency, sometimes called property *alpha* (see Herlitz 2019b):

> *Basic contraction consistency*: If an alternative, x, is permissible in a choice set C, then it is also permissible in a choice set C' that is a subset of C and contains x.

It can be easily seen that basic contraction consistency is violated by the two-step model. Consider once again x, $x+$, and y, which relate to each other with respect to P in the following way:

$$x \approx y; \; x+ \approx y; \; x < x+$$

and which relate to each other with respect to Q in the following way:

$$x+ < y; \; y < x; \; x+ < x$$

According to the two-step model, y should be chosen (and is the only permissible option) if the choice set is $\{x, x+, y\}$. This is because only $x+$ and y are determinately maximal with respect to P, and $x+$ is worse than y with respect to Q. Since x is worse than $x+$ with respect to P, it is not determinately maximal and is therefore discarded in the first step; it is irrelevant how x and $x+$ relate to each other with respect to Q.

However, x should be chosen (and is the only permissible option) according to the two-step model if the choice set is $\{x, y\}$. Both x and y are determinately maximal with respect to P when $x+$ is not in the choice set, and y is worse than x with respect to Q. Since $x+$ is not in the choice set, there is no option that is better than x with respect to P in the choice set; there are no grounds for ruling out x in the first step, so how good x is with respect to Q becomes highly relevant.

Obviously, $\{x, y\}$ is a subset of $\{x, x+, y\}$, and whereas y is permissible in $\{x, x+, y\}$, it is not permissible in $\{x, y\}$, which contains y. In other words, basic contraction consistency is violated.

The two-step model would violate basic contraction consistency if x, $x+$, and y are the three wines; P reflects how enjoyable they are; and Q reflects how environmentally friendly the production processes of the wines are, as described in the previous section. When choosing between all three wines, the approach suggests that one first discards the worse white wine, x, and then ranks the better white wine, $x+$, and the red wine, y, with respect to how environmentally friendly their production processes are. Since the production process of the red wine is more environmentally friendly than that of the better white wine, the approach suggests that one will choose the red wine. When choosing only between the less enjoyable white wine and the red wine, all options are nondeterminate in their ranking with respect to how enjoyable they are, so all options are evaluated in terms of how environmentally friendly their production processes are. Since the production process of the less enjoyable white wine is more environmentally friendly than that of the red wine, the approach suggests that one ought to choose the more environmentally friendly red wine.

5 Avoiding Violations of Rationality Requirements

The previous sections reveal that nondeterminacy in principle can cause agents who adopt seemingly attractive and plausible decision strategies

to violate what appear to be basic requirements of rationality. An agent who adopts the strategy "choose any determinately maximal alternative" might form choice sequences that lead to choices that are not determinately maximal. An agent who attempts to impose conventional comparative relations on pairs of options that are nondeterminate in their ranking with respect to the reasons that apply to the situation by invoking some new reason might end up generating cyclical evaluations of her options. An agent who discards alternatives that are not determinately maximal and compares the remaining options that are determinately maximal with respect to some new reason might violate basic contraction consistency. What can we learn from this?

First, it can be noted that both the argument that illustrated how nondeterminacy can spawn cyclical evaluations and the argument that illustrated how two-step models violate basic contraction consistency rely on an important assumption. For these arguments to work, it must be assumed that the reasons that admit of nondeterminacy and the reason that is brought in to establish a ranking do not relate to each other in specific ways.

If one, by contrast, puts constraints on what might qualify as a valid reason that plays this role, the arguments can be refuted (see Herlitz 2019b). As an illustration, consider again the case in which x, $x+$, and y are three wines, and there is nondeterminacy in their ranking when one ranks them with respect to how enjoyable, P, they are ($x \approx y$ and $x+ \approx y$). Assume further that the two white wines, x and $x+$, are enjoyable because they are light and refreshing, and the red wine, y, is enjoyable because it has a rich taste. Enjoyable, P, admits of nondeterminacy because both wines that are light and refreshing and wines with a rich taste are enjoyable, but it is not fully determined how these features relate to each other for overall assessments of how enjoyable different wines are. Insofar as the reason one uses to establish conventional rankings or to choose between options in a second step, Q, is entirely unrelated to how enjoyable the wines are, for instance, if this reason relates to how environmentally friendly the production processes are, imposing conventional comparative relations and the two-step model run into problems (as illustrated in the previous sections). By contrast, if one holds that the reason used to establish a conventional ranking must be a reason, Q^*, that determines how being light and refreshing and being rich in taste relate to each other with respect to how the features contribute to overall assessments of how enjoyable different wines are, one can avoid these problems. For instance, Q^* could be "light and refreshing wines are better than wines with a rich taste," or "wines with a rich taste are better than light and refreshing wines." Such a reason, Q^*, will generate rankings of the less enjoyable white wine and the red wine as well as of the more enjoyable white wine and the red wine such that if it determines that the less enjoyable white wine is better than the red wine, then also the more enjoyable white wine will be better than the red wine, and if the more enjoyable white wine is worse than the red wine, then also the less enjoyable white wine will be worse than the red

wine. It follows from the constraints on Q^* that if $y < x$ with respect to Q^*, then $y < x+$ with respect to Q^*, and if $x+ < y$ with respect to Q^*, then $x < y$ with respect to Q^*. This means that there will be no cycle in the ranking if one imposes conventional comparative rankings with reference to Q^* and relying on Q^* in a two-step model will not lead to violations of basic contraction consistency.

The point can be generalized within specific explanatory frameworks of nondeterminacy. Consider, for instance, how one can impose constraints on Q with reference to the supervaluationist approach to vagueness within the framework that holds that nondeterminacy is due to vagueness. Assume that x and y are nondeterminate in their ranking with respect to P and that $x+$ and y are nondeterminate in their ranking with respect to P because P is vague so that it is indeterminate whether x is more P than, less P than, or equally as P as y and indeterminate whether $x+$ is more P than, less P than, or equally as P as y. According to the supervaluationist approach to vagueness, a sentence that contains a vague term is "supertrue" if it is true on all admissible precisifications and "superfalse" if it is false on all admissible precisifications (Fine 1975; Broome 2009; Andersson 2017). An admissible precisification is one that is in accordance with ordinary language and that respects so-called penumbral truths (Fine 1975). If one holds that Q must be such that it provides an admissible precisification of P, imposing conventional comparative rankings will not generate cyclical rankings, and the two-step model will not violate basic contraction consistency. This is because introducing and relying on a Q that would lead to cyclical rankings or violations of basic contraction consistency would mean that one precisifies P in a way that contradicts penumbral truths, which is inadmissible.

Insofar as one commits to using *the same Q* that provides an admissible precisification of P to determine a ranking of the alternatives in *all* situations in which some alternatives are indeterminate in their ranking with respect to P (which is vague), one also avoids the risk of forming sequences of choices that lead to options that are determinately worse (with respect to P) than some available choice sequence that is available. If an agent commits to determine what to do when the options are indeterminate in their ranking with respect to P by ranking the options with respect to Q and if Q provides an admissible precisification of P, then if $x+ < y$ at t_1 then $x < y$ at t_2. An agent who commits to using the same Q that provides an admissible precisification of P to determine a ranking of the alternatives will, in other words, either choose $x+$ at t_1 and $x+$ at t_2 or y at t_1 and y at t_2, neither of which leads to a choice that is determinately worse with respect to P than any available alternative.

It is obvious from the aforementioned that one can avoid the problems outlined in the previous sections of this chapter if one accepts that nondeterminacy is due to vagueness and imposes constraints on what qualifies as an admissible Q in terms of admissible precisifications. Somewhat similarly, proponents of the fitting-attitudes approach to value can within

their framework impose constraints on what is an admissible Q in terms of permissible preference orderings. For instance:

> *Yields permissible preference orderings over the choice set: Q* is admissible if and only if it yields preference orderings over a choice set that are permissible given the rankings established by *P*.
>
>   (see Rabinowicz 2008; Herlitz 2019b)

In other words, it is noteworthy that, *given certain explanations of non-determinacy*, there are views of how to make rational choices when reasons fail to fully determine a best alternative by introducing new reasons to establish conventional rankings that avoid the problems outlined in the previous section. These views say that one must impose constraints on what the new reasons that establish conventional rankings might be like, somewhat similar to how supervaluationist approaches to vagueness impose constraints on what constitutes admissible precisifications of vague terms.

6 Understanding How Reasons Cannot Admit of Violations of Rationality Requirements

Given that one can avoid choice-theoretical challenges by putting constraints on reasons that are introduced to create conventional rankings of options that are nondeterminate in their ranking, one can perhaps learn something about when and how nondeterminacy occurs by studying how decision strategies might violate rationality requirements. It is possible that reasons cannot admit of nondeterminacy in such a way that doing what is permissible according to these reasons can mean that one violates basic requirements of rationality. This section explores this possibility.

Consider four views of the problems introduced in the previous sections. First, it is of course possible to refute the relevance of the arguments:

> *The accept irrationality view*: Reasons can admit of nondeterminacy, and this means that acceptable decision strategies sometimes lead to violations of basic requirements of rationality. However, this is not a problem.

Although this is certainly a possible position, I believe most people would favor views that do not entail cyclical rankings or the violation of basic contraction consistency.

Here are three views that achieve this:

> *The properties of nondeterminacy view*: Reasons can admit of non-determinacy, but the phenomenon – nondeterminacy – has certain

properties that constrain what it means to respond appropriately to these reasons.

On this view, nondeterminacy has certain properties with substantive implications for how it is appropriate to respond to the reasons that admit of nondeterminacy when some options are nondeterminate in their ranking. The relation "*x* is at least as good as *y*" has a property that implies that it is appropriate to choose *x*. By contrast, the relation "*x* is nondeterminate in its ranking with *y*" might have a property that implies that it is appropriate to use a certain method to choose between *x* and *y*, and this method has formal constraints in line with the constraints discussed in the previous section.

On the properties of the nondeterminacy view, the features of nondeterminacy explain that one ought to respond to nondeterminacy in a way that avoids violating basic requirements of rationality. The nondeterminacy phenomenon as such constrains rational agents that face it.

> *The not all reasons view*: Reasons can admit of nondeterminacy, but the domain of valid reasons cannot be such that responding to the reasons means that one violates basic requirements of rationality.

On this view, the domain of valid reasons has certain properties such that if the valid first-order reasons admit of nondeterminacy, the only second-order reasons that might be used to establish conventional rankings that are valid are reasons that do not lead to violations of basic requirements of rationality. On this view, basing one's choice of wine – in the example used above – on how environmentally friendly the production process is amounts to basing one's choice of wine on a second-order reason that is invalid.

On the not all reasons view, the domain of valid reasons is such that the only second-order reasons that are valid are those that do not lead to violations of basic requirements of rationality, which explains how one ought to respond to nondeterminacy in a way that avoids violations of basic requirements of rationality. The domain of valid reasons ensures that rational agents that face nondeterminacy can never act on valid reasons and violate rationality requirements because the domain of valid reasons has an architecture that renders this impossible.

> *The rationality obliges view*: Reasons can admit of nondeterminacy, but for an agent to be rational, she must adopt a decision strategy that does not violate basic requirements of rationality.

On this view, rationality has certain properties that constrain how one can rationally respond to nondeterminacy. Accordingly, choosing a wine in the recurrent example with reference to the environmental friendliness of the production process is irrational because norms of rationality

require that rational agents avoid decision strategies that violate basic requirements of rationality. On the rationality obliges view, that this is what it means to be rational explains how one ought to respond to nondeterminacy in a way that avoids violating basic requirements of rationality. Rationality thus constrains agents who face nondeterminacy.

The properties of the nondeterminacy view, the not all reasons view, and the rationality obliges view can all explain why a rational agent who acts on valid reasons that admit of nondeterminacy will avoid violating basic requirements of rationality. As such, each of these views offers a solution to the choice-theoretical challenges that arise once one accepts the possibility of nondeterminacy. However, the views explain how rational agents avoid these problems in different ways, and these different ways have very different implications for different areas of interest to decision theorists and philosophers. Those who accept the properties of the nondeterminacy view will have to accept that comparative relations carry far more information than is often thought and that they also can have wide-ranging implications for how rational agents are permitted to make choices. Those who accept the not all reasons view will have to accept that there are formal constraints on the domain of valid reasons that are indirectly inferred from constellations in the domain of reason in complex ways. Finally, those who accept the rationality obliges view will have to accept a view of rationality that is rich enough to have implications that impose constraints on second-order reasons.

7 Conclusion

In this chapter, I have outlined some choice-theoretical challenges that arise if one accepts the possibility of nondeterminacy, i.e., that action-guiding reasons may fail to fully determine a best option. I demonstrated that nondeterminacy can lead agents to form suboptimal choice sequences and that if they use seemingly plausible decision strategies, they might make cyclical rankings of their options or violate basic contraction consistency. I also demonstrated that there are ways to impose constraints on decision strategies in order to avoid these problems but showed that these constraints can be derived either from the phenomenon of nondeterminacy, from the reasons themselves, or from norms of rationality. Whichever way one attempts to derive the constraints, there will be implications for the phenomenon of nondeterminacy, for what the domain of reasons is like, or for norms of rationality.

Notes

2 Within the fitting-attitudes framework for value, saying that x is not worse than any alternative with respect to p amounts to saying that there is no alternative, z, such that it is required to disprefer x to z.

3 When Ruth Chang first introduced this idea, she tied it to parity, her favored explanation of nondeterminacy. However, there is no reason to tie the phenomenon of resolutional remainder to a particular explanation of nondeterminacy (Wasserman 2004).

4 Somewhat relatedly, insofar as nondeterminacy is possible, it is in principle possible that it occurs when one applies one of several decision criteria.

References

Andersson, H. (2017), *How It All Relates: Exploring the Space of Value Comparisons* (Department of Philosophy, Lund University).

Andreou, C. (2016), 'Dynamic Choice', in E. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Center for the Study of Language and Information Stanford University).

Arrhenius, G. (2000), 'An Impossibility Theorem for Welfare Axiologies', *Economics and Philosophy* 16/2: 247–266.

Arrhenius, G. (2009), 'Can the Person Affecting Restriction Solve the Problems in Population Ethics?', in M. A. Roberts and D. T. Wasserman (eds.), *Harming Future Persons: Ethics, Genetics and the Nonidentity Problem* (Springer), 289–314.

Arrhenius, G. (forthcoming). *Population Ethics: The Challenge of Future Generations*.

Broome, J. (1991), *Weighing Goods* (Blackwell).

Broome, J. (1997), 'Is Incommensurability Vagueness?', in R. Chang (ed.), *Incommensurability, Incomparability, and Practical Reason* (Harvard University Press), 67–89.

Broome, J. (2004), *Weighing Lives* (Oxford University Press).

Broome, J. (2009), 'Reply to Rabinowicz', *Philosophical Issues* 19/1: 412–417.

Broome, J. (2021), 'Incommensurability Is Vagueness', in H. Andersson and A. Herlitz (eds.), *Value Incommensurability: Ethics, Risk and Decision-Making* (Routledge).

Carlson, E. (2010), 'Parity Demystified', *Theoria* 76/2: 119–128.

Chang, R. (2002), 'The Possibility of Parity', *Ethics* 112/2: 659–688.

Chang, R. (2013), 'Grounding Practical Normativity: Going Hybrid', *Philosophical Studies* 164: 163–187.

Chernoff, H. (1954), 'Rational Selection of Decision Functions', *Econometrica* 22/4: 422–443.

Daniels, N. (2008), *Just Health: Meeting Health Needs Fairly* (Cambridge University Press).

Duggan, J. (2013), 'Uncovered Sets', *Social Choice and Welfare* 41: 489–535.

Elson, L. (2017), 'Incommensurability as Vagueness: A Burden-Shifting Argument', *Theoria* 83/4: 341–363.

Fine, K. (1975), 'Vagueness, Truth and Logic', *Synthese* 30: 265–300.

Fleurbaey, M., Tungodden, B., and Vallentyne, P. (2009), 'On the Possibility of Nonaggregative Priority for the Worst Off', *Social Philosophy and Policy* 26/1: 258–285.

Fröding, B., and Peterson, M. (2012), 'Virtuous Choice and Parity', *Ethical Theory and Moral Practice* 15: 71–82.

Gustafsson, J. (2016), 'Money Pumps, Incompleteness, and Indeterminacy', *Philosophy and Phenomenological Research* 92/1: 60–72.

Herlitz, A. (2017), 'Comparativism and the Grounds for Person-Centered Care and Shared Decision Making', *Journal of Clinical Ethics* 28/4: 269–278.

Herlitz, A. (2018), 'Committing to Priorities: Incompleteness in Macro-Level Health Care Allocation and its Implications', *Journal of Medicine and Philosophy* 43/6: 724–745.

Herlitz, A. (2019a), 'Nondeterminacy and Population Ethics', in P. Bowman and K. B. Rasmussen (eds.), *Studies on Climate Ethics and Future Generations*, Working paper series 2019: 1–11 (Institute for Futures Studies).

Herlitz, A. (2019b), 'Nondeterminacy, Two-Step Models, and Justified Choice', *Ethics* 129/2: 284–308.

Herlitz, A. (2020a), 'Stable and Unstable Choices', *Economics and Philosophy* 36/1: 113–125.

Herlitz, A. (2020b), 'Nondeterminacy, Cycles and Rational Choice', *Analysis* 80/3: 443–449.

Herlitz, A. (2020c), 'Non-transitive Better than Relations and Rational Choice', *Philosophia* 48: 179–189.

Herlitz, A. (Forthcoming), 'Toward a Hybrid Theory of How to Allocate Health-Related Resources', *Journal of Medicine and Philosophy*.

Herlitz, A., and Sadek, K. (2021), 'Social Choice, Nondeterminacy, and Practical Reasoning', *Res Philosophica* 98/3: 1–25.

Hsieh, N.-H. (2007), 'Is Incomparability a Problem for Anyone?', *Economics and Philosophy* 23/1: 65–80.

McClennen, E. F. (1990), *Rationality and Dynamic Choice: Foundational Explorations* (Cambridge University Press).

Norcross, A. (1997), 'Comparing Harms: Headaches and Human Lives', *Philosophy and Public Affairs* 26/2: 135–167.

Parfit, D. (1984), *Reasons and Persons* (Clarendon Press).

Parfit, D. (2016), 'Can We Avoid the Repugnant Conclusion?', *Theoria* 82/2: 110–127.

Qizilbash, M. (2007), 'The Mere Addition Paradox, Parity and Vagueness', *Philosophy and Phenomenological Research* 75/1: 129–151.

Rabinowicz, W. (1995), 'To Have One's Cake and Eat It, Too: Sequential Choice and Expected-Utility Violations', *Journal of Philosophy* 92/11: 586–620.

Rabinowicz, W. (2008), 'Value Relations', *Theoria* 74/1: 18–49.

Rabinowicz, W. (2019), 'Can Parfit's Appeal to Incommensurabilities Block the Continuum Argument for the Repugnant Conclusion?', in P. Bowman and K. B. Rasmussen (eds.), *Studies on Climate Ethics and Future Generations*, Working paper series 2019: 1–11 (Institute for Futures Studies).

Rachels, S. (1998), 'Counterexamples to the Transitivity of Better Than', *Australasian Journal of Philosophy* 76/1: 71–83.

Raz, J. (1986), *The Morality of Freedom* (Oxford University Press).

Schwartz, T. (1986), *The Logic of Collective Choice* (Columbia University Press).

Sen, A. (1970), *Collective Choice and Social Welfare* (Holden-Day).

Sen, A. (1997), 'Maximization and the Act of Choice', *Econometrica* 65/4: 745–779.

Sen, A. (2017), 'Reason and Justice: The Optimal and the Maximal', *Philosophy* 92/1: 5–19.

Sinnott-Armstrong, W. (1988), *Moral Dilemmas* (Blackwell).

de Sousa, R. (1974), 'The Good and the True', *Mind* 83/332: 534–551.

Temkin, L. (1996), 'A Continuum Argument for Intransitivity', *Philosophy and Public Affairs* 25/3: 175–210.

Temkin, L. (2012), *Rethinking the Good: Moral Ideals and the Nature of Practical Reasoning* (Oxford University Press).

Voorhoeve, A. (2014), 'How Should We Aggregate Competing Claims?', *Ethics* 125/1: 64–87.

Wasserman, R. (2004), 'Indeterminacy, Ignorance and the Possibility of Parity', *Philosophical Perspectives* 18/1: 391–403.

8 Cross-Categorical Value Comparisons

Krister Bykvist

1 Introduction

Without blinking an eye, we make comparisons across distinct value categories. Here are some examples.[1]

(1) This table is a better table than Trump is a president.[2]
(2) You are more prudent than you are moral.
(3) I am more kind than I am beautiful.

This is puzzling since "better table," "more prudent," and "more kind" cannot be read in the usual ways here: "being better as a table," "showing a higher degree of prudence," and "showing a higher degree of kindness," respectively. This has prompted some linguists to think that these constructions are anomalous in some sense, ungrammatical or nonsensical. Others have claimed they are grammatical and meaningful but only metaphorical or metalinguistic (e.g., it is more appropriate to say "this is a good table" than to say "Trump is a good president"). Some linguists, such as Alan Bale (2006, 2008) and Ewan Klein (1980), however, have offered what they think is a meaningful non-metaphorical interpretation. Except for a few brief comments, philosophers have so far been silent on this perplexing issue, which is especially odd since philosophers thrive on puzzles.[3]

In this chapter, I shall argue against the idea that the meaningfulness of cross-categorical comparisons (sometimes called "indirect" or "interadjectival" comparisons) is an illusion. I shall also argue against the existent linguistic ideas about how to make these comparisons meaningful. My focus throughout will be on cross-categorical *value* comparisons. I shall give the rough contours of a new theory of such comparisons (with a hint on how to extend them to their non-evaluative counterparts) and show that they are normatively relevant. In particular, I shall suggest how they can be usefully invoked when comparing well-being across people, when assessing the fittingness of attitudes, and when making cross-theory value comparisons in the context of acting under evaluative uncertainty

("What should I do when I am not certain about the values of my alternatives?"). If I am right, then there is a need to posit a new nonstandard value relation or, more exactly, a family of cross-categorical betterness and equality relations.

Before we start, I need to make an important clarification. I am interested in philosophical analyses of the concept of cross-categorical comparisons: a priori, necessarily true biconditionals in which the left-hand claim, if true, holds in virtue of the right-hand claim. This is not the typical focus in linguistic analysis, which only aims to provide the truth conditions of a natural language sentence. To draw on linguistic proposals, I will therefore read them as philosophical analyses. This also means that I will gloss over issues concerning the surface grammar of English, which often plays an important role in linguistics but less so in philosophical analysis.

2 Cross-Categorical Comparisons in General

What are cross-categorical comparisons (let us call them *cross-cats* for short, which also will remind us of their cross-breed nature)?

First, we can say that they have any of the following forms, where F and G are different gradable adjectives and A and B are nouns:

x is F-er/more F than *y* is G.
x is F-er/more F for an A than *y* is G for a B.
x is an F-er A than *y* is a G B.

However, not all constructions of these forms are cross-categorical in the sense that I am interested in. Take, for instance, this cross-adjectival sentence:

(4) This chair is taller than it is wide.

This is not cross-categorical in the relevant sense, since the tallness and the width of the chair can be directly compared in terms of *length*, in cm, for example.

What I am after are constructions in which F and G are distinct adjectives that do not pick out directly comparable properties. Here are some non-evaluative examples:

(5) Einstein is more intelligent than I am beautiful.
(6) I am more angry than she is happy.
(7) He is taller for a Swede than he is heavy for a Swede.

For none of the pairs (intelligent, beautiful), (angry, happy), and (tall, heavy) are the elements directly compared on one common dimension,

and the same seems to hold for the evaluative examples listed at the beginning of the chapter: (good as a table, good as a president), (prudent, moral), and (kind, beautiful).

One could perhaps object here that appearances are misleading when it comes to the evaluative examples. It may seem as though the evaluative properties cannot be directly compared, but if we look closer, we can see that they can be so compared. For example, why not say that

> (2)   you are more prudent than you are moral,

should roughly be understood as

> (8)   the degree to which I am prudent counts more toward my flourishing or overall virtuousness than the degree to which I am moral?

The idea is that being kind and being moral fall under the same *covering value*, counting toward my flourishing or overall virtuousness.[4] This does not seem to be plausible, however. It is perfectly fine to say that

> (9)   you are more prudent than you are moral, but I have no idea about what makes your life flourish more.

Furthermore, there is nothing strange about saying that

> (9*)   it is unfortunate that you are more prudent than you are moral, since the degree to which you are moral counts more toward your flourishing than does the degree to which you are prudent.

Returning to the matter of delineating cross-categorical comparisons, we can see that, as Bale notes, one crucial difference between cross-adjectival statements that are and those that are not reducible to direct comparisons concerns their conceptual *entailments*.

> (10a)   I am taller than I am wide.
> (10b)   If I am very wide, then I am at least very tall.
>
> (11a)   Einstein is more intelligent than you are beautiful.
> (11b)   If you are very beautiful, then Einstein is at least very intelligent.
>
> (12b)   He is taller for a Swede than he is heavy for a Swede.
> (12c)   If he is very heavy for a Swede, he is at least very tall for a Swede.

Statement 10a does not entail 10b, for 10a can be true even if I am very wide for a man – say, 1 m 30 cm – and short for a man – say, 1 m 32 cm. All it takes for 10a to be true is that my height exceeds my width. In contrast, 11a entails 11b, and 12a entails 12b. That is why these informal inferences all seem fine.

Einstein is more intelligent than you are beautiful.
You are very beautiful.
So
Einstein is at least very intelligent.

He is taller for a Swede than he is heavy for a Swede.
He is very heavy for a Swede.
So
He is at least very tall for a Swede.

The same kind of entailment holds for our evaluative examples:

(13a) This table is a better table than Trump is a president.
(13b) If Trump is a very good president, then this table is at least a very good table. (Given Trump's actual behavior, it is more appropriate to use a counterfactual conditional here!)

(14a) I am more selfish than you are moral.
(14b) If you are very moral, then I am at least very selfish.

(15a) You are more kind than you are brave.
(15b) If you are very brave, then you are at least very kind.

3 Skepticism about Cross-Categorical Comparisons

There is skepticism about whether we should take cross-cats literally. They have been accused of being ungrammatical, meaningless, metaphorical, and metalinguistic. So before we go on to discuss literal interpretations of cross-cats, we need to respond to these skeptical challenges. I think we can put the accusations of ungrammaticality and meaninglessness to one side since it seems obvious that we can use cross-cats grammatically and meaningfully. Furthermore, as we have seen, they also have clear and distinctive entailments.

According to the *metaphorical* interpretation, cross-cats only have metaphorical meaning. It is true that we sometimes use cross-cats to convey hyperbole:

(16) I am older than the universe is big.
(17) He is taller than the Great Pyramid of Khufu is heavy.

But it is one thing to say cross-cats can be used metaphorically and another to say that they can *only* be used in this way. For example, we do not seem to be forced to use the previously listed sentences metaphorically.

(1) This table is a better table than Trump is a president.
(2) You are more prudent than you are moral.
(3) I am more kind than I am beautiful.

They *seem* to have a literal meaning as well. I hope that the account of cross-cats I will present will validate this point – at least for evaluative cross-cats.

According to the *metalinguistic* idea, a cross-categorical sentence of the form

   x is F-er/more F than *y* is G

should be understood as something like

   it is more appropriate to say "*x* is F" than to say "*y* is G."

(The other constructions can be given analogous readings.) This account does not say much until one says *why* it is more appropriate to do this. One possibility is that it is more appropriate to say it because

   x is a clearer example of an F than *y* is an example of a G.

This approach does not work for all cross-cats, however, since it can be true that *x* is F-er than *y* is G, even though *x* is clearly not an F and *y* is clearly not a G

   (18)   I am taller than you are heavy

can be true even though I am a very short person, and thus clearly not a tall person, and you are a very light person, and thus clearly not a heavy person. Perhaps you are very light, and I am short but not very short.

Of course, much more can be said about this and other alternative metalinguistic approaches, but I hope that the account of cross-cats I will present shows that there is no need to "go metalinguistic" (at least not for evaluative cross-cats).

4 Positional Accounts

According to what I shall call *positional accounts*,

   x is F-er than *y* is G iff *x* has a higher position in the F-ranking than *y* has in the G-ranking.

   x is equally as F as *y* is G iff *x* has the same position in the F-ranking as *y* has in the G-ranking.

(The other constructions are given analogous readings.) Applied to our evaluative examples, this means that

   This table is a better table than Trump is a president iff this table has a higher position in the *good as a table* ranking than Trump has in the *good as a president* ranking.

You are more prudent than you are moral iff you have a higher position in the prudence ranking than you have in the morality ranking.

I am more kind than I am beautiful iff I have a higher position in the kindness ranking than I have in the beauty ranking.

This account has a ring of truth, but the crucial question is how to understand the talk about "higher position." Alan Bale (2006, 2008) spells it out as follows. Suppose we are considering n items ranked in terms of F-ness and also n items ranked in terms of G-ness. For each ranking, assign an item the ratio $n/n = 1$ if it is on top in the ranking, the ratio $n - 1/n$ if it is second, $n - 2/n$ if it is third, and so on. In Bale's *universal scale* account,

x is F-er than y is G iff the ratio assigned to x is greater than the ratio assigned to y.

x is equally as F as y is G iff the ratio assigned to x is the same as the ratio assigned to y.

As it stands, this account is severely restricted, however.

First, it only works if both the F-ranking and the G-ranking are *complete*. But we seem to be able to make cross-categorical comparisons when at least one of the rankings has *gaps*. For example, we can say that:

(19)   I am a better teacher than Talentlesso is an artist,

even if we assume that the *good as an artist*-ranking is incomplete. Perhaps Mozart cannot be said to be a better, worse, or equally as good artist as Michelangelo, to use a famous example.[5]

Second, it only works if the F-ranking and the G-ranking have the *same number* of compared items. To see this, suppose there are three Fs and five Gs. Then the second-best item in the F-ranking will be assigned two-third and the second-best in the G-ranking will be assigned two-fifths.

We do not want to say that the second-best item is F-er than another is G just because the F-ranked items are less numerous. For example, we do not want to say that I am kinder than you are brave just because, even though we are both second-best in our respective rankings, there are more brave people than kind ones.

Third, it only works if there is a *finite number* of compared items. But we seem to be able to make true cross-cats when at least one of the rankings has an infinite number of items, for example, when one of the compared properties is *unbounded*:

(20)   I am more moral than you (who are suffering immensely) are well off,

where being well off is an unbounded property (there is no limit to how well off you can be).

Perhaps these flaws can be fixed in a more sophisticated positional account. But there remains one problem that no positional account seems to be able to deal with. To see this, note first that any positional account must say that

if x is on top in the F-ranking and y is at the bottom in the G-ranking, then x is higher up in the F-ranking than y is in the G-ranking.[6]

Assume the top-ranked x is *very bad* in the relevant F-sense and the bottom-ranked y is *very good* in the relevant G-sense. Then it does not seem to be true to say that x is F-er than y is G despite the fact that x is higher up in the F-ranking than y is in the G-ranking. Suppose, for example, a, b, and c are compared in terms of both being good as a philosopher and being kind. They are all very bad as philosophers, but a is less bad than b, and b is less bad than c. They are all very kind but a is kinder than b and b is kinder than c. Then a has a higher position in the *good as a philosopher* ranking than c has in the *kindness* ranking since a is top ranked and c is bottom ranked. But we do not want to say that a is a better philosopher than c is kind.

We can make a similar point without focusing on cases in which the polarity is different (very bad versus very good). Suppose a, b, c are all kind but not very kind, and a is kinder than b, b is kinder than c. Suppose also that a, b, c are all very beautiful, and a is more beautiful than b, and b is more beautiful than c. Then a has a higher position in the kindness ranking than c has in the beauty ranking, but we do not want to say that a is kinder than c is beautiful.

In fact, this example shows that Bale's account is in danger of being *internally inconsistent*. Remember that he wants to validate this inference:

a is kinder than c is beautiful.
c is very beautiful.

So

a is at least very kind.

If Bale wants to insist that a is kinder than c is beautiful because a is higher up in the kindness ranking than c is in the beauty ranking, then he can no longer accept the entailment he claimed was a characteristic feature of cross-cats.

To these objections, Bale could reply that he had in mind cases in which we consider and rank *all* brave people and *all* kind people. If we consider the maximal comparison class of all kind people, the top-ranked ones must be very kind and the bottom ranked not very kind. Similarly, if we consider the comparison class of all beautiful people,

the top-ranked ones must be very beautiful and the bottom ranked not very beautiful. So it cannot be true that the top ranked in terms of kindness is not very kind and the bottom ranked in terms of beautiful is very beautiful.

But then we are back to the problem of unboundedness. This move does not work for cases in which there are no maximal or minimal elements, so the choice is between internal inconsistency or a severely restricted account. I hope we can do better and avoid this dilemma.

5 Further Constraints on Evaluative Cross-Cats

The previous examples show that we have further constraints on a viable account of evaluative cross-cats. Let us use "X-good," "X-bad," "X-neutral," "X-better" as short for "good in the relevant X-sense" (examples of instances would be "prudentially good" and "good as a table"), "bad in the relevant X-sense" (e.g., "prudentially bad" and "bad as a table"), "neutral in the relevant X-sense" (e.g., "finally neutral"), and "better in the relevant X-sense" (e.g., "prudentially better" and "better as a table"), respectively. We can now state the further constraints succinctly (where F and G are distinct senses of value):

> *Polarity*
> If x is F-*good* and y is G-*bad*, then x is F-better than y is G-valuable.
> If x is F-*neutral* and y is G-*bad*, then x is F-better than y is G-valuable.
> If x is F-*good* and y is G-*neutral*, then x is F-better than y is G-valuable.

I use "G-valuable" here rather than "G-good" since it is assumed that y is not G-good but G-bad or G-neutral.

> *"Very" modification*
> If x is very F-good and y is G-good but not very G-good, then x is F-better than y is G-good.

The last constraint can be generalized to other modifiers:

> *Degree modification*
> If x is D F-good and y is D' G-good and D is a higher degree modifier than D', then x is F-better than y is G-good,

where D is a degree modifier, such as *not very, very*, *extremely*, and *moderately*, and

> D is a higher degree modifier than D' = df. for all x, y, and F, if x is D F and y is D', then x is F-er than y.

Here are some illustrations:

(21)   This table is a better table than Trump is a president, since this table is a *good table* and Trump is a *bad president*. (*Polarity*)

(22)   You are more prudent than you are moral, since you are *very prudent* and you are moral but *not very moral*. ('Very' modification)

(23)   I am more kind than I am beautiful, since I am *extremely kind* and I am beautiful but *not extremely beautiful*. (*Degree modification*)

There is one important complication here that needs to be addressed. When I say that x is F-good but not very F-good and y is very F-bad, I take this to imply that x is F-better than y is F-valuable (by the first polarity principle listed above). But it also implies a *cross-polar* comparison: x is *less* F-*good* than y is F-*bad*, or, equivalently, y is *more* F-*bad* than x is F-*good*. This comparison holds, since y is *very* F-bad, while x is F-good but *not* very F-good. Suppose, for example, that x is a good but not very good table, and y is a very bad philosopher; then x is a better table than y is a philosopher. But it is also true that x is less good as a table than y is bad as a philosopher, or, equivalently, y is worse as a philosopher than x is good as a table, since x is good but not very good as a table, and y is very bad as a philosopher.

Note that cross-polar comparisons can be found in the *intra-category* cases too. Suppose that your trifling pleasure is good but not very good, and my suffering is very bad. So, your pleasure is better than my suffering, but it is also true that your pleasure is less good than my suffering is bad, or, equivalently, my suffering is worse than your pleasure is good.[7]

6 Conjunctive Accounts

Let us now consider the so-called conjunctive accounts of cross-cats. The most prominent and elaborated one is Ewan Klein's (1980). This account, adjusted to my terminology, holds that

"x is F-er than y is G" is true iff there is a degree modifier D such that "x is D F, and y is not D G" is true (where "D" is read as "at least D")

Applied to our initial evaluative sentences, (1)–(3), the results would be as follows:

(24)   "This table is a better table than Trump is a president" is true iff there is a degree modifier D such that "this table is a D good table and Trump is not a D good president" is true.

(25)   "You are more prudent than you are moral" is true iff there is a degree modifier D such that "you are D prudent and you are not D moral" is true.

(26) "I am more kind than I am beautiful" is true iff there is a
degree modifier D such that "I am D kind and I am not D
beautiful" is true.

As seen, Klein's account takes *Degree modification* and turns it into a
definition of cross-cats. I think this account is on the right track – after
all, it gives the right verdict about (1) to (3) – but there are still some
issues. First, the account seems to overgenerate cross-cats. Suppose that I
am very prudent and my sock, unsurprisingly, is not moral and thus not
very moral either. Then the account would imply that I am more prudent
than my sock is moral. But this is false since my sock *cannot* be moral. An
easy fix would be to say that "*x* is D F and *y* is not D G" in the analysans
and should be replaced with "*x* is D F and *y is G* but not D G."

But that would lead to other problems, for now the account can no
longer say that this table is a better table than Trump is a president. Since
Trump is a bad president, he does not fulfill the condition "is *good* but
not D good," which he has to fulfill according to the revised account.

Second, Klein's account (including the revised version) has problems
satisfying *Polarity*. Suppose, for instance, that this table is neither good
nor bad as a table but is simply *neutral* as a table. We still seem to be able
to say truthfully that this table is better as a table than Trump is good as a
president. But there is no D such that this table is D *good*, for it is neutral.

Third, it is not clear how to define cross-categorical equality in Klein's
account.

7 My Account

My account builds on Klein's but avoids its limitation. To see what my
account amounts to, let us first introduce the notions of a polarity function,
degree modifier function, and modified polarities. A *polarity function* takes
as an input a value kind and spits out a polarity property based on that
kind. There are three such functions, *pos*, *neg*, and *neut*. The output of *pos* is
a positive polarity property, a goodness property based on that input value
kind, so *pos*(F-value) = F-goodness. The output of *neg* is a negative polar-
ity property, a badness property, so *neg*(F-value) = F-badness. The output
of *neut* is a neutral polarity property, so *neut*(F-value) = F-neutrality. So,
for instance, *pos*(final value) = being finally good, *neg*(final value) = being
finally bad, and *neut*(final value) = being finally neutral. I am not assuming
that all three functions are defined for all kinds of values. Perhaps some
kinds of values do not have neutral values (are there neutral hammers?),
and perhaps other kinds do not have negative values (e.g., 'moral worth').

A *degree modifier* function is a function that takes a polarity prop-
erty as input and spits out a degree-modified polarity property. Examples
of degree modifiers functions are *very*, *moderately*, *barely*, *somewhat*,
and *not very*. So, *very*(F-good) = being very F-good, *moderately*(F-bad)

= being moderately F-bad, *barely*(F-good) = being barely F-good, and *not very* (F-good) = being F-good but not very F-good. I shall assume that there is also a *trivial* degree modifier function, *td*, such that *td*(F-good) = being F-good, *td*(F-bad) = being F-bad, and *td*(F-neutral) = being F-neutral. *td* is the only degree modifier function that is defined for neutral value properties since neutrality does not come in degrees.[8] Finally, d is a higher degree function than d' iff, for all x, y, and F-value, if x falls under d(F) and y falls under d'(F), then x is F-better than y.

A degree-modified polarity property can be seen as a *range property* since it carves out a set of items that could differ in degrees of the relevant value but share the property of being within the range. So, the items that are very F-good can differ in F-value, but they all share the property of falling within the range defined by being very F-good. For example, among the very good presidents, some are better presidents than others. A degree-modified polarity property is also *non-positional*, since whether an item is top, bottom, or *n*th in the relevant ranking is not determined by whether it exemplifies under it. That a flute is very good does not say anything about whether it is best, worst, or *n*th in the ranking of flutes.

It is important to note that the way a value adjective, such as "very F-good," picks out a range property can be a highly contextual matter. What falls under "very F-good" in one context need not fall under "very F-good" in another context. I can count as a very good basketball player in a context in which the comparison class is my elderly relatives, but not in a context in which the comparison class is the Swedish national basketball team. In my statements of the polarity and degree modifier functions, I have ignored these contextual issues to avoid unnecessarily cluttering the exposition.

Here are some examples of degree-modified polarity properties:

> *Degree-modified positive polarity properties*: being kind, being very kind, being finally good, being moderately finally good, being barely prudentially good, being good as a president, being extremely good as a president

> *Trivially degree-modified neutral polarity properties*: being finally neutral, being prudentially neutral

> *Degree-modified negative polarity properties*: being unkind, being very unkind, being selfish, being moderately selfish, being finally bad, being extremely finally bad, being prudentially bad, being somewhat prudentially bad, being very bad as a president, being extremely bad as a president

Finally, a *modified polarity* function, *mp*, is a function composed of a polarity function, *p*, and a degree function, *d*: $mp(\cdot) = d(p(\cdot))$. Examples are *very*(*pos*(·)), *moderately*(*neg*(·)), and *barely*(*pos*(·)).

We now define cross-categorical betterness in the following way:

x is F-better than *y* is G-valuable = df. there are modified polarity functions *mp*, *mp*' such that *x* falls under *mp*' (F-value), and *y* falls under *mp*'(G-value), and *mp* is of a higher rank than *mp*'.

mp is of a higher rank than *mp*' = df. for all *x*, *y*, F, if *x* falls under *mp*(F-value) and *y* falls under *mp*' (F-value), then *x* is F-better than *y*.

Note that this definition ensures that *higher rank* is irreflexive and asymmetric, given a certain *instantiation* condition: for any pairs of *mp* and *mp*', there are F, *x*, and *y* such that *x* falls under *mp* (F-value), and *y* falls under *mp*'(F-value). We can then use the irreflexivity and asymmetry of F-betterness to show that *higher rank* must also be irreflexive and asymmetric. To show irreflexivity, suppose for reductio that *mp* is of a higher rank than itself. Then, given the instantiation condition, there are F, *x* such that *x* falls under *mp* (F-value), and, given the definition of higher rank, *x* is F-better than itself, which is impossible. To show asymmetry, suppose for reductio that *mp* is of a higher rank than *mp*' and *mp*' is of a higher rank than *mp*. Then, given the instantiation condition, there are F, *x*, *y* such that *x* falls under *mp* (F-value), and *y* falls under *mp*'(F-value), and, given the definition of higher rank, *x* is F-better than *y* and *y* is F-better than *x*, which is impossible.

If we also assume that for any n-tuple of modified polarities, mp_1, mp_2, mp_3, …, mp_n, there are F, x_1, x_2, x_3, …, x_n such that x_1 falls under mp_1(F-value), x_2 falls under mp_2(F-value), x_3 falls under mp_3(F-value),…, and x_n falls under mp_n(F-value), we can also show that higher rank must be acyclical, if we also assume that F-betterness itself must be acyclical, which seems plausible for any F.[9] Suppose for reductio that mp_1 is of a higher rank than mp_2, mp_2 is of a higher rank than mp_3, …, but mp_n is of a higher rank than mp_1. Suppose also that x_1 falls under mp_1(F), x_2 falls under mp_2(F), x_3 falls under mp_3(F), …, and x_n falls under mp_n. Then, by the definition of higher rank, x_1 is F-better x_2, x_2 is F-better than x_3, x_3 is F-better than x_4, …, but x_n is F-better than x_1, which is impossible.

Admittedly, these instantiation conditions assume a certain richness in value structure that could be questioned. An alternative way to ensure that *higher rank* is well-behaved is simply to *postulate* that the relation is irreflexive, asymmetric, and transitive.

Note also that *Cross-Cat Betterness* ensures the following:

Value Category Harmony
For all F, *x*, and *y*, if there are *mp* and *mp*' such that *x* falls under *mp*(F-value), and *y* falls under *mp*' (F-value), and *mp* is of a higher rank than *mp*', then there are no *mp*'' and *mp*''' such that *x* falls under *mp*''(F), and *y* under *mp*'''(F), and *mp*''' is of a higher rank than *mp*''.

For if this conditional is false, it follows from *Cross-Cat Betterness* that for some F, x, and y, x is F-better than y and y is F-better than x, which is impossible since F-betterness is asymmetric.

What about cross-categorical *equality*? We cannot just say that

x is as F-valuable as y is G-valuable iff there is an *mp* such that x falls under *mp* (F-value) and y falls under *mp* (G-value).

This account of equality would entail that

x is as kind as y is beautiful, whenever x is kind and y is beautiful, no matter whether x is kind but not very kind and y is very beautiful.

x is as good a table as y is a president, whenever the x is a good table and y is a good president, no matter whether x is not a very good table and y is a very good president.

This captures a very coarse-grained notion of cross-cat equality. At the other extreme, we have *exact* cross-cat equality:

x is as F-valuable as y is G-valuable iff there are *mp* and *mp'* such that x falls under *mp*(F-value), and y falls under *mp'*(G-value), and *mp* has the same rank as *mp'*.

mp has the same rank as *mp'* = df. for all x, y, and F, if x falls under *mp*(F-value) and y falls under *mp'*(F-value), then x is exactly as F-valuable as y.

This is a very restrictive notion of cross-cat equality, however. Indeed, it seems as though *mp* has the same rank as *mp'* only if *mp* = *mp'* = *dt*(*neut*(·)), since any other modalized polarity function will allow for variation in the degree of F-value. If this is correct, x is as F-valuable as y is G-valuable only if x is F-*neutral* and y is G-*neutral*.

Here is a better account that avoids this problem:

Cross-Cat Equality
x is as F-valuable as y is G-valuable iff, for all *mp*, x falls under *mp*(F-value) iff y falls under *mp*(G-value) (and both x and y fall under some *mp*).

This definition makes cross-categorical equality transitive, which is an attractive feature. Note also that it entails that Mozart is as good a composer as Michelangelo is a sculptor (they are both exceptionally good in their own fields), which seems to be a welcome result. Indeed, since this holds even though we cannot say that Mozart is a more creative artist than Michelangelo, or that Michelangelo is a more creative artist than Mozart, or that they are equally creative, one may suggest that this captures a notion of *parity*. (I say *a* notion rather than *the* notion since there seem to be many different value concepts one can have in mind when one talks about parity.)[10]

Let us now turn to *inter-polar* cross-cats. I suggest the following definitions:

x is more F-*good* than *y* is G-*bad* = df. there are modified polarity functions *mp*, *mp*', such that *mp* = *d*(*pos*(·)) and *mp*' = *d*'(*neg*(·)), *d* is a higher degree modifier function than *d*', and *x* falls under *mp*(F-value), and *y* falls under *mp*'(G-value).

x is more F-*bad* than *y* is G-*good* = df. there are modified polarity functions *mp*, *mp*', such that *mp* = *d*(*neg*(·)) and *mp*' = *d*'(*pos*(·)), *d* is a higher degree modifier function than *d*', and *x* falls under *mp*(F-value) and *y* falls under *mp*'(G-value).

x is as F-*bad* as *y* is G-*good* = df. for all *mp* and *mp*' such that *mp* = *d*(*neg*(·)) and *mp*' = *d*'(*pos*(·)), *d* = *d*', *x* falls under *mp*(F-value) iff *y* falls under *mp*'(G-value).

Since my aim in this chapter is modest – just to outline an account of evaluative cross-cats – I will not say much about how to generalize my account to all non-evaluative cross-cats. Suffice it to say that the central notions, i.e., polarity functions and degree modifier functions, which together define modified polarity functions, can easily be generalized to *non-evaluative* properties. I have already done so for degree modifier functions, but polarity is also a general aspect of gradable adjectives (and the expressed properties). For example, length has both a positive polarity, being *tall*, and a negative one, being *short*. Similarly, weight has a positive polarity, being *heavy*, as well as a negative one, being *light*. Felt temperature seems to have three polarities: *warm*, *cold*, and *neutral*. (Of course, in all these cases, polarity is a contextual matter.) The polarity functions *pol*, *neg*, *neut* can thus be extended to have non-evaluative kinds as inputs, and the previous definitions can be extended to apply to non-evaluative cross-cats, *mutatis mutandis*.

The advantages of my account are the following:

(1) It has the right entailment.
If *x* is F-better than *y* is G-good and *y* is very G-good, then *x* is at least very F-good.

(2) It applies even when there are gaps in one ordering.
If *x* falls under *mp*(F-value) and *y* falls under *mp*'(G-value), and *mp* is of a higher rank than *mp*', then *x* is F-better than *y* is G-valuable, even if either the F-value ranking or the G-value ranking is gappy.

(3) It applies even when the number of F-value ranked items is not the same as the number of G-value ranked items.
If *x* falls under *mp*(F-value), and *y* falls under *mp*'(G-value), and *mp* is of a higher rank than *mp*', then *x* is F-better than *y* is G-valuable, even if the number of items that have F-value is different from the number of items that have G-value.

(4) It applies even when the F-value ranking or G-value ranking is unbounded.
If x falls under mp(F-value) and y falls under mp'(G) and mp is of a higher rank than mp', then x is F-better than y is G-valuable, even if there is no maximum or minimum in one or both of the rankings.

(5) It satisfies *Polarity*.
This is obvious.

(6) It satisfies *Degree modification*.
This is also obvious.

(7) It deals with inter-polar cross-cats.
This is also obvious.

8 Normative Applications

Now that we have a better grasp of evaluative cross-cats, it is time to see whether they have any normative relevance. Even though these new value relations would of course be usable in purely axiological matters when we are just interested in making value judgments, it would be disappointing if these value relations could not be put to any normative use at all.

One possible application is normative decisions based on comparisons of what is good for different people.[11] Suppose you are convinced, as many economists still seem to be, that there is no common value dimension on which we can compare what is good for people, as well-being is interpersonally incomparable. Then saying that x is better for me than y is for you amounts to a cross-cat. You can then use my analysis to decide what you ought to do in a choice between realizing A, which is very good for me, and realizing B, which is good but not very good for you (and no other person is affected). Since, according to my analysis, A is better for me than B is for you, this can be seen as providing a reason to choose A over B.

A second application is this. We have two positions and one applicant. One position requires prudence and the other being moral. We need to offer the candidate one of the positions. If the candidate is more prudent than moral – perhaps because she is very prudent but not very moral – then we should offer her the job that requires prudence. Now some might object that we are invoking another value consideration here: finding the best match between the candidate and the job. But this does not take away the significance of the cross-categorical value comparison since we are not using the "best match" value to define cross-categorical value comparisons; rather, the "best match" value is *based* on the cross-categorical comparison.

Another application relies on the link between values and attitudes. Even though it is controversial to define values in terms of attitudes, it is reasonable to assume that, at least for some kinds of values, if something

has a certain value, it is fitting to have a certain attitude toward it. What attitude it is fitting to have toward it depends on what kind of value it has. Sometimes it is obvious what attitude it would be fitting to have: for example, it is fitting to admire what is admirable, envy what is enviable, praise what is praiseworthy, blame what is blameworthy, and desire what is desirable.

Cross-cats can be put to use here. If x is more admirable than y is enviable – perhaps because x is very admirable and y is enviable but not very enviable, then to admire x but not envy y seems a more fitting combined stance than not to admire x but envy y. Furthermore, the stance "admire x very much, envy y but not very much" is more fitting than the stance "admire x but not very much, envy y very much." More generally, if x is more admirable than y is enviable, it is more fitting to admire x more than one envies y than to envy y more than one admires x. Note that this claim invokes yet another kind of cross-categorical comparison – namely, one of *attitudinal* strength since the strength of admiration and the strength of envy do not seem to be reducible to one common kind of attitudinal strength.[12]

The final application is evaluative uncertainty. The obvious obstacle to taking evaluative uncertainty seriously is that we can be uncertain about the truth of two radically different theories that employ different kinds of values. For example, we can be undecided between (a) a certain virtue theory (e.g., Philippa Foot's) according to which there is no impartial value – we can only talk about virtuous or vicious acts, and (b) a utilitarian theory according to which actions can be (instrumentally) impartially good, but no actions are virtuous or vicious. Suppose that the choice situation is like the following (without well-defined probabilities):

Actions	Virtue Theory	Utilitarianism
A	Very vicious	Very impartially good
B	Very virtuous	Impartially good but not very impartially good

Action A might be the joyful killing of an innocent that brings about billions of extra very happy lives. Action B is the joyful omission of killing an innocent that brings about a few extra very happy lives created by the agent out of the most virtuous motives. There are no direct value comparisons to be made across the theories since the theories employ different value concepts, but we still seem to be able to say the following:

A is impartially better, according to utilitarianism, than it is virtuous, according to virtue theory, since it is impartially *very impartially good* according to utilitarianism and *very vicious* according to virtue theory.

B is more virtuous, according to virtue theory, than it is impartially good, according to utilitarianism, since it is *very virtuous* according

to virtue theory and *impartially good but not very impartially good* according to utilitarianism.

B is impartially better, according to utilitarianism, than A is virtuous, according to virtue theory, since B is impartially *good but not very impartially good* according to utilitarianism and A is *very vicious* according to virtue theory.

These cross-cats seem to suggest that B would be a more reasonable choice since it "maximizes the minimum cross-categorical value."

More generally, the idea is that it is reasonable to choose B in a situation of the following kind (without well-defined probabilities).

Actions	Theory T1	Theory T2
A	Very T1-bad	Very T2-good
B	Very T1-good	T2-good but not very T2-good

Of course, maximin is not uncontroversial, but this is just one example of how cross-cats can be used to guide action in certain cases of evaluative uncertainty. Other principles that take into account both (qualitative) probability or plausibility and ordinal value information can be used instead. The scope of application of this approach to decision-making under evaluative uncertainty goes beyond that of existing approaches, since they tend to assume either that no value comparisons can be made across theories or that they can made only when the theories employ exactly the same value concept.[13]

9 Concluding Remarks

The conclusion is that, notwithstanding the old saying, we *can* compare apples with oranges! Joking aside, there seems to be room for cross-cats that can be put to normative use. Of course, much more needs to be said to bolster this claim and show how it fits into a general account of cross-cats, both evaluative and non-evaluative. I have provided only a very rough sketch of how to make sense of evaluative cross-cats. One issue that needs to be addressed is that cross-cats seem more natural and plausible when they involve similar value concepts, such as good as a table and good as a president, or prudentially good and morally good; indeed, one could claim that there is a specific *second-order* value concept that unifies the value concepts in each pair: goodness of its kind, and goodness of persons, respectively. Cross-cats that invoke very different concepts and very different value bearers seem less natural and plausible. For example, it seems odd to say that my excellent wine is a better wine than my having a dull life is a good state of affairs. So, one option is to add

restrictions to my account. For *x* to be cross-categorically comparable to *x* and *y*, they have to fall under similar (but nonidentical) value concepts and be similar value bearers. Of course, for this to be useful, we need to decide what "similar" means.

Another remaining question is why cross-cats seem more compelling when they involve objects with *extremal* values. It is very compelling to think that *x* is F-better than *y* is G-valuable when *x* is extremely F-good and the very best F one can imagine, whereas *y* is extremely G-bad and the very worst G one can imagine.[14] We know this cannot just be a question of their relative positions, so we need some other explanation of the compellingness.

Even though there are remaining issues, I hope that I have at least shown that cross-categorical comparisons should be taken seriously – and not simply dismissed as ungrammatical, meaningless, or merely metaphorical – and that my account is an improvement on the existing ones.[15]

Notes

1 I assume that prudent, moral, kind, and beautiful are thick evaluative terms.
2 It might be clearer to say, "than Trump is *good* as a president."
3 Kraut (2011: 153) claims that it makes sense to say, "This watermelon is a better watermelon than that poet is a poet," but he does not provide an account of such comparisons.
4 For more on the notion of covering value, see Chang (1997).
5 See Chang (2002).
6 The (in)famous zero-one rule for interpersonal comparisons of well-being, according to which everyone's top-ranked options are assigned 1 and everyone's bottom-ranked options are assigned 0, satisfies an analogous principle and faces problems analogous to those that afflict the positional account. For more on critical discussion of the zero-one rule, see Hausman (1995).
7 Here "less good" does not mean "worse," which it can do in some contexts.
8 These degree modifier functions are the standing meanings of their *linguistic* counterparts "very," "moderately," "barely," and so on. A linguistic degree modifier, such as "very," retains its meanings across different constructions, even if it is applied to different evaluative adjectives. For example, "very" has the same meaning in "very prudentially good" as in "very finally good."
9 At least this holds for the notion of betterness that is the comparative of "good," which is my focus here. For an alternative non-transitive notion of betterness, see Temkin (2012: ch. 1).
10 For more on the notion of parity, see Chang (2002).
11 I would like to thank Gustaf Arrhenius and Michael Morreau for suggesting this application.
12 See Rabinowicz (2012: 153) for a brief discussion of what he calls "cross-kind" commensuration of attitudes.
13 MacAskill et al. (2020) offers an opinionated introduction to existing accounts of decision-making under moral uncertainty.
14 I would like to thank Joe Roussos and Wlodek Rabinowicz for alerting me to this.

References

Bale, A. (2006), *The Universal Scale and the Semantics of Comparison*, PhD thesis (McGill University).

Bale, A. (2008), 'A Universal Scale of Comparison', *Linguistics and Philosophy* 31: 1.

Chang, R. (1997), 'Introduction', in R. Chang (ed.), *Incommensurability, Incomparability, and Practical Reason* (1–34). Harvard University Press.

Chang, R. (2002), 'The Possibility of Parity', *Ethics* 112/4: 659–688.

Hausman, D. (1995), 'The Impossibility of Interpersonal Utility Comparisons', *Mind* 104/415: 473–490.

Klein, E. (1980), 'The Semantics of Positive and Comparative Adjectives', *Linguistics and Philosophy* 4: 1–45.

Kraut, R. (2011), *Against Absolute Goodness* (Oxford University Press).

MacAskill, W., Bykvist, K., and Ord, T. (2020), *Moral Uncertainty* (Oxford University Press).

Rabinowicz, W. (2012), 'Value Relations Revisited', *Economics and Philosophy* 28/2: 133–164.

Temkin, L., (2012), *Rethinking the Good* (Oxford University Press).

9   What Does Incommensurability Tell Us About Agency?[1]

Luke Elson

Terminology is a minefield here: one must choose between "incommensurate," "incommensurable," and "incomparable," at least. The terminologies are incommensurate, but "incommensurate" has stuck, so I use it. Two items or options A and B are *V-incommensurate* when with respect to some value V, A is not better than B, B is not better than A, and A and B are not equally good. V-incommensurability is the failure of all three *trichotomous* comparisons (better, worse, and equally good) with respect to V.[2]

Examples of incommensurability almost always involve two options or items that each do better along different dimensions of the relevant standard or value. Cooking at home is healthier and cheaper, but ordering delivery is tastier and more convenient. In some classic cases – Mozart versus Michelangelo in creative terms (Chang 1997: 15) or Stonehenge versus St. Peter's in terms of impressiveness (Broome 1997) – it's not even clear how to specify the relevant dimensions in an informative way.

We could be epistemicists about incommensurability: maybe in all of these cases one of the comparisons does apply, but we don't know which. A version of this view concerning the Moorean good has been defended by (Regan 1997) and deserves respect, but I set it aside to ask a conditional question: if none of the trichotomous comparisons determinately holds, then what are the implications for agency?

Very often, we don't simply investigate which option is better with respect to some value – we also choose between options. When acting, it's very rare that we have no choice at all (our options can include stalling for time, doing nothing, or doing something completely outrageous). Much of the time, though, the best option is so obvious that we don't even waste any time consciously deliberating or thinking it over. See Arpaly (2002) for a classic discussion of the prevalence – or otherwise – of deliberation in our actions.

I will assume that we are making a choice with respect to some value, but I will suppress mention of the value, which doesn't matter for my purposes. The crucial point is the connection between the value and

our reasons: we ought to choose the best option (if there is one), and if A is better than B, then we have a stronger all-things-considered reason to choose A than B. Hence, I will write interchangeably of values and reasons.[3]

When one of the trichotomous comparisons obtains, and given these stipulations, the situation is fairly clear. If A is better than every other option – if A is *strongly dominant* – then only Aing is permissible and doing anything else instead would be a mistake. But sometimes, one option A is merely *weakly dominant*: at least as good as every option. This can be because every option is equally good, or because there are two or more options (including A) that are as good as each other but better than all others. When we are choosing cans of beans of the same brand at the supermarket, there are likely to be many cans tied for best and perhaps some that are worse (the dented ones). Buying the worse beans is impermissible – not morally so, but it is forbidden by our reasons. It ought not be done.

If the set of options is not infinite (and the betterness relation is not cyclic), then we know that there must be at least one option that is not strongly dominated: there is no option better than it. But under incommensurability, there may be no weakly dominant option. There may instead be (at least) two options A and B, incommensurate with each other – and so not weakly dominant – but better than all other options.

It may be, for example, that cooking for yourself and ordering takeaway are each better than the other options – going hungry, foraging – but are incommensurate with each other. Or if you are engaging in a spot of post-pandemic travel, the ranking of your options may have a fast express train and a cheaper slow train incommensurate with each other but better than all other options (walking, a fast express train in the wrong direction, … most of our options are so silly that we don't even bother to think about them).

In such cases, two pieces of phenomenology seem to pull in opposite directions. First, either option seems permissible – certainly, neither is criticizable. If neither cooking nor ordering takeaway is better than the other, aren't both acceptable? And when quizzed about your choice by a nosy neighbor, "well cooking would have been healthier, but delivery was more convenient so I went for that" could be a perfectly reasonable justification. (Your neighbor might be suggesting that you should value your health more than you do, but that's a different point, and even more nosy.)

Second, except perhaps in the most trivial choices (such as what to have for dinner), we don't stop worrying when we reach the conclusion that two options are incommensurate. We typically continue to agonize and deliberate about a choice between them, at least in major choices. Even if we have to choose *right now*, and therefore pick arbitrarily, we might retrospectively think the choice over. If you are like most people, when house hunting, there are several options open to you that are incommensurate

with each other but dominate all others: the smaller place closer to work, the bigger place further away, the fixer-upper, and so on. But with such a major financial commitment as a mortgage or lease, coming to think that your remaining options are incommensurate does not simply stop your deliberation. Most of us do not at that point simply roll a die. Instead, we think it over. (We might roll a die in the food case, but not all of us would in the train case. How much you agonize about your travel options has a lot of individual variation; I would probably enjoy travel more were I to simply relax about such choices as which train to take.)

These two pieces of phenomenology are not contradictory. The former is about the permissibility of actions, whereas the latter is about the decision procedure – and not even about which decision procedures are permissible or appropriate but about which ones we in fact employ. But the two are certainly in tension: if both options are permissible, then why worry and deliberate about the choice?

Joseph Raz and Ruth Chang have each drawn quite far-reaching conclusions about the nature of human agency from incommensurability, and in particular from something like these two pieces of phenomenology. After criticizing their arguments, I'll argue that if we see incommensurability as vagueness – a view defended by (Andersson 2017), (Elson 2017), and most famously (Broome 1997) and Broome in this volume – then its implications for agency are minimal. I don't pretend to offer a rigorous defense of my alternative but simply to explore how I think about choice under incommensurability, and to sketch a reasonably plausible picture.

1 Raz's Classical Conception

Joseph Raz distinguishes two views of human agency. The first is what he calls *rationalism*: "Paradigmatic human action is taken because, of all the options open to the agent, it was, in the agent's view, supported by the strongest reason" (Raz 2002: 47).

Rationalism doesn't say that in paradigmatic human action, we do the action that is *actually* supported by the strongest collection of reasons. Instead, the action that the agent believes to be most supported by reason. And as we are often mistaken or uncertain, these need not be the same. But as I am setting aside epistemicism, I will focus on cases where her beliefs are accurate. What does "paradigmatic" mean, here? I take it to refer to the central, "normal" case, when things are working as they should. There may be exceptions, but they are either relatively isolated oddities or involve some kind of error:

> **Rationalism.** Normally, rational actions are those where the agent chooses what she believes to be the strongly dominant option because she believes it to be strongly dominant.

The rationalist picture is that there are many reasons for and against various actions, and the action one ought to do is that one supported by the strongest collection of reasons, after all the weights are in. (I'm setting aside here "enticing reasons," which would complicate this picture, and of which I am skeptical. See (Dancy 2006) for a defense of such reasons.) A huge part of the metaethics of the past 25 years or so consists of arguments about what these reasons are: are they just desires (or grounded in desires), or facts about impartial welfare, or particularist facts, or …? Rationalism says that there is little distinctive role for the will other than acting in compliance with the reasons we have.

There can be some exceptions to rationalism's "normally." Cases where there are multiple weakly dominant options – such as Buridan's Ass – are not a major challenge because those options remain, in Raz's words, "supported by the strongest reason." Weak dominance is not a serious threat to the rationalist view.

The main threat is incommensurability of an uncompromising form:

> **Hard Incomparability**. A and B are hard-incomparable when each of the three trichotomous comparisons determinately fails to apply – A is determinately not better than B, B is determinately not better than A, and A and B are determinately not equally good – and in virtue of this, *no* comparison holds between A and B.

In hard incomparability, all three trichotomous comparisons *determinately fail to apply*, and no other comparison applies either. The position of "determinately" is crucial here because the broader definition of incommensurability says only that the three trichotomous comparisons each *fail to determinately apply*. This is the difference between "she is not determinately tall" (which leaves room for it to be indeterminate whether she is tall) and "she is determinately not tall" (which doesn't).

(Raz 1986) argues that incommensurability is hard incomparability. If this is so, and if incommensurability is commonplace, then rationalism starts to look untenable because there will often be no option that's even weakly dominant. As hard incomparability becomes more widespread, the picture of normal action as reason narrowing down the options to a few dominant ones looks increasingly strained. Some instances can be accepted – if they are, as (Raz 2002: 48) puts it, "relatively rare anomalies" – but rationalism is a distortion if there not being any weakly dominant option is the *more* common situation.

The more we move in this direction, the more support is lent to Raz's second picture of agency, which he calls the *Classical Conception*: "paradigmatic human action is one taken because, of all the options the agent considers rationally eligible, he chooses to perform it" (Raz 2002: 47). Here is my gloss:

Classical Conception. Normally, rational actions are those where the agent chooses from among actions she believes to be rationally eligible.

To say that an action X is "rationally eligible" is to say that X is not strongly dominated, and so to choose X would not go against our reasons: it's not the case that we ought not X.

On the Classical Conception, even after the strongly dominated options have been eliminated, there typically remains a wide range of rationally eligible options. If there are five incommensurate takeaway food options in Reading, but each of these five is better than all other restaurants, then reason will eliminate all but those five options. They are rationally eligible.

The pictures have quite different underlying moral psychologies. Rationalism implies that in normal cases the will has no role beyond executing one (strongly dominant) or perhaps one of several (weakly dominant) choices. Raz argues that pervasive incommensurability construed as hard incomparability makes this picture a distortion because the will must often choose from several rationally eligible but incommensurate (and thus not even weakly dominant) options.

He argues that incommensurability is indeed widespread because desires are the only feasible candidate reasons (or source of reasons) that could remove all – or nearly all – incommensurabilities, but desires don't provide reasons. Here is how I understand the master argument of (Raz 2002):

(1) Desires don't provide reasons.
(2) But desires are the only plausible candidate for commensurating values.
(3) So reasons normally leave a number of hard-incomparable options.
(4) Which supports the Classical Conception.

I've defended the inference from (3) to (4) as plausible, but I've not engaged with the rest of the argument. Next, I'll argue that if we assume that incommensurability is vagueness – in particular, that it's often *indeterminate* which option is better – then widespread incommensurability can be reconciled with something close to rationalism, because (3) is false.

2 Chang's Hierarchical Voluntarism

Raz draws from incommensurability the thought that the will can choose without (and after) reason. Ruth Chang draws a different lesson: agency may involve creating reasons where there were none.

Chang argues that the trichotomy doesn't exhaust the available comparisons:

> **Parity**. Sometimes, options are trichotomously incomparable but nevertheless comparable because a fourth comparative relation called "Parity" holds between them.

According to Parity, at least sometimes, incommensurate options are *on a par*. Though central to Chang's overall view, the Parity claim is not crucial to the deliberation argument that is my topic.

I will here discuss her view as defended across a series of papers, including (Chang 2013) and (Chang 2017), but my main focus is the fascinating argument of (Chang 2009). That argument for voluntarist creation rests on the angstiness and continued deliberation discussed earlier: coming to think that our best options are incommensurate with each other doesn't usually stop deliberation in its tracks, as coming to think that they are precisely equally good does. Importantly for Chang, we don't always think this continued deliberation under incommensurability *irrational*, as, for example, a waste of mental resources would be.

To explain the rationality of continued deliberation under incommensurability, Chang argues that deliberation has more functions than we thought. I join her in a broad conception of deliberation as including "discovering, recognizing, investigating, appreciating, and engaging with the reasons there are," but she adds to this list the voluntarist *creation* of new reasons (Chang 2009: 259). There are other voluntarists, but her voluntarism is distinctively hierarchical: there are "given" (non-voluntary) reasons, but they sometimes leave incommensurabilities. Voluntarist reasons can only rationalize or justify actions that are not strongly dominated, considering only the given reasons.

Here is how I picture it: the given reasons mark out the fences of the playground within which we must act, but they don't tell us what to do within that playground. Whereas for Raz the will must now simply choose what to do, for Chang, the faculty of reason can – through its capacity to deliberate – *create* or endorse reasons to do one thing rather than another, remaining in the playground. So the view is rationalist in a certain sense: paradigmatic action involves acting for what we take to be the strongest reasons, but it's (within limits) up to us which reasons are strongest. Voluntarist creation is not required, and instead, one may simply "drift" – (Chang 2017: 19) writes that "law schools are populated with drifters."

Why be suspicious about voluntarism? Because it says that we can create reasons. Of course, we *can* sometimes create reasons: I can make it so that I have reason to give you £10 where before I did not by promising to give you £10. But this is changing the facts of the matter – I have created a promise where before there was none. Voluntarism typically involves the more radical claim that I can change or create the reason-giving force of the facts. But it's far less clear – and more objectionable to anti-voluntarists – is that I could create a reason to give you £10 without doing

some action like promising. Promising is quite different to an act of "willing [that] creates normativity by creating new reasons whose normativity derives from the very act of will" (Chang 2009: 255).[4]

Chang's view has the advantage that because one can only create voluntarist reasons within the area marked out by the given reasons, there is no way to create voluntaristic reasons to commit murder or do other heinous things (presuming that they are forbidden by the given reasons). If the anti-voluntarist objection to normative powers is really a worry that such powers threaten to allow us to make all sorts of unfortunate reasons – such as to commit murder – then Hierarchical Voluntarism looks immune to it – see (Chang 2009: 270). Murder is outside the playground. But if the worry is instead an expression of discomfort with us having the ability to create *any* reasons, outside of promising and the like, then the hierarchical view doesn't mitigate the worry, or not very much.

But my main focus is not Hierarchical Voluntarism on its own terms but whether deliberation under incommensurability supports it.

2.1 The Deliberation Argument

Chang argues for Hierarchical Voluntarism by appeal to our second piece of phenomenology – the angstiness that attends choice under incommensurability and continued deliberation. Here is my understanding of the argument of (Chang 2009: 249ff):

(1) Sometimes our reasons run out.
(2) When they do, we often keep deliberating.
(3) Unlike with continued deliberation under betterness or equal goodness, we (often) consider this continued deliberation under incommensurability to be rational.
(4) The best explanation for (3) is that unlike under betterness or equal goodness, deliberation under incommensurability is not a waste of time.
(5) The best explanation for (4) is that deliberation under incommensurability involves the voluntaristic creation of reasons.

Though it captures the logical structure of the argument, this presentation overstates the argument's intended strength, which is intended to be more suggestive than abductive.[5] And Chang defends the hierarchical or hybrid voluntarist view elsewhere too, but I focus on this argument because though I am not convinced by it, it deserves investigation and is (along with Raz's Classical Conception) one of the few attempts I am aware of to seriously explore the connections between agency, deliberation, and incommensurability.

My objection will be to premise (3): I'll argue that we sometimes rationally continue deliberating under trichotomy, and the explanation for this

fact is similar to why we rationally continue deliberating under incommensurability and doesn't involve normative creation. Deliberation is an activity with its own costs – typically, the opportunity cost of spending time on it, but also other psychological downsides – but also its own benefits, which include but go beyond a greater chance of choosing the best option.

Since part of deliberation's point is information-gathering, I won't assume when discussing the deliberation argument that the agent in question has full information about her choice. If the stakes are high enough, then deliberation can continue to be rational even when one option appears best. When I presented this chapter at a conference in Stockholm, I was also planning a trip to Texas to visit family and could only find extraordinarily expensive travel insurance. After a reasonable amount of time, I found what seemed to be the best insurance policy ("Policy X") – the cheapest one that met some quality thresholds. But I continued to deliberate: rather than simply buying that apparently best policy, I continued to search price-comparison websites and so on, thinking about which policy would really be best.

Finally, I bought Policy X. I knew this was likely to be the outcome; in the past, I have not usually been able to beat price-comparison websites. Does this mean that my continued deliberation was irrational? I'm not sure. Certainly, there was an opportunity cost to it. But deliberation served several functions. I *looked* for a cheaper policy, spent time wondering how comprehensive my insurance really needed to be (what my quality thresholds should be), and so on. Assuming that my continued hemming and hawing didn't simply distort my preferences or lead me into irrationality (a non-trivial assumption), the continued deliberation meant that I engaged more with the reasons for and against various policies and searched for policies I had not already seen, increasing the likelihood that I would find the best policy. I believed that Policy X was the best policy, but I was not certain, and the deliberation raised my credence in that proposition.

I also became surer that I was not wasting money, that the policy I settled on was in fact best, and that I wasn't making a costly mistake. This reassurance is worth something and may rationalize some continued deliberation. So perhaps I would add "confirming" to Chang's list of the functions of deliberation (discovering, recognizing, investigating, appreciating, and engaging with the reasons there are, and – perhaps – creating new reasons).

Sometimes the costs of making the wrong choice are so high as to swamp the cost of continued deliberation. Neither costs nor benefits need be financial: some of us find comparing insurance policies oddly fascinating, while also deploring its necessity as a feature of gotcha capitalism. All told, sometimes it can make sense to stay and re-check one's calculation: there is a definite cost to this, not least opportunity cost but also the

prospect of a gain. Continued deliberation can bring net benefits, either by making it more likely that we choose the correct option or because of deliberation itself having non-causal benefits.

On the other hand, sometimes even when it's clear that continued deliberation *would* produce a better choice, deliberation is not worth it and thus not rational. I buy my fruit and vegetables from the local Aldi. Aldi is a discounter, so – inevitably – there are some bruised or otherwise damaged bananas in the large tub that one must reach down into. I *could* achieve a better banana-outcome by deliberating for longer about which banana to take. But Aldi has decent quality controls, so I know that the truly undesirable bananas are rare, and such deliberation takes time. So it would be (even given my relative frugality) irrational for me to deliberate about each banana, not to mention about every item in my weekly shop. The food outcome would be better, but the opportunity cost and mental strain of the deliberation would be large. A quick survey to eliminate any obviously inferior bananas is the best strategy. Applied to nearly every item, this strategy will get me out of the store in a decent amount of time, with several bags of acceptable produce.

But only *nearly* every item. There are some goods in the shop that repay careful examination. The most obvious one is the avocado: a bruised or damaged avocado is near-unrecoverable, and so avocados need careful inspection. I'm laboring this point because it illustrates one thing: the difference between avocados and bananas is *not* that the stakes are higher when choosing the former (they are, but only a little). The difference is that things are more likely to go wrong with avocados, so even given roughly equal stakes, the *expected payoff* of deliberating about them is higher. Even under undoubted cases of trichotomy, sometimes deliberation is rational and sometimes not, and the stakes are not the only things that vary.

What about continued deliberation under equal goodness? It must be conceded that if we are certain that two options are precisely equally good, then it would be perverse to continue to deliberate between them. Buridan's Ass should simply plump for one of the meals.[6] But when it comes to complex choices, we almost never encounter cases of equal goodness. The cans of beans at the supermarket might be precisely equally good, but basically any complex choice with multiple dimensions can engender incommensurability. I challenge you to think of a multidimensional comparison where two different options – no tricks, no implausible stipulations – are precisely equally good, despite being better along different dimensions. (And even in the can of beans case, if we look closely at the shelf, we might notice they vary slightly in appearance…) All this is by way of arguing that if two options are precisely equally good then continued deliberation would nearly always be irrational but that this is likely only to happen in cases where the options are qualitatively identical.

My argument has been basically epistemic: the main benefits of deliberation are more knowledge (or certainty) about our situation and our reasons, and the consequent ability to act better; sometimes when the stakes are high, these can justify continued deliberation even when the facts are apparently all in.

Chang, of course, is not blind to the epistemic possibility. She argues that we can be *practically certain* that we are facing a hard choice, and that the deliberation argument can be run from this point: if it is rational to continue to deliberate beyond practical certainty, then there must be something else going on:

> Although we may never be in a position to know, in some strong sense of "know," that our reasons have run out in any particular case, we can, however, be *practically certain* that they have. If you are practically certain that p, it is irrational for you to act on the assumption that not p.
>
> (Chang 2009: 250–251)

But need this state of practical certainty rule out further deliberation as irrational? Deliberation about a choice involving p is not straightforwardly acting on the assumption that not p. And even if I'm *absolutely* certain that p, deliberating about what to do (where this depends on whether p) need not be irrational. As I've tried to argue, whether it is depends on the costs and benefits of continued deliberation, and they are not just epistemic. Consider:

> **Prison Escape.** I'm locked in a cell and must wait an hour to make my escape attempt. I'm certain that it'll be better to bash one guard over the head with a flowerpot rather than lock him in the bathroom, and these are my only two options, but the opportunity to do either will not arise for an hour.

Is it really irrational for me to continue to deliberate? I don't think so. Of course there is the opportunity cost of not composing a prison sonnet or proof, but there are benefits too, such as confirmation that I am taking the right course and planning for various contingencies. (Some of this might stretch the boundaries of "deliberation," but I think they fall within the broad scope of *deciding and planning how to act*.)

Prison Escape is extreme, but the point is general: granting Chang's definition of practical certainty, practical certainty *need not* rule out deliberation as irrational. So even if we are practically certain that we are dealing with incommensurability, the apparent rationality of continued deliberation doesn't show that deliberation includes more than fact-finding or other non-voluntarist activities. And they seem rational even under trichotomy.

That is half of my response to the deliberation argument for Hierarchical Voluntarism. The other half is much shorter: if the stakes are small enough – or the expected payoff of deliberation small enough, more generally – then even under incommensurability, continued deliberation is not rational. Think again about bananas. Suppose that you are going to buy one bunch of bananas, and each bunch costs £1. One bunch looks in just slightly better condition: a little riper, a little less bruised; the other bunch is bigger. Both are perfectly usable, but they are each better along different dimensions. Assuming that this is just one item on a regular weekly shop, should you continue to deliberate among these incommensurate bunches of bananas? Of course not: getting a notebook and food scale and estimating some kind of quality-adjusted weight (akin to quality-adjusted life years in health care) for each banana would clearly be an irrational waste of your time.

That example was silly, but the overall argument is that there is no general connection between the presence of incommensurability and the rationality of continued deliberation beyond the point of well-founded belief about how your options compare.

There are several ways Chang could go here, such as restating the argument with a tweaked definition of practical certainty that p, as implying explicitly that continued deliberation about whether p is irrational. But I think this is not a plausible route. As the Prison Escape case shows, this kind of "deliberation-forbidding" practical certainty is far rarer. Moreover, one of the assumptions of the current dialectic is that deliberation-forbidding practical certainty does *not* hold during many hard cases: the apparent rationality of continued deliberation under incommensurability is what we are trying to explain.

3 Indeterminist Rationalism

I'll now assume that incommensurability is vagueness, and that vagueness is indeterminacy: in the standard case, it's neither true nor false that (i.e., indeterminate whether) A is better than B or that B is better than A. (It may be either false or neither true nor false that A and B are equally good.)

But what upshots does this view have for agency? What I'll call *Indeterminist Rationalism* says that rationalism is true, but it's often indeterminate what we have most reason to do:

> **Indeterminist Rationalism (IR).** Rationalism is true, but it is often indeterminate which options (strongly and weakly) dominate others.

IR is somewhere between the rationalist and classical views. For Raz (and for hard incomparability in general), there is often determinately *no* best option because the ranking is incomplete. Under the indeterminacy view,

on the other hand, it is indeterminate *which* complete ranking is the correct one: there are multiple candidate rankings. Under the broadly supervaluationist view I favor, this has the upshot that "there is a complete ranking of the options" is true because it's a supertruth: it is true on every candidate ranking, though the rankings disagree about what that complete ranking is. (Compare: "there is a precise minimum number of hairs to be non-bald" is supertrue according to supervaluationism, because true on every sharpening of "bald"; the sharpenings disagree about that number, however.)

I'll argue that IR can neatly explain the two pieces of previously identified phenomenology – permissibility and rational continued deliberation. First, the permissibility intuition. Any sensible account of rational action under indeterminacy will say that when it's indeterminate whether A or B is best (and determinate that no other option is best), Aing is *acceptable* and Bing is acceptable. Why the new terminology? Because there is some dispute about whether in such cases Aing and Bing are each permissible or merely borderline permissible. The most liberal kinds of rules say that both are permissible because E-admissible (that is to say, best on one sharpening); (Rinard 2015) has defended a view where in such cases Aing and Bing are each indeterminately permissible. But even on her view, all non-E-admissible options are impermissible.

Moving beyond an appeal to authority, I will say a few words about why I find a liberal account of action under indeterminacy plausible. First, looking at the structure of the case, if it's indeterminate whether A or B is best, but determinate that every other option is strongly dominated, then if you A or B, then it's indeterminate whether you've done the best thing. If you do anything else, it's determinate that you did not do the best thing. Since there is no option that is determinately best, if you A or B, there is no option to which a critic can point and say "you ought to have done that instead," and say something determinately true. Whether you A or B, you have done the best you can in your situation because the best you can do is what is indeterminately best (to flirt with paradox).

Second, permissive judgments about actions seem clearly right in everyday cases of indeterminacy. Consider a sorites forced march: you are walked along a row of men, starting with the very shortest, and each taller than the next by 1 mm. Eventually, you are walking past some very tall men indeed. You are asked to comment accurately on whether each man is tall or not. Let's assume that you do comment correctly on the clearly not-tall and the clearly tall men. If that is not the case, then you have determinately failed to complete your task correctly. Beyond that assumption, our issue concerns the penumbra, of borderline-tall men: your answer to "is this man tall?" should turn from No to Yes at some point in that penumbra. I think it is intuitively obvious that you are not criticizable if your verdict changes somewhere in the penumbra.

We see how aspects of the Classical Conception are retained. Under a liberal decision rule, there will very often be choices not determined by reasons between options that are indeterminately, weakly dominant and thus plausibly described as rationally eligible.

Our second piece of phenomenology is the connection between incommensurability and continued deliberation. As I argued earlier, there is no *necessary* connection: sometimes deliberation is rational under trichotomy, and sometimes it's irrational under incommensurability. IR simply appeals to the thought that deliberation is an activity with its own costs and benefits.

But it would be churlish to deny that there does seem to be *some* connection between incommensurability and something like continued deliberation. IR explains this by appeal to some peculiar features of indeterminacy.

First, deliberation is rational when its expected benefit exceeds its expected cost. But under indeterminacy, it's often non-obvious or even indeterminate whether this is so: the net expected payoff of each choice will itself be indeterminate, and thanks to the phenomenon of second-order vagueness it may not even be clear *whether* we are facing an instance of incommensurability. It may also often be indeterminate how much deliberation costs (how much do you value your time, as against one of the goods at stake in the choice you face?) and so often indeterminate whether continued deliberation will bring some expected benefit.

Second, what looks like continued deliberation may simply be puzzlement. Indeterminacy is a puzzling, paradoxical phenomenon (at least, paradoxical in those cases where it engenders a sorites) and it shouldn't be surprising that we might linger when confronted with a practical manifestation of it. We can see this by confronting avowed cases of choice under vagueness: suppose, for example, that I ask you to choose the shortest tall man in this crowded room. Assuming that there is some indeterminacy about who is that man, it may seem obvious that you should simply pick a borderline-tall man. And yet I think in normal circumstances, you will "deliberate" for some time, perhaps leading yourself on some sorites forced marches (to make this vivid, suppose that each extra mm of height costs you £1). Eventually, you will likely settle on an arbitrary man in the penumbra of "tall man": it is indeterminate whether he is the tallest short man, and it is indeterminate whether the slightly taller man next to him is the tallest short man.[7] Your arbitrary choice is no better than the choice you could have made when initially confronted with the problem, and yet we would not judge you irrational for spending some time on it. The slogan view of IR is that it makes us stop and think, for much the same reason as indeterminacy everywhere can make us stop and think and puzzle. And as I'll argue in the next section, if Humean accounts of reasons are true, then a broadly voluntaristic explanation of continued deliberation is indeed sometimes plausible.

It might be objected that I'm simply assuming that questions of action under indeterminacy don't themselves create any puzzling questions about agency. There is some truth to this but perhaps not when it comes to rationalism: if the indeterminacy *in our reasons* requires some theorizing about agency, it will nevertheless not take us away from the core rationalist claim that rational action is acting for what we take to be the strongest set of reasons.

4 Indeterminist Humeanism and Naturalistic Voluntarism

To finish, I will explore the upshot of IR when paired with the view that desires provide reasons:

> **Indeterminist Humeanism (IH)**. IR is true, and all reasons are provided by desires.

There are many versions of "Humean" or "internalist" accounts of reasons, but they center on the idea that somehow, an agent has a reason to do A *iff* Aing would promote the satisfaction of one of that agent's desires. The classic of the genre is (Williams 1981), but there are many internalists of different kinds.

I won't engage in detail with arguments against desires as reasons, except to concede that they need an answer, but I will mention one of them. A common objection to Humeanism is that it would mean that deliberation about what to do is at base consideration of what we desire most, insofar as we deliberate about our reasons. This is often felt to be implausible, and in response, some versions of Humeanism say that though reasons depend on desires, it is not true that the desires *are* the reasons. See (Schroeder 2007) for an influential and sophisticated Humean response to this worry.

I do not find it implausible that all deliberation is at least partly about what we desire. For example, I was unsure whether to travel by train or by plane to Stockholm for the conference. I dislike flying, both for its hassle and its environmental damage, but this rail journey would take approximately two full days, and with a child at home, I decided in the end to fly. I deliberated about this for some time, and I have no difficulty construing this deliberation in terms of desires: "do I really want to be stuck on a train for two days, burning money on food out of boredom?" seems to me a wholly accurate description of my deliberative activity. It's not *obviously* false that we always deliberate about what we desire most.

But enough about the Humeanism. What about the indeterminacy? IH will then locate the source of – at least some – incommensurability in indeterminacy in our beliefs and desires. If my desires are imprecise, then they may be sharpened in various ways, and on some sharpenings of my

desires I prefer one house and have most reason to buy that one, and on other sharpenings I prefer the other and have most reason to buy *that* one. I want both a comfortable place to read and a short commute to work, but precisely how do these desires weigh against each other? How many commuting minutes am I willing to give up for an extra square meter of writing space? And so on. IH says that the answers to these questions are indeterminate, and so choices that depend on those answers will also manifest indeterminacy, in a version of IR.

IH engenders a *naturalistic quasi-voluntarism* because what I desire is at least partly within my control – and if my desires change, then my reasons change.

It is voluntarism, in the sense that if I can change my desires, I can change the normative valence of some fact, without changing the facts of the matter, unlike in the promising case. If I can give myself a desire that some car is red, then I can give myself a reason to get out the spray paint. But it is naturalistic and only *quasi*-voluntaristic because changing my reasons relies on causal, contingent mechanisms for changing my desires – there is no direct rational "endorsement" of some feature or fact that makes it a reason. To give myself a reason to eat vegan food, I may undergo hypnosis or choose to reflect deeply and vividly on the methods of industrial farming, believing that this will strengthen my desire to avoid many kinds of animal products. But the naturalistic nature of the view means that the process may not be predictable: perhaps I will simply be inured to the brutal realities. Animal farmers are not typically vegan, though there could be many explanations of that fact.

I have limited voluntary control over how my reasons evolve. As we have just seen, our desires may evolve in unpredictable ways – though ones that can seem to be retrospectively inevitable – we may move to the countryside in search of more space and then either develop a taste for rural life or come by grim experience to hate commuting. These developments are not only unpredictable but also chaotic and random: it may be that you hate commuting this year because of the pandemic and because your next-door neighbor – who takes the same train as you – is obnoxious. But had you moved to the countryside next year, after the vaccine and after said neighbor moved away, you'd grow to appreciate the peace of a commute.

The point is that you can at least partly affect what you desire, and thus what you have reason to do. But doing so will not always be permissible. In particular, "strategic" deliberation, with the aim of changing your desires and thus your reasons, will often be irrational. If at time t_0 you have most reason to A, which strongly dominates all your other options, then it is impermissible to engage in a path of deliberation that will make you less likely to A. If you have most reason to quit smoking, then it's impermissible to engage in visualizations about how horrible withdrawal will be, visualizations that make it less likely that you will

quit. This is because the deliberation *dis-promotes* the outcome that you quit smoking: it makes it less likely that you comply with your reasons.

But, of course, if your deliberation is successful, then your desires may have sufficiently changed that you now at t_1 have most reason to not-A: most reason to keep smoking. Yesterday, strategic deliberation in favor of smoking was impermissible; today, strategic deliberation in favor of *quitting* is impermissible. The deliberation has been a sort of transformative experience, similar to that described by (Paul 2014). But this shouldn't be too much of a surprise because according to the Humean view I'm appealing to, our reasons may often change.

Indeterminacy permits a little more strategic deliberation. If at t_0 it is indeterminate whether I have most reason to A or to B, then assuming a permissive decision rule, not only is it permissible for me to A or to B, it's also permissible for me to go to a therapist who will strengthen my desire to A or to her colleague who strengthens B-desires. But doing so is optional: indeterminist "drifting" is often respectable. And taking the wider view, fixing my reasons is not always permissible – the cost of therapy may exceed any benefit of having determinate reasons.

Here we see a kind of hierarchy in the voluntarism: if it's indeterminate whether I have most reason to A or to B, I may only permissibly do things that make it more likely that I A or that I B. Typically, there will be a general causal connection between "getting more reason to X" and "becoming more likely to X," especially given a Humean view. So typically, if I ought not C, then I ought not engage in therapy that will foreseeably strengthen my reasons to C. And there will be exceptions to even this.

It's possible to break out of the hierarchy. At t_1, if the set of permissible options has been changed, then – even if the process that begun the change at t_0 was impermissible at t_0 – the change stands. Deliberating to create these reasons was impermissible, but we are where we are, and so acting upon them at t_1 is not. If all reasons are grounded in desires, then there is no hard boundary to the playground.

Of course, not only deliberation can change our desires. The actions I take now may well determine my reasons. But since the process is naturalistic and often unpredictable, we may reject the following argument due to Raz:

> [T]here is no reason for incommensurabilities among the options open to the agents, for when push comes to shove, the need to choose will concentrate the minds of the choosers, who will realize (or think that they do) that they want one of the options more than the others.
> (Raz 2002: 49)

As a claim about the phenomenology of deliberation, it seems false to me. Perhaps there are biases that reconstruct desires, but I often seem to

plump without forming desires. Retrospectively, I can see that it would have been quite consistent with my wants to have gone the other way – but I had to choose. In principle, this is little different choice under the weak dominance of Buridan's Ass.

The more general point of this section has been that once we admit that desires can provide reasons – and especially if *only* desires can provide reasons – and that there can be indeterminacy therein, a kind of unpredictable naturalistic non-hierarchical voluntarism becomes an option.

5 Conclusion

Both of the authors I've considered draw conclusions about agency from incommensurability. Raz argues that agents can't create normativity, but the will can choose between incommensurate options, meaning a somewhat beefed-up role for the will, which does more than simply execute the verdicts of reason. Chang, by contrast, argues that agents can create normativity but only within the space of rational freedom marked out by the given reasons. Here the faculty of reason and deliberation is beefed up to cope with cases where that space is non-trivial.

I have argued, *pace* Raz and Chang, that incommensurability and related phenomena need not inspire radical views about agency. According to the two indeterminist views I've considered, incommensurability brings many quirks and oddities but does not fundamentally affect the nature of deliberation.

Notes

1 For incisive written comments, I am indebted to Henrik Andersson, John Broome, Ruth Chang, Anders Herlitz, Brad Hooker, David Hull, and Philip Stratton-Lake. I'm also grateful to participants at the 2019 "Incommensurability: Vagueness, Parity and other Non-Conventional Comparative Relations" conference in Stockholm, an audience at Oakland University, and participants in the Reading philosophy Graduate Class.
2 If the vagueness view I discuss next is correct, then "incommensurate" may be a sensible usage since at least some of the time, incommensurability involves its being vague about how options are to be compared on the same scale – how they are to be measured against each other.
3 I'm grateful to John Broome and Brad Hooker for forcing me to be clearer here. Hooker points out that supererogation may complicate claims such as "we ought to choose the best option." But the arguments to follow can focus only on cases where that claim is uncomplicatedly true.
4 I'm grateful to Brad Hooker and Philip Stratton-Lake for discussion of this point.
5 I'm grateful to Ruth Chang for discussion on this point.
6 I'm grateful to Anders Herlitz for suggesting this connection.
7 Unless we are at the edge of the penumbra, but that possibility is complicated by second-order vagueness, where "borderline" itself has borderline cases.

References

Andersson, Henrik. (2017). *How It All Relates: Exploring the Space of Value Comparisons*. Lund University.

Arpaly, Nomy. (2002). *Unprincipled Virtue*. Oxford University Press doi: 10.1093/0195152042.001.0001

Broome, John. (1997). 'Is Incommensurability Vagueness?', in Ruth Chang (Ed.), *Incommensurability, incomparability, and practical reason* (67–89). Harvard University Press.

Chang, Ruth. (1997). 'Introduction', in Ruth Chang (Ed.), *Incommensurability, Incomparability, and Practical Reason* (1–34). Harvard.

Chang, Ruth. (2009). 'Voluntarist Reasons and the Sources of Normativity', in David Sobel & Steven Wall (Eds.), *Reasons for Action* (243–271). Cambridge University Press.

Chang, Ruth. (2013). 'Grounding Practical Normativity: Going Hybrid'. *Philosophical Studies*, *164*(1), 163–187. doi: 10.1007/s11098-013-0092-z

Chang, Ruth. (2017). 'Hard Choices'. *Journal of the American Philosophical Association*, *3*(1), 1–21. doi: 10.1017/apa.2017.7

Dancy, Jonathan. (2006). 'Enticing Reasons', in R Jay Wallace, Philip Pettit, Samuel Scheffler, & Michael Smith (Eds.), *Reason and Value: Themes from the Moral Philosophy of Joseph Raz* (91–118). Oxford University Press.

Elson, Luke. (2017). 'Incommensurability as Vagueness: A Burden-Shifting Argument'. *Theoria*, *83*(4), 341–363. doi: 10.1111/theo.12129

Paul, L. A. (2014). *Transformative Experience*. doi: 10.1093/acprof: oso/9780198717959.001.0001

Raz, Joseph. (1986). Value Incommensurability: Some Preliminaries. *Proceedings of the Aristotelian Society, New Series*, *86*, 117–134.

Raz, Joseph. (2002). 'Incommensurability and Agency', in *Engaging Reason*. doi: 10.1093/0199248001.001.0001

Regan, Donald. (1997). 'Value, Comparability, and Choice', in *Incommensurability, Incomparability and Practical Reason*. (129–150). Harvard University Press.

Rinard, Susanna. (2015). 'A Decision Theory for Imprecise Probabilities'. *Philosophers' Imprint*, *15*(7), 1–16.

Schroeder, Mark. (2007). *Slaves of the Passions*. Oxford University Press. doi: 10.1093/acprof:oso/9780199299508.001.0001

Williams, Bernard. (1981). 'Internal and External Reasons', in *Moral Luck* (101–113). Cambridge University Press.

Part IV

Incommensurability, Risk, and Uncertainty

10 Incommensurability Meets Risk

Wlodek Rabinowicz

The problem I discuss in this chapter concerns interaction between incommensurability in value and risk.[1] More specifically, the chapter focuses on value comparisons between risky actions whose outcomes are bound to be mutually incommensurable whatever state the world might be in. It is seemingly obvious that, in such circumstances, the actions themselves must be incommensurable in their value. Relatedly, it is seemingly obvious that one action cannot be better than another if its outcome would not be better than that of the other action, whatever state the world might be in. But these intuitions, as we shall see, can well be challenged. Indeed, they should be challenged, as there are cases in which they lead us astray.

The problem, in its main outline, was identified by Caspar Hare (cf. Hare 2010; see also Hare 2009, 2013: 45–57). It was subsequently discussed by Temkin (2012), Schoenfield (2014), and Bales et al. (2014).[2] While Hare views it as a problem concerning rational preferences and rational choice between risky actions, I discuss it as a challenge for formal axiology, more specifically for a formal account of value relations. The general question is how axiology should deal with situations in which value incommensurability interacts with risk. This is also Temkin's perspective on the problem. Schoenfield and Bales et al. combine these two perspectives, i.e., the perspective of axiology and that of rational choice. They present it as a problem for rational choice guided by value comparisons. All these authors' contributions will be briefly commented on in what follows, though primarily in footnotes.

In the main text, I am going to focus on the problem itself, as I view it, and on how I think it should be solved. I will describe the problem in Section 1. In Section 2, I will present the solution I oppose, and then, in Section 3, I will present my own solution. In Section 4, I will defend my solution against a prima facie objection, and then, in Sections 5–7, I will provide an extended argument against the solution I oppose and in support of the one I favor. This argument will be based on a formal account of value relations inspired by the Fitting-Attitudes Analysis of value. In Section 8, I will identify a limitation in the scope of my argument, a

limitation that leads to a residual issue I address in Section 9. This issue is one I do not know how to resolve.[3]

1 The Problem

To set up the problem in its general form, consider four possible outcomes, a, b, a^+, and b^+, such that a^+ is better than a but not better than b, while b^+ is better than b but not better than a.

   By the transitivity properties of betterness and equal goodness, it follows from the aforementioned that no other betterness relationships obtain between these four outcomes, nor are any of them equally good. Thus, a and a^+ are incommensurable both with b and with b^+, where by incommensurability I mean the following value relation:

> Two items are *incommensurable* iff neither item is better than the other nor are they equally good.

To illustrate, suppose that a and b are two attractive holiday options: two holiday trips to exciting, but very different, destinations. Say, a is a trip to the Galápagos Islands, while b is a trip to Peru; a^+ and b^+ are the same trips at a small discount (or simply two slightly improved trips to the same locations as a and b, respectively). The destinations are different enough to make a and b incommensurable in value. The discounts (slight improvements) make a^+ better than a and b^+ better than b, but they are not significant enough to make a^+ better than b or b^+ better than a.[4]

   Let *S1* and *S2* be two equiprobable states of the world (two "states of nature"), exactly one of which is actually going to obtain. These states determine the outcomes of the available actions. We assume that the states per se are axiologically neutral (i.e., they are devoid of any value, either positive or negative) and, furthermore, that they do not contribute any value, positive or negative, to the outcomes they determine. Suppose, for example, that a fair coin is to be tossed and that the states are the two possible results of the toss: *S1* is heads and *S2* is tails. We assume that heads and tails both have the same objective probability of one-half.

   The two states form a partition: they are mutually incompatible and jointly exhaustive. I will interpret actions in Savage's manner, as functions from states to outcomes. These are the outcomes they would cause, in different states. This way of understanding actions seems unproblematic as long as the states are causally and probabilistically independent of the actions, which I am going to assume in what follows. Indeed, I am going to assume throughout that we only consider state partitions in which the states are action-independent and are devoid of any value, whether intrinsic or contributive.

   Now, consider two actions, *X* and *Y*, and their outcomes in states *S1* and *S2*, respectively:

	S1	*S2*
X	a^+	b^+
Y	b	a

Thus, action *X* yields a^+ in *S1* and b^+ in *S2*, while *Y* yields *b* in *S1* and *a* in *S2*.[5]

The problem to be discussed in this chapter is simple to pose: is *X* better than *Y*? It might well seem so since both outcomes of *X* are improvements on the outcomes of *Y*.[6] But it might also be argued that *X* is *not* better than *Y*, since no outcome of *X* is better than the outcome *Y* would have in the same state. Indeed, in each state, the outcomes of *X* and *Y* are incommensurable.[7] I will now consider the arguments on both sides in more detail: in the next section, I will focus on the view that *X* is not better than *Y*, and in the section after that, the opposing view will be considered.

In what follows, I shall refer to the issue at hand as *Hare's Problem*, in honor of its originator.[8] In Hare's Problem, in the form in which I pose it, incommensurability meets risk, but solving this problem does not, of course, give us anything like a complete account of the interactions between risk and incommensurability. Attempting to offer such a complete account goes far beyond the goals I have set myself in this chapter. Still, solving Hare's Problem should give us at least some opening to a more comprehensive theory.

2 *X* Is not Better than *Y*

Assuming that states that form a given partition are causally and probabilistically independent of actions, the following principle is obviously correct:

> **(Statewise) Dominance**: (i) An action is better than another action if, in every state, its outcome is better than that of the other action. (ii) An action is at least as good as another action if, in every state, its outcome is at least as good as that of the other action.[9]

Now, consider another principle that seems close in spirit to Dominance:

> **Complementary (Statewise) Dominance**: An action is *not* better than another action if, in every state, its outcome is *not* better than that of the other action.

Or, to put it more succinctly, an action is not better than another if there is no state in which it has a better outcome.[10,11]

While the first clause of Dominance links better outcomes to better actions, Complementary Dominance postulates the corresponding link

between outcomes and actions as regards the *complement* of betterness: the relation of not being better. If the outcomes are not better, in any state, then the action is not better.

There is also an important connection between Complementary Dominance and the second clause of Dominance: in the absence of incommensurabilities, "not better" is equivalent to "at most as good" – the converse of "at least as good." Consequently, in the absence of incommensurabilities, Complementary Dominance would collapse to the second clause of Dominance, which can also be expressed as follows: an action is at most as good as another action if, in every state, its outcome is at most as good as that of the other action. But if incommensurabilities are allowed, as in the case presented here, Complementary Dominance represents an independent constraint.

In Hare's Problem, Complementary Dominance entails that *X is not better than Y*, since there is no state in which the outcome of *X* is better than that of *Y*.[12]

That *X* is not better than *Y* would also follow if we were to accept a principle that links incommensurability of outcomes to incommensurability of actions:

> **Incommensurability**: Two actions are incommensurable if their outcomes are incommensurable in every state.

While Incommensurability does not entail Complementary Dominance, nor is it entailed by it, the two principles are closely related in spirit. Whoever accepts one of these principles should be willing, I think, to accept the other.

It might seem that principles such as Complementary Dominance and Incommensurability are irresistible, as irresistible as Dominance. Here is how an argument for these principles might proceed.

Consider the value of an action conditional on (the obtaining of) a particular state. Arguably, this conditional value can be reduced to the value of that action's outcome in the state in question. This suggests that the same should hold for conditional value relations:

(i) The value relation between two actions that is conditional on a particular state reduces to the value relation between the outcomes of these actions in that state.

But it would seem that

(ii) If two actions stand in the same value relation conditional on each state, then this value relation must also hold between them unconditionally.

Premises (i) and (ii) imply that

(iii) If the outcomes of two actions stand in the same value relation in each state, then this value relation must also hold between the two actions unconditionally.[13]

If this argument is correct, then principles such as Complementary Dominance and Incommensurability must be valid, as they are special cases of (iii), along with Dominance.

But the argument is not correct, at least not as it stands. Its conclusion (iii) is not generally true for all value relations. To see this, consider a value relation such as "not equally as good." It might well be that two actions do not stand in this relation even though their outcomes in each state do. For example, suppose you can bet on heads or on tails and gain $1 if you win. The two bets are equally good assuming that heads are equally as probable as tails. But in every state, whether it is heads or tails, the outcomes of the bets are not equally good: one bet is won and the other is lost.

One might object that the two actions in this betting example are *not* equally good: one of them is better than the other, but which it is depends on the state that actually obtains. If the coin is going to fall heads up, then betting on heads is better than betting on tails, and if it is going to fall tails up, then betting on tails is better than betting on heads. This objection assumes that the value of an action depends on the actual state of the world.[14] It is certainly one way of looking at an action's value, but it is not the only way. When we consider an action ex ante, before learning the state the world is in, and think of it as a potential object of choice, its outcomes in all positively probable states should play a role in determining its value – or so I would argue. However, arguing for this ex ante approach to the evaluation of actions and, more generally, for an ex ante approach to the evaluation of risky prospects would take us too far afield in this chapter. Instead, here I take the ex ante perspective for granted. The question I discuss is therefore the following: what is the right solution to Hare's Problem given the ex ante approach to evaluation?[15]

If (iii) is not generally true, then premise (i) or premise (ii) in the previous argument must be rejected. I am inclined to accept (i) but reject (ii). Given (i), it follows that conditionally on each state the two actions in the betting example are not equally good, because their outcomes are not equally good. But still, unconditionally, the actions *are* equally good, which provides a counterexample to premise (ii).

Since the previous argument, which issues in (iii), is incorrect and its conclusion is not generally true, Complementary Dominance and Incommensurability would need to be defended in some other way. As long as these principles are in doubt, we cannot conclude that, in Hare's Problem, action *X* is not better than *Y*.

3 *X* Is Better than *Y*

That *X is* better than *Y* is the position I accept myself. I will argue for it in this section and defend this argument further in the section that follows. I also need to explain what is wrong with Complementary Dominance and will do so in Sections 6 and 7 after some necessary preparations in Section 5.

Here, then, is how an argument for *X* being better than *Y* can be set up. I will first do so using very strong assumptions and then show how these assumptions can be successively weakened.

If states of nature are assigned objective probabilities, we can compare actions in terms of the *lotteries* on outcomes that the actions give rise to. The lottery l_x generated by an action *x* is a probability distribution over outcomes *o* such that for at least one state, *S*, in the state partition, $x(S) = o$. For each such outcome *o*, $l_x(o)$ is the sum of the probabilities of the states in which *x* yields this outcome:

$$l_x(o) = \Sigma_S P(S)(x(S) = o),$$

where *S* ranges over the states in a given partition.
As is easy to see,

$$l_X = (a^+, .5; b^+, .5), \text{ while } l_Y = (b, .5; a, .5).$$

In this notation, we describe a lottery by listing its possible outcomes, with each outcome being followed by its probability. The ordering in which the outcomes are listed does not matter, so, for example, (b, .5; a, .5) and (a, .5; b, .5) describe the same lottery.

The von Neumann–Morgenstern expected utility theory assumes that preferences for lotteries satisfy the axiom of Independence: replacing an outcome in a lottery with another outcome that is a more preferred results in a more preferred lottery. What follows is the corresponding principle for value comparisons:

> **Independence**: For all outcomes *o*, *o′*, *o″* and all probabilities $p > 0$, $(o, p; o″, 1-p)$ is better than $(o′, p; o″, 1-p)$ iff *o* is better than *o′*.

According to this principle, different lottery prizes (outcomes) make independent value contributions to the value of a lottery: improving a prize always improves the lottery.

Since a^+ is better than *a* and b^+ is better than *b*, Independence implies that $(a^+, .5; b^+, .5)$ is a better lottery than $(a^+, .5; b, .5)$ and that the latter lottery is better than $(a, .5; b, .5)$. Since betterness is transitive, it follows that $l_X = (a^+, .5; b^+, .5)$ is better than $l_Y = (a, .5; b, .5)$.

Suppose we now assume:

> **Reduction of Actions to Lotteries**: For all actions x and y, x is better than (equally as good as) y iff l_x is better than (equally as good as) l_y.

According to this principle, the value of an action is determined by the value of the associated lottery.

It follows from Independence and Reduction of Actions to Lotteries that action X is better than action Y, contrary to what is implied by Complementary Dominance.

An argument for X being better than Y can be constructed in another way as well, without relying on such strong principles as Reduction of Actions to Lotteries and Independence. (Independence, in particular, is highly problematic. As is well-known, its preferential version has counterintuitive implications, as exemplified by the famous Allais' Problem.) Instead, we could appeal to Dominance, which we should accept in any case, together with the following weakening of Reduction of Actions to Lotteries:

> **Weak Reduction of Actions to Lotteries**: If two actions give rise to the same lottery, they are equally good.[16]

Now, consider adding to X and Y a third action, Z:

	S1	*S2*
X	a^+	b^+
Y	b	a
Z	a	b

Y and Z give rise to the same lottery, $(a, .5; b, .5)$. (Remember that the ordering of the outcomes in the description of a lottery does not matter.) Thus, Y and Z are equally good by the Weak Reduction of Actions to Lotteries. But by (the first clause of) Dominance, X is better than Z since, in every state, X's outcome is better than Z's. Since betterness is transitive across equal goodness, this means that X must also be better than Y.

Indeed, the detour through lotteries is not needed to establish the result we are after. Instead, we can base our argument on a very simple principle that is even less demanding than Weak Reduction of Actions to Lotteries.

> **Switch**: If for some equiprobable states S and S', an action x can be obtained from an action y just by switching outcomes between S and S', then actions x and y are equally good.[17]

In other words, a mere switch of outcomes between equiprobable states does not affect the value of an action.[18]

Now, it is easy to see that action *Z* results from *Y* by such a switch of outcomes between equiprobable states *S1* and *S2*. Thus, Switch implies that *Y* and *Z* are equally good and, therefore, again by (the first clause of) Dominance and the transitivity of betterness across equal goodness, *X* is better than *Y*, just as *X* is better than *Z*.[19] I find this argument from Switch both simple and compelling.[20]

Bales et al. (2014) pose an objection to Hare's (2010) dominance-based argument for the impermissibility of *Y*, an argument in which Hare also introduces a third action *Z* that is obtained from *Y* by switching outcomes.[21] Their objection can be adapted to my argument as well, as follows: action *Z* would still be dominated by *X* even if states had unequal probabilities. However, if the states' probabilities were unequal, we would no longer be able to use Switch (or Weak Reduction of Actions to Lotteries) to prove that *Y* and *Z* are equally good. We would thus be unable to establish that *X* is better than *Y*.[22] This seems unsatisfactory.

I do not find this objection very worrying. That the argument does not extend to cases in which the state probabilities are unequal is as it should be. If the difference in state probabilities were non-negligible, then it would be reasonable to judge that *X* and *Y* *are* incommensurable. Indeed, the modeling of value relations that I am going to present next (beginning in Section 5) bears this out. On the other hand, it might be reasonable to expect that *X* would still be better than *Y* if the difference in state probabilities were very small. The argument I have presented cannot be used to prove it, but this limitation in scope does not make the argument less compelling.

I do recognize, however, that friends of Complementary Dominance will not be swayed by the argument but will reject one of its premises. Since they are committed to Dominance and also, I take it, to the transitivity of betterness across equal goodness, they will have to reject Switch. Indeed, this move is perfectly natural for a friend of Complementary Dominance. As noted earlier, accepting that principle goes hand in hand with accepting Incommensurability, but Incommensurability is incompatible with Switch. Since outcomes *a* and *b* are incommensurable, actions *Y* and *Z* lead to incommensurable outcomes in every state. Therefore, Incommensurability implies that that *Y* and *Z* themselves are incommensurable and not equally good, as Switch would have it. In other words, from this perspective, Switch must be rejected. (The same, of course, goes for the stronger principles, such as the Reduction, or Weak Reduction, of Actions to Lotteries.) Thus, there is still work to do as long as we haven't yet provided a *direct* argument against Complementary Dominance and Incommensurability. I will attempt to do it in Sections 6 and 7; then I will also provide reasons for accepting Switch.

4 The Same Outcomes?

First, however, we need to consider and put to rest another potential objection to the argument from Switch. The application of Switch to the case at hand assumes that the outcome of *Y* in *S1* is the same as that of *Z* in *S2*, and – analogously – that the outcome of *Y* in *S2* is the same as that of *Z* in *S1*. This might be questioned.

One way to question it is by arguing that an outcome includes the state in which it occurs. Thus, a more complete description of, say, the outcome of *Y* in *S1* is '*b* in state *S1*', while a more complete description of the outcome of *Z* in *S2* is '*b* in state *S2*'. Conceived in this way, these two outcomes are not identical. Thus, it is not true that *Z* is obtained from *Y* by a mere outcome switch between equiprobable states. (For the same reason, the lotteries that *Y* and *Z* give rise to are not identical either.)

But this non-identity objection is ungrounded. It is perfectly reasonable to exclude from the outcomes those parts that are axiologically irrelevant. And we have assumed that the states in Hare's Problem are irrelevant in this way. While they determine the outcomes of the actions, they do not contribute any value, positive or negative, to the outcomes under consideration. Consequently, there is no reason to incorporate such states into outcomes.

A more challenging version of the non-identity objection distinguishes between seemingly identical outcomes by taking into account their *modal* features. In particular, if we evaluate the outcome of an action, one relevant consideration might be *what the action could have brought about instead if the state of the world had been different*. Think of a bet in which you lose a sum of money. For the evaluation of this outcome, it might well play a role whether the bet could have resulted in a large win. This modal fact may be relevant (positively or negatively) to the disvalue of your loss, so it may be reasonable to require that it should be included in the outcome specification.

This kind of appeal to what might have been does not, however, distinguish between the outcomes in the problem at hand. Thus, consider *b*, which is the outcome of action *Y* in *S1*. In the other, equally probable state, *Y* would have resulted in *a*. The situation is exactly similar concerning *b* as the outcome of action *Z* in state *S2*: in the other, equally probable state, *Z* would have resulted in *a*. Therefore, in this respect, there is no difference between *b* as the outcome of *Y* and *b* as the outcome of *Z*. The same applies, mutatis mutandis, to outcome *a*, as brought about by *Y* or by *Z*.

An appeal to the modal features of outcomes might, however, also take another form: in evaluating an outcome of an action, it is one thing to consider what this action might have brought about if the state had been different. But we could also consider what would or could have

happened in the same state *if another available action had been performed instead.* It might be argued that a loss resulting from a risky action is less bad for the agent if, in a given state, it could not have been prevented by acting differently: under such circumstances, there is no reason to regret what one has done, and this might make the outcome easier to bear.[23]

In the problem we are considering, there is a modal difference of this kind between the otherwise identical outcomes of *Y* and *Z*, a difference that *might* be relevant to their value. Thus, the outcome *b* of *Y* in *S1* is not worse than the outcome in that state of any other action at the agent's disposal. In *S1*, actions *X* and *Z* would have brought about a^+ and *a*, respectively, and neither of these outcomes is better than *b*. By contrast, the outcome *b* of *Z* in *S2 is* worse than the outcome b^+ of *X* in the same state. Because of this modal difference, it is arguable that the outcome *b* of *Z* in *S2* is worse than the outcome *b* of *Y* in *S1*. The same applies to *a*. The outcome *a* of *Z* in *S1* is arguably worse than the outcome *a* of *Y* in *S2*: in *S1*, if the agent had chosen *X* instead of *Z*, she would have brought about a better outcome, a^+. But in *S2*, if the agent had chosen *X* instead of *Y*, she would not have brought about a better outcome, as b^+ is not better than *a*. This forms a basis for the objection that action *Z* is worse than action *Y*: each outcome of *Z* is worse than the seemingly identical outcome of *Y*. Consequently, even though, by Dominance, *X* is better than *Z*, it does not follow that *X* is better than *Y*.

The objection just presented is controversial since it is debatable whether the modal features of outcomes should be seen as relevant to their value. However, there is a way to finesse the objection instead of meeting it head on: we can simply change the example by introducing a fourth action, *U*, that is related to *Z* in the same way as *X* is related to *Y*:

	S1	S2
X	a^+	b^+
Y	b	a
Z	a	b
U	b^+	a^+

By introducing *U*, we make *Y* and *Z* perfectly symmetrical in their modal properties: we make their corresponding outcomes modally identical. Just as for *Z* there is another action, *X*, that in every state has a better outcome, so it is for *Y* in our modified example: there is another action, *U*, that in every state has a better outcome. Thus, *a* as the outcome of *Y* and *a* as the outcome of *Z* now have the same modal features, as do *b* as the outcome of *Y* and *b* as the outcome of *Z*.

Thus, in the presence of this fourth action, modal considerations no longer pose any obstacle to applying Switch to *Y* and *Z* and concluding

that these two actions are equally good. In combination with Dominance and the transitivity of betterness across equal goodness, we can then draw the conclusion that *X* is better than *Y*.

Someone might object at this point: your argument from Switch shows that *X* is better than *Y* when *Z* is available, and you have now improved the argument by bringing in the fourth option, *U*. But how is this supposed to show that *X* is better than *Y* when no other alternatives are on offer, as is the case in Hare's original problem?[24] I agree that the argument does not show this if one believes that the value of an action might depend on what the agent could have done instead. Note, however, that Complementary Dominance implies that *X* is *not* better than *Y*, *even* if *Z* and *U are* available – that is, even in the case in which the argument from Switch cannot be faulted on modal grounds. Consequently, this argument at least shows that Complementary Dominance must be rejected if Switch is accepted.

In what follows I will simplify my discussion by assuming that the modal features of actions are *not* relevant to their value. On this assumption, it does not matter for the comparison of *X* with *Y* whether or not they are the only alternatives on offer. The argument from Switch can then be used to draw the conclusion that *X* is better than *Y* quite generally, independently of what other actions might be available. However, Complementary Dominance, which might seem at least as compelling as Switch, implies that this conclusion is false. Therefore, if we want to hold on to the view that *X is* better than *Y*, we need to confront the opposition head on and take a closer look at Complementary Dominance. Is this principle as compelling as it seems to be at first sight?

5 Value Relations Analyzed

To approach this issue, it is helpful to ask first how value relations should be analyzed. In Rabinowicz (2008), I proposed such an analysis in the spirit of the Fitting-Attitudes account of value. According to this account, to be valuable is to be a fitting target of a pro-attitude. An item *i* is valuable iff it is fitting (appropriate, suitable) to have a pro-attitude toward *i*. Or, to put it in more standard normative terms, *i* is valuable iff one ought to have a pro-attitude toward *i*.[25] Different kinds of value correspond to different kinds of fitting pro-attitudes. Items that exhibit "thick" values, that are, say, admirable or delightful, call for "thick" pro-attitudes: admiration, delight. "Thin" values, such as goodness, correspond to "thin" pro-attitudes: approval or appreciation. For thin value relations, such as betterness, the relevant pro-attitude is preference: an item *i* is better than an item *j* iff one ought to prefer *i* to *j*. Analogously, *i* and *j* are equally good iff one ought to equi-prefer *i* and *j*, i.e., be indifferent between them. Consequently, *i* and *j* are incommensurable iff one neither ought to prefer one to the other nor ought to be indifferent. The kind

of preference and indifference that is relevant in this context is meant to be a relatively stable and considered stance. Betterness and equal goodness are thus analyzed in terms of normative requirements regarding considered preferential attitudes. These normative requirements might in some cases leave leeway for different preferences: if divergent preferences regarding i and j are permissible (not unfitting), then these items are incommensurable.

Let **I** be the domain of items that are being compared and suppose that **K** is the non-empty class of all permissible preference orderings of that domain. We assume that in every ordering in **K**, weak preference (i.e., preference-or-indifference) is a reflexive and transitive relation. Preference and indifference are definable in terms of weak preference in the standard way, as, respectively, the asymmetric and symmetric parts of that relation. This implies, for the orderings in **K**, that preference is transitive and asymmetric, while indifference is transitive, symmetric, and reflexive. We can now define different value relations on **I** in terms of **K**. For any two items i and j in **I**,

   i is *better* than j iff in every ordering in **K**, i is preferred to j,

   which is another way of saying that one ought to prefer i to j, and

   i is *equally as good* as j iff in every ordering in **K**, i and j are equi-preferred,

   which is another way of saying that one ought to equi-prefer i and j.

These definitions imply that the relation of betterness-or-equal goodness is transitive and reflexive because weak preference has been assumed to be transitive and reflexive. In its turn, this entails that betterness is transitive and asymmetric, while equal goodness is transitive, symmetric, and reflexive. It also implies that betterness is transitive across equal goodness, i.e., that i is better than j if i is better than k and k and j are equally good. We were assuming this property of betterness in our argument from Switch in Section 3: X is better than Y if X is better than Z and Z and Y are equally good.

Given this modeling of value relations,

   i and j are *incommensurable* iff
   (i)   there are orderings in **K** in which i is not preferred to j,
   (ii) there are orderings in **K** in which j is not preferred to i, and
   (iii) there are orderings in **K** in which i and j are not equi-preferred.

   Clauses (i)–(iii) amount, respectively, to three conditions: i is not better than j, j is not better than i, and i and j are not equally good.

Note that these necessary and jointly sufficient conditions of incommensurability are satisfied if (though not only if) the following holds:

> **K** contains two preference orderings such that, in one, i is preferred to j, and in the other, j is preferred to i.

In Rabinowicz (2008), I suggested that this condition can be used to define a very common form of incommensurability that Ruth Chang calls "parity" (cf. Chang 2002). Accordingly,

> i and j are *on a par* iff
> it is permissible to prefer i to j and likewise permissible to prefer j to i[26]

Cases like this may arise when the comparison between two items is based on weighing several relevant dimensions, or aspects, of comparison against each other. One item might rate higher than the other on some of the dimensions and lower on others. Different assignments of weights to dimensions are often admissible in such cases, giving rise to different permissible all-things-considered preference orderings of items. **K** is meant to be the class of all such permissible all-things-considered orderings. If item i comes higher than j given some admissible assignments of weights to the dimensions of comparison, while j comes higher than i given some other admissible weight assignments, then **K** will contain preference orderings in which i is ranked above j and other preference orderings in which j is ranked above i. We will thus have a case of parity.

6 Deconstructing Complementary Dominance: The Argument from Ordinal Distances

Consider what this analysis of value relations in terms of permissible preference orderings implies for the problem at hand. We let the item domain **I** consist of outcomes and actions[27] and we take **K**, as before, to be the class of permissible preference orderings of **I**. If Dominance and Switch are to be valid, then their preferential variants must hold for *every ordering* in **K**:

> **Dominance (P):** (i) An action is preferred to another action if, in every state, its outcome is preferred to that of the other action. (ii) An action is weakly preferred to another action if, in every state, its outcome is weakly preferred to that of the other action.[28]

> **Switch (P):** If an action x can be obtained from an action y just by switching outcomes between two equiprobable states, then x and y are equi-preferred.[29]

Here and in what follows, "(P)" after the name of a condition stands for *preference*, to distinguish such conditions on permissible preferences from the analogous principles for value relations.

Now, what about Complementary Dominance? For this principle to be valid, **K** would need to satisfy the following rather clumsy condition on permissible preferences:

> (*) If for every state *S* there is some ordering in **K** in which the outcome of action *x* in *S* is not preferred to the outcome of action *y* in *S*, then in some ordering in **K**, *x* is not preferred to *y*.[30]

Note that (*), unlike Dominance (P) or Switch (P), is not a constraint on each preference ordering in **K** considered separately, but instead a condition on class **K** taken as a whole: it requires that **K** contain an ordering with a certain property if **K** contains orderings exemplifying certain other properties.

While (*) is both necessary and sufficient for Complementary Dominance, it lacks immediate appeal and, indeed, it seems to lack intuitive support. To see this, consider, as an illustration, a simple model that represents the setup of Hare's Problem. There are just four outcomes in **I** in this model: a^+, which is better than *a*, and *b+*, which is better than *b*, but a^+ is not better than *b*, nor is b^+ better than *a*. Suppose that class **K** consists of three preference orderings, *P1*, *P2*, and *P3*, that order the four outcomes as follows (in each ordering the higher-ranked outcomes appear higher up):

P1	*P2*	*P3*
a^+	b^+	a^+, b^+
a	*b*	*a, b*
b^+	a^+	
b	*a*	

While *P1* and *P2* order the four outcomes linearly, *P3* contains two ties: a^+ and b^+ are equi-preferred in *P3* and so are *a* and *b*.

Given our specification of **K**, it follows that a^+ is better than *a* (as it is preferred to *a* in every ordering in **K**), b^+ is better than *b* (for the corresponding reason), but a^+ is not better than *b* (*b* is preferred to a^+ in *P2*), nor is b^+ better than *a* (*a* is preferred to b^+ in *P1*), just as we wanted.

P3 is not really needed for our purposes: if we removed it from **K**, the model would equally well represent the assumed betterness relationships between outcomes. But adding *P3* does not hurt. Indeed, instead of making **K** smaller, we could make it larger: we could add further orderings to **K**, as long as in all of them a^+ is preferred to *a* and b^+ is preferred to *b*. We assume for the time being, however, that all the orderings in **K** are

complete, i.e., contain no preferential gaps. I am going to consider incomplete orderings in the next section.

The orderings in **K** extend to all the items in the domain **I**. Since we have assumed that **I** also contains actions, we need to say something about how *P1–P3* order actions *X*, *Y*, and *Z*. We need to do this in order to assess Complementary Dominance. How compelling is this principle, or – which amounts to the same thing in our analysis – how compelling is its preferential version (*)?

The answer is that it is not compelling at all. The model described earlier can serve as an illustration. The antecedent of Complementary Dominance holds for *X* and *Y* in this model: there is no state in which *X* has a better outcome than *Y*. Or, what amounts to the same thing, the antecedent of (*) is satisfied by *X* and *Y*: for every state, there is a preference ordering in **K** in which the outcome of *X* in that state is not preferred to the outcome of *Y* (for state *S1*, *P2* has this feature since it ranks *a⁺* below *b*, while *P1* has this feature for *S2* since it ranks *b⁺* below *a*). What about the consequent of (*) then? Is it satisfied by *X* and *Y*? Is there an ordering in **K** in which *X* is not ranked above *Y*?

Note that in every preference ordering in **K**, at least one of the possible outcomes of *X*, if not both, is preferred to the outcome of *Y* in the same state. In *P1* this applies to *a⁺*, the outcome of *X* in *S1*, in *P2* it applies to *b⁺*, the outcome of *X* in *S2*, while in *P3* it applies to the outcomes of *X* in both states. Thus, in each of these permissible orderings, the preference for *X* over *Y* is supported by the preference for at least one of *X*'s possible outcomes over the corresponding outcome for *Y*.

Admittedly, in orderings *P1* and *P2* (though not in *P3*), the opposite preference, for *Y* over *X*, is also supported by the preference for one of the outcomes of *Y* over the corresponding outcome of *X*. But, as I will now argue, this support for *Y* is *weaker* than the support for *X*. To see this, first consider *P1*:

$$
\begin{array}{|c|}
\hline
P1 \\
\hline
a^{+} \\
a \\
b^{+} \\
b \\
\hline
\end{array}
$$
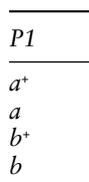
In this ordering, the preference for *X* over *Y* is supported by the preference for *X*'s outcome over *Y*'s outcome in *S1*, i.e., in *P1*, *a⁺* is preferred to *b*. The preference for *Y* over *X* is supported in *P1* by the preference for *Y*'s outcome over the outcome of *X* in *S2*, i.e., *a* is preferred to *b⁺*. But it is easy to see that this support for the preference for *Y* is weaker – less pronounced – than the support for the preference for *X*. In a very intuitive sense, the ordinal distance between *a⁺* and *b* in *P1* is greater than the ordinal distance between *a* and *b⁺*. The former items occupy in this

ordering the first and the fourth place, respectively, while the latter two items occupy, respectively, the second and the third place: they are thus closer to each other.

While intuitive, this idea of being closer in an ordering can be made more precise. An assignment of numbers to items *represents* an ordering of the items iff higher numbers are given to items that are placed higher up in this ordering.[31] Now, clearly, in every assignment that represents *P1*, the absolute difference between the numbers assigned to a^+ and b must be greater than the absolute difference between the numbers assigned to a and b^+, simply because of their placement in the ordering. In this sense, then, the *ordinal distance* between a^+ and b in *P1* is greater than that between a and b^+.

To put it more generally, for any complete ordering *P*, the ordinal distance between items i and j is greater than (equal to) the ordinal distance between items k and l iff for every assignment v of numerical values to the items in *P*'s domain, if v represents *P*, then the absolute difference between $v(i)$ and $v(j)$ is greater than (equal to) the absolute difference between $v(k)$ and $v(l)$. To put it differently, when we talk about ordinal distances, we focus only on those distance comparisons that are *invariant* under all monotonic (order-preserving) transformations of the measurement scale.

Obviously, comparisons of ordinal distance are extremely gappy. For this reason, the notion of ordinal distance has very limited use. Thus, for example, the ordinal distance between items that occupy, say, the first and second places in an ordering is neither greater nor smaller than that between items that come in the second and fourth places. Nor are these distances equal; instead, they are incommensurable. As we have seen, however, some ordinal distances *are* greater or smaller than others. These differences in ordinal distances give rise to the differences in support. The support *P1* offers to *X* in *S1* is stronger than the support it offers to *Y* in *S2*, because, in *P1*, the ordinal distance between a^+ and b is greater than that between a and b^+.

The argument with respect to *P2* is exactly analogous.

$$\begin{array}{c} \hline \textit{P2} \\ \hline b^+ \\ b \\ a^+ \\ a \\ \hline \end{array}$$

In *P2*, just as in *P1*, the support that the preference for *Y* receives in one state (*S1*) is weaker, in terms of the ordinal distance, than the support that the preference for *X* receives in the other state (*S2*). But then, since the states are equiprobable, it should not come as a surprise that *X* is preferred to *Y* in both *P1* and *P2*, and indeed in every ordering in **K**.

(In *P3*, the preference for *Y* receives no support at all; the preference for *X* over *Y* is supported in both states.) This means that the consequent of (*) is false, or – which amounts to the same thing – that Complementary Dominance fails to hold in this model.

Indeed, we can establish a stronger conclusion. The argument from ordinal distances generalizes: in *any* complete preference ordering in which (i) a^+ is ranked above a and (ii) b^+ is ranked above b, the support for the preference for *X* over *Y* is stronger, in terms of ordinal distance, than the support for the preference for *Y* over *X*.[32] Unless (i) and (ii) hold, the ordering is not permissible given our assumptions about the betterness relationships between these four outcomes. Therefore, in every permissible complete ordering, *X* is preferred to *Y*. If all preference orderings in **K** are complete, it follows that in all of them, *X* is ranked above *Y*. Which means, on our analysis of value relations, that *X* is better than *Y*. Thus, Complementary Dominance fails in all models of Hare's Problem in which **K** is a set of complete preference orderings. In all of them, *X* is better than *Y*. Note, by the way, that this argument also implies that Incommensurability fails in the same way. While the outcomes of actions *X* and *Y* are incommensurable in every state, these actions themselves *are* commensurable: *X* is better than *Y*.

A similar argument from ordinal distances can be used to defend Switch. As we know, this condition is equivalent to Switch (P). If an action is obtained from another action just by switching outcomes between equi-probable states, then, according to Switch (P), the two actions must be equi-preferred in every ordering in **K**. To see how we can argue for Switch (P), let us use, as an illustration, the comparison between *Y* and *Z*. Once again, we only consider, for the time being, complete preference orderings. In any such ordering, either (1) outcomes *a* and *b* are tied, and then the ordering will rank *Y* and *Z* equally, or (2) one of these outcomes is ranked above the other: for example, *a* is ranked above *b*. (The opposite ranking can be dealt with in the same way.) Then (i) *Y*'s outcome is preferred to that of *Z* in *S2* and (ii) *Z*'s outcome is preferred to that of *Y* in *S1*. The support that the preference for *Y* over *Z* gets from (i) is exactly equal to the support the preference for *Z* over *Y* gets from (ii); the ordinal distance from *a* to *b* is the same as that from *b* to *a*. Since *S1* and *S2* are equiprobable, it follows that *Y* and *Z* will be equi-preferred in any such permissible ordering, just as in the permissible orderings considered in case (1). It follows, then, that Switch (P) must be satisfied by any permissible complete preference ordering. Therefore, in all models in which **K** only contains complete orderings, Switch will be valid.

Let us return to Hare's Problem. I have argued that in every permissible complete ordering, *X* will be ranked above *Y*. This argument depends on the equi-probability of states. To see this, consider *P1* as an example. If *S2* were more probable than *S1*, the support for the preference for *Y* over *X* in *P1*, while less pronounced in terms of ordinal distance than the

support for the opposite preference, would gain weight: it would come from the preference for the outcome of *Y* in the more probable state (*S2*). Correspondingly, the support for *X* over *Y* in *P1*, while more pronounced, would lose weight: it would come from the preference for the outcome of *X* in the less probable state (*S1*). This could lead to *Y* being ranked above *X* in *P1*. In *P2*, on the other hand, the support for *X* would gain weight and thus would even more easily outbalance the less pronounced support for *Y*. As a result, if the states had unequal probabilities, *X* and *Y* might well be on a par, with each of them being preferred to the other in some permissible preference ordering.

Does this mean, then, that the case in which state probabilities are equal is exceptional, a kind of singularity? That as soon as they are not equal, *X* and *Y* become incommensurable? This suggestion would go too far. After all, there might be cases in which all permissible preference orderings are based on cardinal representations of outcome utilities and on the associated expectational representations of the utilities of actions. In all such orderings, *X* will still be preferred to *Y* even after a slight perturbation of state probabilities: expected utility is continuous in both utilities and probabilities. The argument from ordinal distances presented earlier does not, of course, presuppose the expected utility interpretation of permissible preferences. The comparison of ordinal distances on which this argument relies is much less exacting than the interval measurement of preference strength needed for determining expected utilities. Still, the ordinal modeling I rely on does not exclude such interval measurement: it is in principle compatible with the expected utility interpretation of permissible preferences. Thus, even if one state becomes slightly more probable than the other, *X* might well remain a better action than *Y*.

7 Incomplete Preference Orderings

There is a natural line of defense available to the friends of Complementary Dominance and Incommensurability. In the argument from ordinal distances, I have been assuming that all the preference orderings in **K** are complete. But this need not of course be the case. Indeed, in Rabinowicz (2008), I stressed that some orderings in **K** might well contain preference gaps. In Hare's Problem, a permissible preference ordering might be partly incomplete in its ranking of the four outcomes a^+, a, b^+, and b, as long as it ranks a^+ above a and b^+ above b. It might contain gaps when it comes to all the other comparisons between these four outcomes. In such an incomplete ordering, none of the possible outcomes of *X* is preferred to the outcome that *Y* would have in the same state. Consequently, the preference for *X* over *Y* is not supported in this ordering by preferences for outcomes.[33] A critic of my argument could therefore insist that this permissible incomplete ordering is not going to rank *X* above *Y*: instead, it will contain a preference gap with respect to these two

actions. The presence of such an ordering in **K** will therefore imply that X is incommensurable with Y. Thus, Complementary Dominance and Incommensurability will be rehabilitated.

This is certainly a weak point in my argument. But can I respond to the critic? I think so. To begin with, gaps in preference orderings might be permissible, but they do not *have* to be permissible. In a model in which preferential gaps among the outcomes of X and Y are not permissible, the argument from the preceding section shows that Complementary Dominance and Incommensurability are violated. Second, even in models with permissible preferential gaps, it might well be possible to fill in the gaps in permissible ways. To illustrate, I might, permissibly, have no preference regarding the two holiday trips, to the Galápagos Islands and to Peru, but it certainly is permissible for me to form a definite preference for one trip or the other. Likewise, it is permissible for me to form an equi-preference for these two trips. In other words, a permissible incomplete preference ordering can be *permissibly completed*, in various ways.

Consider where this idea can take us. Let G be a set of items in the domain **I** and P a permissible preference ordering that contains gaps between some of the items in G. Suppose that these gaps can be permissibly filled, i.e., that there is a permissible *completion of P with respect to G*, a permissible ordering P' that agrees with P but is complete over G.[34] If such a permissible completion of P with respect to G exists, then it is reasonable to expect that P is the common part of some set of its permissible completions with respect to G. In other words:

> If there is a permissible completion of P with respect to G, then P is the intersection of some non-empty set of permissible completions of P with respect to G.[35]

If this condition is satisfied by P with respect to every subset G of **I**, we shall say that P is *well-rounded*. Well-roundedness is a very plausible constraint on permissible orderings. It expresses the intuition that, if the gaps in a permissible preference ordering can be permissibly filled in, then this ordering represents the preferences of someone who remains neutral when it comes to different permissible ways of filling the gaps. It contains only what is common to the different completions.[36]

If we accept the constraint of well-roundedness on the orderings in **K** and accept that gaps in permissible orderings can be permissibly filled in, then the argument from ordinal distances can be extended to incomplete orderings in **K**. If a permissible ordering P is gappy within the set $\{a, a^+, b, b^+\}$, but these gaps can be permissibly filled in, then P is the intersection of a set of its permissible completions with respect to this outcome set. But we already know that in all permissible orderings that are complete with respect to $\{a, a^+, b, b^+\}$, action X is preferred to action Y. Since P is the

common part of a set of such orderings, the same must apply to *P* itself, and even in this incomplete ordering, *X* must be preferred to *Y*.

This means that the critic's objection fails. Complementary Dominance and Incommensurability are not satisfied even if incomplete orderings are allowed in **K**.

The argument I have just presented can also be used to show that Switch (P) (and thus also Switch) holds even if **K** contains incomplete preference orderings. If an action *x* can be obtained from an action *y* by a mere switch of outcomes between equiprobable states, then it does not matter that in some permissible orderings *P* there is a preference gap between these outcomes. As long as it is permissible to fill this gap, well-roundedness implies that *P* will be the intersection of complete permissible orderings. We already know that *x* and *y* are ranked equally in all these orderings, so they will also be ranked equally in *P*.

At this point, let me digress. Complementary Dominance has an exact analogue for permissible preferences stating that for all orderings in **K**, the following holds:

> **Complementary Dominance (P):** An action is not preferred to another action if there is no state in which its outcome is preferred to that of the other action.

Unlike (*), this principle is neither necessary nor sufficient for Complementary Dominance.[37] But it might appear to be as compelling as the latter. Are the appearances misleading even in this case? The answer is yes. As seen earlier, in a permissible incomplete ordering according to which there is no state in which the outcome of *X* is ranked above the outcome of *Y*, *X* still must be ranked above *Y*, provided that the gaps in the ordering can be permissibly filled in and the constraint of well-roundedness is satisfied.[38] However, it should be noted that Complementary Dominance (P) is perfectly innocuous if the preference ordering under consideration is complete. Then this condition immediately follows from the second clause of Dominance (P). This is exactly analogous to the observation made in Section 2: if the value ordering is complete (i.e., if there are no incommensurabilities), then Complementary Dominance follows from the second clause of Dominance.

8 Incomparability

Do we now have a satisfactory general solution to Hare's Problem? Is an action such as *X* necessarily better than its competitor *Y*? Well, perhaps not. The kind of case I have so far been considering is one in which potential gaps in a permissible preference ordering of outcomes can be permissibly filled in. I have considered a case in which outcomes *a⁺* and *a* are *on a par* with outcomes *b⁺* and *b*. Even if preferential

gaps regarding these outcomes might be permissible, it is permissible to prefer any of the former outcomes to any of the latter and vice versa. Parity is a form of incommensurability, in fact, the most common one. It is exemplified by such comparisons as the one between holiday trips, or by Hare's own comparison between different restaurant options. Arguably, it also is exemplified by another case he considers, one in which I have to choose which of my possessions should be saved from a fire: a Fabergé egg or an album with my wedding photos (cf. Hare 2010: 237f).[39]

At least in principle, though, there might also exist cases of incommensurability in which it is *not* permissible to prefer one item to the other or to be indifferent.[40] In such cases what is required is a preferential gap. In Rabinowicz (2008), I refer to such a form of incommensurability as *incomparability* in value. Clearly, incomparability should be expected to obtain between items belonging to different ontological categories, say, between persons and events or between abstract entities such as properties or states of affairs and concrete things. But if the compared items belong to the same category, postulating incomparability is considerably more problematic. In particular, if what we compare are different action outcomes, then it is fair to say that incomparability between such items is much rarer than parity; indeed, whether it can ever occur may be disputed.

An example of Hare's Problem that might involve incomparable outcomes is provided by Bales et al. (2014). Their original example is somewhat problematic, but here is what I think is a more convincing variant of this example:[41] outcomes *a* and *b* are, respectively, the very painful death of my mother and the very painful death of my father. In the "+"-versions of these outcomes, my mother/my father suffers slightly less when dying. *a*[+] is slightly better (less bad) than *a* and *b*[+] is slightly better than *b*, but from my perspective, both *a*[+] and *a* are incommensurable with *b*[+] and *b*. It might be argued that this incommensurability is of the extreme kind, that it is a case of incomparability in (personal) value. On this reading, it is impermissible for me to prefer the death of my mother (whether very painful or slightly less so) to the death of my father, and vice versa, nor is it permissible for me to be indifferent. What is required on my part is a preferential gap. Arguably, what characterizes genuine choice dilemmas is such incomparability between the alternatives.[42]

Are such examples as the preceding one convincing cases of incomparability of outcomes? This might be questioned. Perhaps, instead of a preferential gap being required, it might be permissible for me to be indifferent between the painful death of my mother (*a*) and the painful death of my father (*b*). I balk at this thought because I think that I can be indifferent only if I do not care which of them is going to die. But is this what indifference necessarily involves? It is not obvious.[43] If being indifferent between two alternatives is compatible with caring

very much which alternative is going to obtain, then the example under consideration might not be a case of incomparability: it might then be permissible for me to be indifferent rather than remain in a state of a preferential gap.

Note, however, that I should prefer my mother to die slightly less painfully if possible: the preference for a^+ over b is one I surely ought to have. Since preferences in permissible orderings are transitive across indifference,[44] it would follow that I ought to prefer my mother's slightly less painful death (a^+) to the painful death of my father (b) *if* I am indifferent between her very painful death (a) and his very painful death (b). This, it appears to me, is a counterintuitive implication. I ought not to have this preference in any permissible preference ordering: I ought not to prefer her death to his death (or vice versa), even if it would be slightly less painful. So, it does seem that being in a state of a preferential gap is the only permissible stance toward a and b in this case. If that is true, then it *is* a case of incomparability.

If some gaps in a preference ordering are due to incomparabilities, then such gaps cannot be permissibly filled in. But this means that in a case like this the argument from ordinal distances cannot be extended to incomplete preference orderings. This extension assumes that it is permissible to fill in preferential gaps. It is only if there is a permissible completion of a preference ordering with respect to an outcome set $G = \{a, a^+, b, b^+\}$ that this ordering, by the constraint of well-roundedness, can be taken to be the intersection of a set of its permissible completions with respect to G.

9 Residual Paradox: Incomparability Meets Risk

So, where does this leave us? I have provided an argument that explains why Complementary Dominance must be rejected: one action might well be better than another even if there is no state in which it has a better outcome. I have shown how this can be possible when the outcomes of two actions are on a par in every state. But my argument does not apply to a weakening of Complementary Dominance, according to which one action cannot be better than another if there is no state in which its outcome is better than, *or even comparable with*, the outcome of the other action. Indeed, the following principle seems compelling:

> **Incomparability**: Two actions are incomparable if their outcomes are incomparable in every state.

This means, however, that we still confront a paradox: Incomparability is intuitively compelling, but so is Switch, and these two principles are mutually incompatible if incomparabilities between outcomes are possible. If they are possible and Incomparability holds, then Switch fails to

hold. Actions *Y* and *Z* exemplify this point: if *a* and *b* are incomparable, then the outcomes of *Y* and *Z* are incomparable in every state. But then Incomparability implies that *Y* and *Z* themselves are incomparable. They are not equally good, as Switch would have it.

However, Switch is no less compelling when the outcomes are incomparable. Even then it seems intuitive that switching outcomes between equiprobable states should not affect the value of an action. Why should it matter whether I obtain outcome *a* if the coin falls heads up and outcome *b* if it falls tails up, or vice versa? Or *does* it matter if *a* and *b* are incomparable? Even though, as I hope to have shown, it would not matter otherwise.

These questions remain with us. But then, despite the progress we have made when it comes to interactions between parity and risk, we still are left with the question concerning the interaction between risk and *incomparability*. We seem to have on our hands something that looks like a genuine paradox: both Incomparability and Switch are intrinsically compelling, none of them seems vulnerable to a direct criticism, but if incomparable outcomes are possible, these two principles cannot both be correct.

It is a paradox I do not know how to resolve.

Notes

1 This chapter has been long in the making: its first version was circulated as early as 2016, and it was essentially completed by 2017. Over the years, I presented it at philosophy seminars at York University, London School of Economics, Lund University, the Research School of Social Sciences at the Australian National University in Canberra, Humboldt University in Berlin, and at a workshop on formal theories of value at the University of Colorado at Boulder. I am indebted to the participants in these events and to several other colleagues for very helpful comments and suggestions. Special thanks are due to Matt Adler, Staffan Angere, Adam Bales, Luc Bovens, Richard Bradley, John Broome, Krister Bykvist, Daniel Cohen, Franz Dietrich, Johan Gustafsson, Alan Hájek, Christian List, David Makinson, Jake Nebel, Graham Oddie, Christian Piller, Toni Rønnow-Rasmussen, Mozaffar Qizilbash, Thomas Schmidt, Katie Steele, Orri Stefánsson, and John Weymark. I also received very helpful comments from the editors of this volume, Henrik Andersson and Anders Herlitz. Finally, I want to thank Stephen Sanborn at Proper English AB for a very thorough and competent language review.
2 Three more recent articles should also be mentioned, by Bader (2018) and Doody (2019a, 2019b). They are all well worth studying but will not be commented on here, as they appeared after this chapter had been essentially completed.
3 I have been waiting to get this chapter off my chest, mainly because I continued to hope, in vain as it happens, that I would find a way to solve this residual problem. Now, however, it is time to present what I have and leave the outstanding issues to the reader.
4 It is not essential for the problem at hand that it is the same kind of improvement in both cases or that the size of the improvement is exactly the same (if

it is meaningful to compare these sizes). If a^+ is a trip to the Galápagos Islands at a discount, then b^+ could be a trip to Peru with an extra day in Cuzco.

5 In his comments on this chapter, Anders Herlitz has suggested how this choice problem could arise in real life. Suppose X and Y are two vacation packages one can buy: X, departing on a Tuesday, is slightly cheaper than Y, departing on a Wednesday. Whether you fly on a Tuesday or on a Wednesday does not matter to you, but the price difference does matter, though only marginally since it is quite small. Both packages promise to take you to either the Galápagos Islands or Peru, but the destinations you get to by choosing X and Y, respectively, are entirely dependent on whether the airline decides to fly to Quito on Tuesdays and to Lima on Wednesdays ($S1$) or the other way round ($S2$). The airline has not yet made its decision, and it may go either way, with equal probability. Vacation package X involves departure on a Tuesday, which means it will take you to the Galápagos Islands via Quito (a^+) if the airline decides to fly to Quito on Tuesdays and to Lima on Wednesdays, but it will take you to Peru (b^+) if the airline makes the opposite decision. Vacation package Y involves departure on a Wednesday, which means it will take you to Peru (b) if the airline decides to fly to Quito on Tuesdays and to Lima on Wednesdays, but it will take you to the Galápagos Islands via Quito (a) if the airline makes the opposite decision.

6 In terms of the illustration used (see the preceding endnote), why not opt for the cheaper trip if it does not matter whether you fly on a Tuesday or a Wednesday?

7 Thus, Bales et al. (2014) and Schoenfield (2014) all argue that X and Y are equally permissible in a case like this. Hare (2010), on the other hand, argues that it would be irrational to opt for Y; rationality requires that you choose X.

8 However, as mentioned previously, for Hare it was a problem of rational choice and not an issue in axiology. I will also sometimes use the label "Hare's Problem" for the choice situation itself in which the agent confronts actions X and Y specified as above – I hope that this will not create any confusions. Bader (2018) rather aptly calls such choice situations cases of "opaque sweetening."

9 I have three comments: (i) this principle, like several others that follow, refers to two distinct value relations: one between outcomes and the other between actions. The former relations can be reduced to the latter if we associate each outcome with the "constant" act that assigns this outcome to every state in the partition. Then the betterness relation between outcomes is reduced to the betterness relation between associated constant acts. (ii) As a general principle, Dominance is plausible only if the states in the partition considered are independent of actions (causally or probabilistically independent, depending on the kind of decision theory we accept). In what follows, this act-independence of states will be assumed throughout. (iii) Even with act-independence, Dominance can fail if the number of states is infinite and the outcomes themselves involve elements of risk or uncertainty. A case in point is the well-known Two-Envelopes Problem (cf. Broome 1995; Dietrich and List 2005). Here we only consider cases in which the space of possibilities is partitioned into a finite number of alternative states.

10 Bales et al. (2014: 460) defend a principle of rational choice that is closely related to Complementary Dominance but applies not merely to binary choices. It is framed in terms of "not worse" instead of "not better," but this difference does not matter:

> *Competitiveness*: An action is rationally permissible if there is no state in which its outcome is worse than that of some alternative action.

Assuming that an action is rationally permissible iff it is not worse than any alternative action, Competitiveness follows from Complementary Dominance, though not vice versa. But in binary choices, such as Hare's Problem, the two principles are equivalent.

11 It is easy to see that Complementary Dominance entails the following principle:

> **Neither Nor**: An action is neither better nor worse than another action if there is no state in which its outcome is better or worse than that of the other action.

In what follows, I am going to argue against Complementary Dominance, but my argument will also apply to Neither Nor.

Schoenfield (2014: 267) adopts a principle of rational choice that corresponds to the conjunction of the first clause of Dominance and Neither Nor:

> **Link**: In cases in which considerations of value are the only ones that are relevant, if you are rationally certain that one option, *A*, will bring about greater value than the alternative option, *B*, you're required to choose *A*. If you are rationally certain that neither of the two options will bring about greater value than the other, it's not required that you choose *A*, and it's not required that you choose *B*.

12 Hare (2010), who does not discuss value relations but instead focuses on rational preferences and rational choice, considers, but in the end rejects, a principle of *deference* to one's "better-informed self" (p. 242). In a comparison between two actions, such as *X* and *Y*, my better-informed self, who knows which state obtains, would have no preference in favor of either of these actions. His preferences for actions would be determined by their actual outcomes, and I know that he would not prefer the outcome of one action to that of the other, whichever state actually obtains. My better-informed self would have the same preferences regarding final outcomes as I have, and Hare assumes that I do not prefer a^+ to b or b to a^+ and, similarly, that I do not prefer b^+ to a or a to b^+. But then, if I know that my better-informed self would have no preferences in favor of *X* or *Y*, the principle of deference implies that it is (at least) rationally permissible for me to have no such preferences either. Clearly, this conclusion should be welcomed by friends of Complementary Dominance. (It might be noted that the preferential version of Dominance could be defended along the same deferentialist lines.)

13 Temkin (2012: section 8.3, p.139) considers and defends (with some important qualifications, see endnote 19) *The State-by-State Comparison Principle*:

> For any two prospects [or actions] *A* and *B*, if the value of *A*'s outcome stands in a particular comparative relation, *R*, to the value of *B*'s outcome for *each* possible state of nature, then prospect *A* stands in relation *R* to *B*.

This principle corresponds to (iii), and Temkin's defense of the principle relies on an argument similar to the one I outlined earlier.

14 If this state is not yet determined (before the toss of the coin), then, according to this view, it is not yet determined *which* of the bets is better, although it is determined that they are not equally good.

15 One should bear in mind, though, that some of the defenders of the view that, in Hare's Problem, *X* is not better than *Y* might take this view precisely because they adhere to the ex post approach to the evaluation of actions.

16 This principle is identical to what Temkin (2012: section 8.3: 238) calls *The First Principle of Equivalence*:

> For any two prospects [or actions] *A* and *B*, if for every possible outcome […] that might arise with a given probability […] if *A* is chosen, the same outcome […] might arise with the same probability […] if *B* is chosen, and vice versa, then prospects *A* and *B* are equally good […].

17 Remember that the states under consideration are assumed to be *axiologically irrelevant*, i.e., they do not contribute any value, positive or negative, to outcomes. Without this assumption, Switch would not be a plausible principle.

18 To put it more formally, actions *x* and *y* are equally good if for some equiprobable states *S* and *S′*, $x(S) = y(S')$, $x(S') = y(S)$, and for all other states *S″* in the state partition, $x(S'') = y(S'')$. While Switch is a very simple condition with a highly limited scope of application, it is equivalent to a principle whose application is wider:

> **Permutation:** Let π be a probability-preserving permutation on states in a partition. If for every state *S*, the outcome that action *x* assigns to *S* is assigned by action *y* to $\pi(S)$, then *x* and *y* are equally good.

Switch is a special case of this more general principle; it is a case in which π is a simple transposition of one state into another that is equally probable. At the same time, this special case implies the general principle since any probability-preserving permutation on a finite set of states is the relative product of a finite sequence of probability-preserving transpositions. Consequently, by the transitivity of equal goodness, Switch entails Permutation.

19 An essentially similar argument has been presented by Temkin (2012: section 8.3). The example Temkin considers differs somewhat from the one offered by Hare. In Temkin's example, *Z* results from *Y* by an outcome switch, but action *X* leads to an improved outcome as compared with *Z* only in one state and to the same outcome in the other state. Consequently, in his argument that *X* is better than *Z*, Temkin needs to rely on a stronger principle of dominance than the one I have been assuming. Temkin concludes from this argument that the First Principle of Equivalence (which implies that *Y* and *Z* are equally good) and the State-by-State Comparison Principle (which, as we have seen, implies that *X* is *not* better than *Y*) are incompatible with each other. He finds the former principle particularly compelling. He therefore suggests that the State-by-State Comparison Principle should be restricted to accommodate the First Principle of Equivalence: "Perhaps the most natural way to do that would be to limit the scope of the State-by-State Comparison Principle to all and only those cases where it would not directly conflict with the First Principle of Equivalence" (Temkin 2012: 242). Temkin admits, however, that in the absence of justification, it is an ad hoc solution.

20 An even shorter, more direct argument for the same conclusion appeals to Stochastic Dominance as its only premise (see Bader 2018).

> **Stochastic Dominance:** An action *x* *stochastically dominates* an action *y* iff (i) for every possible outcome *o*, the probability that *x* will result in an outcome at least as good as *o* is at least as great as the corresponding probability for *y*, and (ii) for some *o*, this probability is greater. If clause (i) is satisfied, *x* is at least as good as *y*. If (i) and (ii) are satisfied, *x* is better than *y*.

It is easy to see that *X* stochastically dominates *Y* and thus is better than *Y* according to this condition. But it may be questioned whether Stochastic

Dominance is as intuitively compelling as the principles I have used in my argument from Switch. Stochastic Dominance is in any case much stronger than Statewise Dominance and Switch taken together. Admittedly, appealing to it in the argument for X being better than Y obviates the need to assume that betterness is transitive across equal goodness. Although this may be seen as an advantage by some, the advantage is minimal if one finds, as I do, this transitivity condition extremely intuitive.

In private communication, Orri Stefánsson has suggested another principle that would immediately establish that X is better than Y:

> An action x is better than an action y if the outcome of x would in every state be better than that of y if y's outcomes were switched between some equiprobable states.

This condition is much weaker than Stochastic Dominance and yet it also makes it unnecessary to appeal to the transitivity of betterness across equal goodness. But there is a price to pay: the condition is ad hoc, being especially designed to immediately get us the conclusion we are after. Jumping to the conclusion in this way is unattractive; it is preferable to proceed to it by separate steps that can be subjected to independent scrutiny.

21 Hare (2010) defines the notion of *prospects* associated with an action: each such prospect is a possible outcome of that action paired with the agent's credence that the action would yield the outcome in question. The set of all prospects associated with an action is thus just like a lottery but with objective probabilities replaced by credences. Hare then assumes the following principle:

> *Prospects Determine Permissibility*: Facts about what it is rationally permissible for me to do are determined by facts about the prospects associated with the options available to me.

> (Hare 2010: 240)

This principle is similar in spirit to the Reduction of Actions to Lotteries.

Hare then considers two choice situations, one in which we choose between X and Y (with equal objective probabilities of states replaced by equal credences) and the other in which the choice is between X and Z. He implicitly assumes the preferential version of Statewise Dominance as a criterion of permissibility, and thus concludes that Z is impermissible in the latter choice problem. This follows even if the agent has incomplete preferences, as long as she prefers a^+ to a and b^+ to b. But since the prospects associated with Y and Z are the same and since in both choice problems the alternative action is the same (X), if Prospects Determine Permissibility, Y must be impermissible in the choice between X and Y, just as Z is impermissible in the choice between X and Z.

Hare also puts forward a general theory of rational choice, Prospectism, which as a special case implies that Y is impermissible if X is available. That theory, however, rests on strong assumptions, much stronger than those I am willing to rely on in this chapter. According to Prospectism, an action, interpreted as a set of prospects, is permissible iff it maximizes expected utility (as compared with other available actions) with respect to some utility function on outcomes that is compatible with the agent's possibly incomplete preferences over outcomes. Needless to say, the view that an action is permissible only if it is in this way rationalizable in terms of expected utility maximization is contentious. At the same time, the theory might be criticized for being too permissive in that it imposes no restrictions on admissible utility

functions apart from the requirement that they must be compatible with the agent's outcome preferences. In the case we consider, if the agent's preferences have gaps when it comes to comparing a and a^+ with b and b^+, Prospectism implies that action Y becomes permissible as soon as the agent ever so slightly increases her credence for one of the states. For this objection to Prospectism, see Nissan-Rozen (2015).

22 Nor would it be possible to use Stochastic Dominance to establish this conclusion. It would no longer be the case that X stochastically dominates Y. To see this, suppose that $S1$ is more probable than $S2$. Then the probability that X results in an outcome at least as good as b is lower than the corresponding probability for Y. Analogously, if $S2$ is more probable than $S1$, the probability that X results in an outcome at least as good as a is lower than the corresponding probability for Y.

23 For a discussion of different kinds of regret and ways of allaying them, and in particular for the distinction between *outcome regret*, which arises when we reflect on what our action could instead have brought about, and *action regret*, which comes from reflecting on what would have been brought about if another action had been chosen, see Bovens and Rabinowicz (2015).

24 I am indebted to Anders Herlitz for pressing this point.

25 This formulation ignores various qualifications and provisos we might want to impose, for example, that an agent ought to hold a pro-attitude toward i only if she knows what i is like, that the "ought" in question is an ought of recommendation rather than a strict demand, or that the reasons for holding a pro-attitude (i.e., the considerations that explain why one ought to hold it) should be of the "right kind" – they should invoke value-making features of the item under consideration rather than extraneous factors (for this last point, see Rabinowicz and Rønnow-Rasmussen 2004). In what follows, I make a tacit assumption that these qualifications are satisfied.

26 This definition does not exclude, of course, that in cases of parity it might also be permissible to be indifferent between the items in question or even to have no preferential attitude as far as they are concerned. Indeed, in typical cases of parity, all these possibilities can be expected to be permissible (cf. Rabinowicz 2008).

27 As previously, we think of an action as a function from states in a state partition to outcomes. In some cases, particularly in the cases we focus on, the elements of the state partition are assigned objective probabilities. In what follows, whenever two actions are being compared, they are assumed to be based on the same state partition.

28 That the first clause of Dominance (P) holds for all orderings in **K** is both necessary and sufficient for the corresponding first clause of Dominance. However, that the second clause of Dominance (P) holds for all orderings in **K** is not sufficient for the second clause of Dominance. Using the former principle, we can prove that if the outcomes of x are at least as good as the outcomes of y in every state, then x is weakly preferred to y in every ordering in **K**. But this falls short of x being at least as good as y. To be at least as good as y, x must be better than y or equally as good as y, i.e., it must be preferred to y in every **K**-ordering or equi-preferred with y in every **K**-ordering. It is not enough if it is preferred in some orderings and equi-preferred in others. This problem would be avoided if one *redefined* the notion of being at least as good: i could be said to be at least as good as j iff i is weakly preferred to j in every ordering in **K** (cf. Rabinowicz 2008). While "at least as good" on this redefinition would still be entailed by "better or equally good," the converse entailment would no longer hold.

29 That this condition holds for all orderings in **K** is both necessary and sufficient for Switch.

30 Or, by contraposition: if action x is preferred to action y in every ordering in **K**, then there is some state S such that the outcome of x in S is preferred to the outcome of y in S in every ordering in **K**.

31 More generally, for any complete ordering P of the items in a certain domain, an assignment v of numerical values to the items in this domain *represents P* iff v assigns higher numbers to items placed higher in P and the same numbers to items that are equal ranked in P.

32 In fact, *P1* and *P2* are the only permissible complete orderings in which the preference for Y over X receives *any* support at all. Proof: the preference for Y over X receives some support only if (1) a is preferred to b^+ or (2) b is preferred to a^+. Since a^+ is preferred to a and b^+ is preferred to b, (1) implies, by the transitivity of preference, that the four items are ordered as in *P1*, while (2) implies, again by transitivity, that they are ordered as in *P2*.

   At the same time, in any complete ordering, the preference for X over Y must receive *some* support. Proof: if a^+ is ranked above a and b^+ is ranked above b, then, by completeness, (i) a^+ is preferred to b or (ii) b^+ is preferred to a. If (i) holds, the preference for X over Y receives support in *S1*. If (ii) holds, this preference receives support in *S2*.

33 But could the preference for Y over X be supported in a permissible incomplete ordering without the preference for X over Y being supported? No, it could not. If the preference for Y over X is supported, this must mean that the ordering ranks (i) a above b^+ or (ii) b above a^+. Being permissible, it ranks (iii) a^+ above a and (iv) b^+ above b. By the transitivity of permissible preferences, (i), (iii), and (iv) imply that the ordering in question is identical to *P1*, while (ii), (iii), and (iv) imply that it is identical to *P2*. And we already know that in both these complete orderings, the preference for X over Y is supported and, indeed, that it is more pronounced than the preference for Y over X.

34 *P′ agrees with P* iff for all items i and j in **I**, if i is preferred to (equi-preferred with) j in P, then i is also preferred to (equi-preferred with) j in $P′$. Note that agreement is not a symmetric relation: if $P′$ agrees with P, P might not agree with $P′$. Indeed, this will always be the case if $P′$ fills some gaps in P.

35 This condition can also be equivalently stated as follows:

   If the set $\Delta(P, G)$ of all permissible completions of P with respect to G is not empty, then $P = \cap\Delta(P, G)$.

   To establish this equivalence, it is sufficient to note that if P is the intersection of a non-empty set of orderings, then adding to that set further orderings that agree with P will not make any difference to the intersection.

36 A complete ordering is of course trivially well-rounded since it does not contain any gaps.

37 It is not sufficient since the antecedent of (*) does not require that the antecedent of Complementary Dominance (P) should be satisfied by any ordering in **K**. Thus, Complementary Dominance (P) cannot be used to prove the consequent of (*) from its antecedent. Nor is Complementary Dominance (P) necessary for (*) since the consequent of (*) is relatively weak: it only stipulates that there exists *some* ordering in **K** in which action x is not preferred to action y. We could have a model in which some such ordering exists, even though **K** also contains *another* ordering that violates Complementary Dominance (P) with regard to actions x and y.

38 This counterexample also applies to what might be seen as a kind of preferential analogue of Incommensurability, i.e., the principle that requires a preference gap between two actions if there are preference gaps between their outcomes in every state. It should be kept in mind, though, that

incommensurabilities in value and preference gaps are not the same thing. If a preference gap between two items is permissible, this entails that the items in question are incommensurable, but the opposite implication need not hold.
39 Note, however, that Hare, who is interested in the preferential version of his problem, assumes in his examples that the agent lacks preferences regarding these options. My interest is in the value relations, so I allow both the absence of preference and its presence, although in the latter case, I assume that it is permissible to have a preference either way.
40 Or this impermissibility might at least apply to considered preferential attitudes Both considered preference and considered equi-preference might be disallowed.
41 I am indebted to Luc Bovens for suggesting this variant.
42 In Rabinowicz (2008), I suggested that *Sophie's Choice* might be an example of a situation in which the alternative outcomes are incomparable from Sophie's perspective: she must choose which of her children to save, but it is impermissible for her to prefer one child rather than the other to survive, nor is it permissible for her to be indifferent between these outcomes.
43 I am indebted to Graham Oddie for pressing this point.
44 That preferences in permissible orderings are transitive across indifference is a crucial assumption in my model. Without it, the model would not validate the condition that betterness is transitive across equal goodness.

References

Bader, R. M. (2018), 'Stochastic Dominance and Opaque Sweetening', *Australasian Journal of Philosophy* 96/3: 498–507.
Bales, A., Cohen, Daniel, and Handfield, T. (2014), 'Decision Theory for Agents with Incomplete Preferences', *Australasian Journal of Philosophy* 92/3: 453–470.
Broome, J. (1995), 'Two Envelope Paradox', *Analysis* 55/1: 6–11.
Bovens, L., and Rabinowicz, W. (2015), 'The Meaning of 'Darn It!'', in I. Hirose and A. Reisner (eds.), *Weighing and Reasoning: Themes from the Philosophy of John Broome*, Oxford University Press, Oxford: 129–139.
Chang, R. (2002), 'The Possibility of Parity', *Ethics* 112/4: 659–688.
Dietrich, F., and List, C. (2005), 'The Two-Envelope Paradox: An Axiomatic Approach', *Mind* 114/454: 239–248.
Doody, R. (2019a), 'Opaque Sweetening and Transitivity', *Australasian Journal of Philosophy* 97/3: 559–571.
Doody, R. (2019b), 'Parity, Prospects, and Predominance', *Philosophical Studies* 176/4: 1077–1095.
Hare, C. (2009), 'Perfectly Balanced Interests', *Philosophical Perspectives* 23/1: 165–176.
Hare, C. (2010), 'Take the Sugar', *Analysis* 70/2: 237–247.
Hare, C. (2013), *The Limits of Kindness* (Oxford University Press).
Nissan-Rozen, I. (2015), 'Review of *The Limits of Kindness* by Caspar Hare', *Economics and Philosophy* 31/3: 493–499.
Rabinowicz, W. (2008), 'Value Relations', *Theoria* 74/1: 18–49.
Rabinowicz, W., and Rønnow-Rasmussen, T. (2004), 'The Strike of the Demon: On Fitting Pro-attitudes and Value', *Ethics* 114/3: 391–423.
Schoenfield, M. (2014), 'Decision Making in the Face of Parity', *Philosophical Perspectives* 28/1: 263–277.
Temkin, L. S. (2012), *Rethinking the Good – Moral Ideals and the Nature of Practical Reasoning* (Oxford University Press).

11 Incommensurability That Can(not) Be Ignored

Katie Steele

1 Decide Based on That Which One Has an Opinion

It seems natural and wise to make decisions on the basis of that which one has opinions, ignoring that which one has no clear opinion. For instance, consider the following very mundane decision problem:

> *Bike Route*. I must decide which way to ride to work – whether the most direct route through the city center or a more convoluted route that is better connected by bike lanes and green spaces. There are a number of salient differences between these routes that matter for their evaluation, concerning such properties as their duration, safety, and aesthetics. There may well be other ways in which the two options differ, but I struggle to bring them to mind, let alone compare the bike routes with respect to them. For instance, how might my taking each route affect the well-being of all those I happen to interact with on the way and all those who interact with those people?

In framing the Bike Route decision problem and deliberating about what to do, it may seem reasonable that I focus just on those aspects or *criteria* for evaluating (pairs of) bike routes about which I have a clear opinion or preference relation with respect to the options. What is meant here by a "clear" opinion or preference relation is that I either (strictly) prefer one option to the other in the pair or else I am indifferent between the options in the pair. To lack a clear opinion or preference in this sense is then, let us say, to find the options *incommensurable*. (Here the term "incommensurable" is treated as synonymous with other terms of art in the literature, like *incomparable* or *on a par*. Any purported differences or ways in which these are distinct forms of lacking a regular preference relation between options are not important to the discussion here.) The thought is that the criteria under which the bike route options are incommensurable are precisely those that are less obvious or salient to me. While criteria under which options are incommensurable may sometimes

be salient, they are typically not, and that is perfectly reasonable, one might think, if such criteria are not important in decision-making.

My Bike Route deliberations seem entirely commonplace. If the aforementioned is an apt characterization of these deliberations, then it follows that we do commonly ignore in our decision-making that which we have no opinion about or criteria for which options are incommensurable since this is as mundane a decision problem as any. That is, it suggests that we invoke, in framing our decisions, some kind of "incommensurability independence" principle, whereby the relative value of options is independent of those aspects of the options that admit of no clear opinion or preference relation between them (namely, strict preference or indifference) such that they are incommensurable. Note that options may be incommensurable with respect to a criterion due to lack of opinion about the desirability of states of affairs by that criterion and/or lack of opinion about how likely it is that the options will result in these various states of affairs or realizations of the criterion. For instance, perhaps I cannot rank the bike routes with respect to the comfort of other riders because I cannot determine whether it is more or less or equally comfortable for a bike route to be more crowded.[1] Otherwise (or in addition), perhaps I cannot rank the bike routes with respect to the comfort of other riders because I cannot assign precise probabilities to the relative size of the crowds on the bike routes on this particularly fresh spring day.

A principle akin to "incommensurability independence" is apparently defended by L. A. Paul in her (2014) book *Transformative Experience*. Paul claims that decisions involving transformative experiences, by definition, are decisions for which the options cannot be compared with respect to many aspects or criteria that one cares about (and thus also happen to be salient, in this case). Consider for instance the decision about whether to take on a new social role, such as whether or not to become a mother or whether or not to train to be a professional astronaut. The status quo and the role-change option may be difficult to compare with respect to criteria like "sense of personal achievement" and "level of anxiety." Nonetheless, Paul suggests that one can reasonably come to a choice based on the remaining aspects or criteria about which one does have an opinion, often simply the "surprise or adventure" that the options afford.[2]

In other discussions too, something like an "incommensurability independence" principle is floated as the only way to proceed in making rational decisions. For instance, Lenman (2000) raises a puzzle for those committed to (the relevance of) consequentialist ethical reasoning. The idea is that even the most foresighted agent has a sense of "cluelessness" about how the consequences of their options compare under the full myriad of ethically relevant aspects or criteria. One will easily become perplexed, for instance, in trying to predict which future persons may possibly exist and what these persons will go on to do and how good this

would be, were some given option pursued. Greaves (2016: 312) says, in light of this problem, "In attempting to take consequences into account in practice in decision-making, we usually focus on those effects – let us call them 'foreseeable' effects – that we take ourselves to be able to foresee with a reasonable degree of confidence." One reading of this common wisdom (although not the one that Greaves favors) is that agents do and indeed rightly focus just on those aspects of the consequences of options that permit a clear comparison, in the sense used earlier, between them.

Finally, it seems that in the very framing of a decision problem, we routinely overlook aspects of options for which they are difficult to compare. Our reasoning concerns only the salient aspects of options or criteria for comparing them, and, as suggested earlier, the many aspects of options that are difficult to pin down and admit of no comparison between them are often not salient. So something like an "incommensurability independence" principle seems to play a role in shaping our very perception of a decision problem – the apparent starting point for our deliberations. Of course, if this really were the starting point for our deliberations it must by definition lie beyond rational principles. But one might suspect that what we typically regard as the starting point for our deliberations is actually the result of some prior (un)principled reasoning.

All told there is arguably a lot of implicit support for and allegiance to an "incommensurability independence" principle, whereby in comparing options, one simply ignores those aspects or criteria under which neither option is preferred to the other and nor is one indifferent between them. This chapter assesses whether this principle for comparing options is plausible. In fact, the chapter will assess just whether a weaker principle, entailed by the aforementioned, is plausible: a choice principle based on a kind of non-dominance relation that accommodates incommensurability between options on any given criterion. The non-dominance principle (or rather two versions of it, weak and strong) is introduced in Section 2. The principle is compared with a somewhat analogous principle in the literature on incommensurability and uncertainty owing to Caspar Hare (2010).

In Section 3, it is argued that neither of the two versions of the non-dominance principle is plausible. The case against the strong version is relatively straightforward – it is easy to construct counterexamples. But the weak version too is subject to counterexamples, albeit of a special and more unusual form. Whether or not any of these counterexamples are genuine, however, might be open to challenge, especially in light of the debate inspired by Hare's analysis. I show in Section 4 that, despite the analogies, there are significant dis-analogies between the multi-criteria and uncertainty settings, and so the counterexamples retain their force. Finally, Section 5 returns to the question of whether we do routinely ignore aspects of our decisions for which we cannot compare options or whether our ordinary deliberations can be cast in a more reasonable light.

2 Making the Principle(s) Precise

To explore the form of reasoning suggested in the previous section, I will introduce a different decision problem, for which, by assumption, there is *salient* incommensurability between the options. That is, the options are incommensurable with respect to certain aspects of their consequences or criteria that are nonetheless salient (perhaps owing to the fact that one of the options is recognizably novel or unfamiliar). When incommensurabilities are explicitly tabled in this way, we can examine whether it is generally coherent to "ignore incommensurabilities" in one's reasoning.

The problem is a very simplistic climate change policy problem. The choice is between the status quo option of business-as-usual or inaction on climate change, and a geoengineering option. We will assume that for the (rather short-sighted) decision-maker at hand, there are only two aspects of the outcome, or two criteria, that matter for the comparison of the options: the welfare of the present generation (labeled "current") and the welfare of the subsequent generation (labeled "future"). These criteria are *separable* (in the sense that they amount to independent, additive contributions to the overall value of the options) and together exhaust all that matters for the decision-maker in comparing the options.

To add some color: the geoengineering option involves the current generation paying for this capability and the subsequent or future generation implementing it. The status quo is thus perceived to be better for the current generation than the geoengineering option since it does not involve costs associated with building geoengineering capability. Hence the utility associated with the former, a, is greater than the utility associated with the latter, b. The two options are, by assumption, incommensurable for the future generation – each has indeterminate utility for this criterion (compatible with incommensurability), represented by "*ind*." (In subsequent models of the problem, further assumptions will be introduced by way of making the indeterminate utilities more precise).

The informal choice principle of "ignoring incommensurabilities" recommends the status quo option in this decision problem. In comparing the options, we ignore the "future" criterion, as the options are incommensurable under this criterion – both business-as-usual and the deployment of geoengineering are, after all, very hard to evaluate, and ultimately compare, for the generation directly experiencing them. We then focus just on the "current" criterion, or what the options are like for the current generation. Under this criterion, the status quo option is preferred.

The reasoning just described in fact amounts to a special case of "ignoring incommensurabilities" – it involves no more than an appeal to (some kind of) dominance of one option over the other. That is, it appears we need only appeal to a choice principle that invokes the following rather compelling non-dominance relationship (the choice of terminology will be explained next):

(**Criterion-Wise**) **Competitive**. An option is *not* worse than another option, and thus (criterion-wise) competitive, if, under every *criterion*, its outcome is *not* worse than that of the other option.

Only the status quo option is (criterion-wise) competitive. The status quo option is not worse than the geoengineering option for those criteria under which the options are incommensurable (here only "future") and it is moreover not worse than the geoengineering option (in fact it is better) under the remaining "current" criterion. As such, the status quo option is surely a permissible option, and plausibly uniquely so.

The choice principles just alluded to can in fact be formalized as weak and strong versions of what we might call "(criterion-wise) competitiveness," as follows:

Weak (Criterion-Wise) Competitiveness. If an option alpha is (criterion-wise) competitive with respect to all other alternative options, it is rationally permissible to perform alpha.

Strong (Criterion-Wise) Competitiveness. If an option alpha is (criterion-wise) competitive with respect to all other alternative options, it is rationally permissible to perform alpha, and it is not rationally permissible to perform an option beta that is not similarly (criterion-wise) competitive with respect to all other alternative options.

Both of these non-dominance or so-called competitiveness principles seem compelling, even the strong version. Moreover, they are analogues of non-dominance principles (across states) that have been recently entertained by Hare (2010) and defended by Bales et al. (2014). Indeed the terminology here follows that of Bales et al. (2014).[3] One can restore the state-wise principles by replacing the aforementioned "criterion" with "state." Start with the very notion of a non-dominated or competitive option:

(**State-Wise**) **Competitive**. An option is *not* worse than another option, and thus (state-wise) competitive, if, under every *state*, its outcome is *not* worse than that of the other option.

This notion then features in the two choice principles (with labels roughly in keeping with those of Bales et al. (2014)):

Weak (State-Wise) Competitiveness. If an option alpha is (state-wise) competitive with respect to all other alternative options, it is rationally permissible to perform alpha.

Strong (State-Wise) Competitiveness. If an option alpha is (state-wise) competitive with respect to all other alternative options, it is

rationally permissible to perform alpha, and it is not rationally permissible to perform an option beta that is not similarly (state-wise) competitive with respect to all other alternative options.

Despite the analogy with plausible state-wise, non-dominance choice principles, neither the strong nor the weak criterion-wise versions are themselves compelling. This will be demonstrated in the following section. We will return to why the two sets of choice principles – state-wise and criterion-wise, respectively – have differing status, due to important dis-analogies, later.

3 Against the Criterion-Wise "Competitiveness" Principles

We can interrogate the strong (criterion-wise) competitiveness principle by initially considering different versions or instantiations of the decision problem described in Table 11.1. This involves examining the different ways in which the options may be incommensurable under the second criterion ("future") and how this is weighed against the preference for the status quo under the first criterion ("present"). Assume that the (expected) utilities in Tables 11.2, 11.3, and 11.4 are comparable and scaled in such a way that the two criteria are weighted equally. Indeterminate (expected) utility is modeled in terms of sets of (expected) utility functions; two options are incommensurable when the (expected) utility functions in the set do not agree on the ranking of the options. Finally, the overall (expected) utilities for the options are determined in the usual way, as a weighted average of the (expected) utilities for the options under the criteria.

Table 11.2 gives one instantiation of the more general geoengineering problem described in Table 11.1. The (expected) utilities a and b (where $a > b$) associated with the present outcomes are specified to be 0 and -100, respectively. And the sets of (expected) utilities for the future generation

Table 11.1 A multi-criteria problem with explicit incommensurability

	current	future
status quo	a	*ind*
geoengineering	b	*ind*

Table 11.2 So far, so good for (criterion-wise) competitiveness

	current	future
status quo	0	$\{-50, -50\}$
geoengineering	-100	$\{-500, 40\}$

outcomes represent incommensurability between these outcomes as required. While the status quo is expected to yield −50 utils for the future generation, the geoengineering option is expected to yield either −500 or 40. (For simplicity, let us assume this is a discrete set containing only two expected utility functions.[4]) However, once we take into account the current generation as well, the two options are *not* incommensurable. The status quo has expected utility of −25 (the result of $0.5 \times 0 + 0.5 \times -50$). Geoengineering has expected utility of {−300, −30}; that is, it either has expected utility of −300 (the result of $0.5 \times -100 + 0.5 \times -500$) or else −30 (the result of $0.5 \times -100 + 0.5 \times 40$). Both of these expected utilities for geoengineering are less than −25 and so the status quo option comes out as clearly preferable overall. This looks good even for the strong (criterion-wise) competitiveness principle. By this principle, the status quo option is uniquely permissible, in line with the numbers.

But we can easily see how the numbers might not have been so favorable for strong (criterion-wise) competitiveness. Consider the version of the geoengineering problem in Table 11.3, which is a slight adjustment of the previous version.

Again, the expected utility of the status quo is −25. But this time, the expected utility of the geoengineering option is {−300, −20}. So by the standard method for determining the overall evaluation of options, the status quo and geoengineering options are incommensurable. According to one expected utility function, the status quo is better since −25 > −300. But according to the other, the geoengineering option is better since −25 < −20. So it is not the case that the status quo is uniquely permissible, as per strong (criterion-wise) competitiveness.

Before moving on, it is important to address a worry about where the (expected) utility values in the above tables and those that follow come from. One might wonder what licenses any given representation of a decision problem. Surely the representation is derived from the agent's overall preference ordering, and what is at issue here are constraints on that very preference ordering. If strong (criterion-wise) competitiveness is a constraint on rational preferences, then Table 11.3 together with the stated calculus for determining the overall evaluation and thus ranking of the options simply does not represent a rational agent. It might be added that an objection to strong (criterion-wise) competitiveness must involve a demonstration that the principle conflicts with an alternative and even

Table 11.3 A *prima facie* violation of strong (criterion-wise) competitiveness

	current	future
status quo	0	{−50, −50}
geoengineering	−100	{−500, 60}

more compelling principle of preference/choice. And that has not been provided here.

A preliminary response to this objection is simply to push the can a little down the road. For now, we are assuming that a rational agent entertaining a decision problem for which the criteria are separable and exhaustive of what matters to her and moreover admit of comparable (expected) utility measures, has overall preferences that can be represented as the weighted average of her expected utilities across the criteria.[5] (We are further allowing that preferences may be incomplete under any given criterion, represented as sets of expected utilities, yielding sets of weighted average expected utilities overall.) The initial question is not whether strong or weak (criterion-wise) competitiveness *should* be regarded as a primitive or basic constraint on rational preferences. Rather, it is the more modest question of whether there is in fact a conflict between strong or weak (criterion-wise) competitiveness and the model of rational preference just described. And we see from Table 11.3 that there is indeed a conflict. This provides some reason to doubt strong (criterion-wise) competitiveness; whether we should rather rethink our supposed model of rational preference will be taken up in the next section.

One way to pose the lesson here is that when a pair of options are incommensurable for one or more criteria in a decision problem, on the model we are supposing, the incommensurability can "swamp" the overall evaluation of the options, even if one option is strictly preferred on the remaining criteria. It is only in special cases that the incommensurability of options on some criteria can be regarded as "minor," in the sense that the associated range of expected utilities does not undermine the strict preference for one option on the remaining criteria. Hence strong (criterion-wise) competitiveness is not generally consistent with the model of rational preferences in the multi-criteria setting that is supposed here.

Let us turn now to the weak (criterion-wise) competitiveness principle. Are there cases where, despite being (criterion-wise) competitive or non-dominated, an option is nonetheless not permissible? The answer is "yes." Again, a small variation on the previous decision problem can show that even weak (criterion-wise) competitiveness is not generally consistent with the model of rational preference presupposed here.

In the new case, described in Table 11.4, the options are incommensurable under both criteria. Hence by weak (criterion-wise) competitiveness, both options are permissible. And yet, by the lights of our supposed model of rational preference, the geoengineering option here is not permissible. The geoengineering option will be either worse than the status quo for the current generation by 100 and yet better for the future generation by 40, or it will be better than the status quo for the current generation by 20, and yet worse for the future generation by 300. As such, geoengineering is overall determinately worse. Indeed, the status quo has an overall expected utility

of −25 while geoengineering has (determinately less) overall expected utility of {−55, −165}; that is, it either has an overall expected utility of −55 (the result of 0.5 × −100 + 0.5 × −10), or else it has overall expected utility of −165 (the result of 0.5 × 20 + 0.5 × −350).

Admittedly, this is a rather peculiar case, insofar as the agent's attitudes go. The peculiarity is the strange complementarity between the agent's preference attitudes across the two criteria *on all candidate expected utility functions*: Either geoengineering is worse for the current generation but better for the future generation, or else geoengineering is better for the current generation and yet worse for the future generation. This pattern is crucial for geoengineering coming out *determinately* worse overall. (All that is further required for that result is that geoengineering is *not sufficiently better* for the current/future generation to compensate for just how much worse it is for the other generation.) In particular, the other two logical possibilities – whereby geoengineering is worse on both criteria or else better on both criteria – must not be live candidates for the agent.

What kind of story might we tell to make the aforementioned model seem plausible? Let the geoengineering option involve the current generation doing a trial implementation and addressing any problems that arise before the future generation does a full implementation. Perhaps the agent is divided between two evaluations because they cannot assign precise probabilities to the possibility that crucial problems will be identified in the trial. If the problems are identified, then the geoengineering option will be worse for the current generation but slightly better, let's say, for the future generation; if the problems are not identified, then the geoengineering option will be better for the current generation but much worse for the future generation when hitherto unforeseen problems arise that are nonetheless expected to be highly detrimental. Assume the agent has a rather strange belief state: for whatever reason, she thinks either the crucial problems being identified is a certainty, or else the crucial problems *not* being identified is a certainty (and nothing in between). Hence there are two possible expected utility representations across the two criteria, as per Table 11.4. One can appreciate from this sketchy narrative that decision problems like those described in Table 11.4 are rather unusual. Nonetheless, they are a conceptual possibility. Indeed, a simpler problem in the same vein will be introduced in the next section.

Table 11.4 A *prima facie* violation of weak (criterion-wise) competitiveness

	current	*future*
status quo	{0, 0}	{−50, −50}
geoengineering	{−100, 20}	{−10, −350}

4 Comparison with the State-Wise Analogues

It was briefly noted earlier that there is room for dispute about the status of the *state-wise* competitiveness principles. Hence one might presume that there should equally be room for dispute about the status of their *criterion-wise* counterparts. As such, one might think that the supposed model of rational (multi-criteria) preference in Section 3 is more contentious than first meets the eye such that the conflict with the (criterion-wise) competitiveness principles is far from devastating for the latter. A closer look at the extent of the analogy between the state-wise and the criterion-wise principles does not support this optimism about the latter, however. That is what we turn to now: a demonstration that the two versions of competitiveness are *not* on a par as constraints on rational preference.

We will focus on the weak form of the competitiveness principles since these are lesser constraints on rational preference. The last section presented a prima facie case that *not even weak* (criterion-wise) competitiveness is a plausible constraint on rational preference. The following decision problem – "Two Transparent Boxes" – presents this case in simpler terms, (even) more contrived as it may be. Indeed, the problem will play a useful role in our subsequent discussion because it has the same tabular form – with the exception of the labeling and meaning of the columns – as the decision problem introduced by Casper Hare (2010) entitled "Two Opaque Boxes." Hare's problem inspired much further discussion of constraints on rational preference where incommensurability and risk are concerned, including Bales et al.'s (2014) defense of what we here refer to as weak (state-wise) competitiveness.

Here is the analogue to Hare's decision problem (with crucial similarities to the final geoengineering case, as will be elaborated next):

> *Two Transparent Boxes.* You show me items *A* and *B*, two dollar coins, two twins Pablo and Penelope, and two transparent boxes, the contents of which must be distributed between the twins in specified ways. I consider *A* and *B* to be incommensurable with respect to both Pablo and Penelope's welfare. Perhaps *A* is a ticket to the opera while *B* is a ticket to an A-league sports game. (Moreover, I consider the twins to derive the same welfare from the same items.) The twins' welfare is all that matters to me and the welfare of each matters equally to me. I see that Box 1 contains *A* for Pablo and *B* for Penelope, while Box 2 contains *B* for Pablo and *A* for Penelope, plus an extra dollar for each. Then you invite me to choose one of the boxes so as to benefit the twins in the specified way.

Table 11.5 depicts our decision problem in tabular form, where *A+* can be interpreted as the outcome *A* plus one dollar, and *B+* can be interpreted similarly. Here the columns represent the independent criteria

Table 11.5 Two transparent boxes

	column 1 – Pablo (0.5)	*column 2 – Penelope (0.5)*
Box 1	A	B
Box 2	B+	A+

or sources of value – Pablo's and Penelope's respective welfare – that matter (equally, by assumption) to the decision-maker. As noted earlier, "Two Transparent Boxes" is constructed to have the same tabular form as Hare's "Two Opaque Boxes" problem, the difference being the meaning of the columns. In Hare's problem, the columns represent rather the possible states of the world (each equi-probable by assumption since they are the results of a fair coin toss). (A further difference then is that the problem in Table 11.5 involves two dollar coins, while Hare's problem requires just one for which the accompanying prize is uncertain.) In both problems it is assumed that A+ is strictly preferred to A, B+ is strictly preferred to B, and A and A+ are each incommensurable with B and B+, independent of the column, whether the criterion or the supposed state of the world.

The decision problem depicted in Table 11.5 – whether the transparent or opaque version – can be regarded as having the same crucial "complementarity" in expected utility values across the columns as the final geoengineering problem above. Assume that the incommensurability between A and B is represented by a set of expected utilities, some of which put A higher in value and some of which put B higher in value. Recall that when a dollar is added to either prize, the two remain incommensurable. On some expected utility functions then, A is better than B plus a dollar, favoring Box 1 in the first column. But on those very same expected utility functions, A plus a dollar is better than B, by an even greater amount. So Box 1 loses by a greater amount in the second column than it wins in the first column. Likewise, for the expected utility functions in which Box 1 wins on the second column, it loses by a greater amount on the first column.

By the weak (criterion-wise) competitiveness principle, both Box 1 and Box 2 are permissible choices in the Two Transparent Boxes problem. Similarly, by weak (state-wise) competitiveness, both boxes are permissible choices in the Two Opaque Boxes problem. That is because, under each of the columns (i.e., each of the criteria/states), the two options are incommensurable. Hence each option is not worse than the other across all the columns and thus each option is competitive. But surely, whichever problem one faces, the choice of Box 2 is better! Moreover, our utility model described above – revealing the "complementarity" between the column utilities (to the detriment of Box 1) for all possible utility functions – apparently reveals why it is that Box 2 comes out determinately better.

Some argue that Box 1 in "Two Opaque Boxes" is nonetheless a permissible choice (for instance, Bales et al. 2014). The major positive argument for this claim (noted already by Hare, 2010) involves appealing to the agent's attitudes *were she fully informed*. In "Two Opaque Boxes" the agent may reason that there are two ways in which the world might go, corresponding to the two states of the world. Or, put differently, if she were fully informed, there are two possible states of knowledge she might have. Either way, her fully informed self will not find Box 1 worse than Box 2 (in fact, she will find them to be incommensurable). So why not defer to her fully informed self and regard the two boxes as permissible choices, neither one worse than the other, even under uncertainty? The general principle might be stated as follows:

> **Defer to Fully Informed Self**. If you know that your fully informed self will find an option no worse than alternatives, then you should now, even under uncertainty, find this option no worse than alternatives and thus permissible.

Something along the lines of the aforementioned is typically granted in the literature as a reason to endorse the weak (state-wise) competitiveness principle. The dispute is about whether this is a *good enough* reason to overcome the reasons *against* endorsing the weak (state-wise) competitiveness principle. But we need not here take a side in this debate (see, however, Rabinowicz, 2021, for further analysis).[6] What is important for our discussion is that this reason in favor of weak (state-wise) competitiveness does not translate well to the multi-criteria setting. The analogue general principle is as follows:

> **Defer to Fully One-Track-Minded Self**. If you know that your fully one-track-minded self will find an option no worse than alternatives, then you should now, even taking into account all criteria, find this option no worse than alternatives and thus permissible.

In the case of "Two Transparent Boxes," the idea is that when the decision-maker focuses on the respective plights of Pablo and Penelope (the "criteria" in this decision problem), she finds neither box worse than the other. Hence, she should find neither box worse than the other overall when all criteria are taken into account.

The crucial problem with this proposed maxim to defer to one's fully one-track-minded self is that this self (or rather these selves), unlike one's fully informed self (or possible selves), has no claim to authority. It simply does not matter what would be my attitudes were I to only care about any single aspect or criterion of choice. That has nothing to do with either my actual or idealized preference attitudes.

These considerations bear on the question of whether "complementarities" in candidate expected utilities across different columns (criteria/states) of a decision problem matter or are even recognized as meaningful. The representation of the agent's attitudes in "Two Transparent Boxes" described earlier, whereby on any candidate expected utility function, Box 2 is better for one person (criterion) by a greater amount than it is worse for the other person (criterion) denies any special primacy to criteria taken in isolation. The idea is that it is simply a further fact about the agent's attitudes that Box 2 is better than Box 1 overall, despite the two boxes being incommensurable with respect to each of the criteria taken in isolation.

Compare the situation with "Two Opaque Boxes." Some argue that here too, there may be facts about the agent's attitudes, to do with "complementarities" in their expected utilities across states, that cannot necessarily be discerned by looking at the states in isolation. Jim Joyce (2010) emphasizes this point in discussing agents who are intuitively best modeled as having "imprecise probabilities." Certain decision problems strongly suggest that the agent considers a number of possible probability functions as live candidates and that complementarities across states, or patterns that are true of all these candidate probability functions, are real and important.[7] But at least in the case of states, it is plausible to argue for the contrary position: that one's attitudes with respect to states taken in isolation are basic, these being the possibilities for one's fully informed self, and that one's overall attitudes across states are derived from one's attitudes toward the states in isolation. In other words, it is plausible to deny any significance to the complementarities or global features shared by all the candidate probability and utility function pairs that may play a role in modeling one's preference attitudes across states. It follows that there is reason to think weak (state-wise) competitiveness is a constraint on rational preference. There are no such reasons favoring weak (criterion-wise) competitiveness.

5 Return to Ordinary Decision-Making Practices

A case has thus been made against choice principles based on a non-dominance relation for the multi-criteria decision setting that extends to incommensurability. The principles in question were dubbed weak/strong (criterion-wise) competitiveness. Any stronger principle – such as a version of independence that extends to incommensurability – is thus also undermined by the analysis here. Recall that the rough idea of the stronger principle is "ignore those aspects of options whereby they are incommensurable, and base one's overall preferences on the remaining criteria." This is a poor guide to choice deliberations. We have seen that an option cannot be ruled out of contention just because it is dis-preferred on all those criteria for which it is not incommensurable with another option.

Moreover, an option that is preferred or at least not dis-preferred on all those criteria for which it is not incommensurable with another option may nonetheless be dis-preferred overall.

Are our ordinary choice deliberations then routinely unreasonable? For all that was said in Section 1, the answer is apparently "yes." The *Bike Route* decision problem was introduced to motivate the idea that we do indeed routinely ignore those aspects of options or criteria for which we cannot compare options. And this decision problem is as mundane as any.

There is, however, a more charitable reading of our ordinary choice deliberations in light of the findings here. It is to insist that appearances can be deceiving. To the extent that we do "ignore incommensurabilities" in assessing options, it is because the incommensurabilities in question are relatively "minor"; they do not "swamp" those aspects of the options for which the agent has clear preferences (as per the decision problem in Table 11.2). In other cases, we may not be ignoring incommensurabilities at all but rather doing something which, again, is typically regarded as faultless (and in accordance with the model of multi-criteria decision-making presupposed in this discussion): we may simply be ignoring criteria for which we are *indifferent* between the options.

The more charitable reading raises further questions, however, about the reasonableness of ordinary choice deliberations. The deliberations as described may well be *consistent*, in the sense of according with the appropriate preference axioms. But one might doubt whether it is reasonable in a more substantive sense to be (more or less) indifferent between options with respect to criteria that are less salient, say, criteria concerning the options' more complicated causally downstream effects.[8] For instance, one might doubt whether it is reasonable to be indifferent between the city bike route and the green bike route when it comes to complicated downstream effects like one's impact on the people one will encounter along the route and how these people will in turn impact on those they encounter and so on.

Here the discussion prompted by James Lenman's worry that we are "clueless" about the consequences of our actions is helpful. Hilary Greaves (2016) offers a limited defense of our being indifferent between pairs of options with respect to their more complicated causally downstream effects. The defense rests on a specific application of "the principle of indifference": roughly that in many cases, there is no reason to think that some coincidental downstream effect of one option will not be equally likely to be a coincidental downstream effect of the other option instead.

The question is whether an appeal of this sort to the principle of indifference is typically apt. Greaves (2016) herself outlines circumstances in which it is not: cases where conflicting considerations bear on how options are ranked with respect to the more difficult choice criteria concerning the options' causally downstream effects. Perhaps Laurie Paul has something like this in mind with respect to decisions involving transformative

experience: for many choice criteria associated with life satisfaction, say, there will be unresolved conflicting considerations regarding how the options should be ranked. In such cases, the reasonable attitude is arguably one of incommensurability between the options under the criteria in question. But note that there seems no reason for this incommensurability, in general, to be relatively minor, such that it can be safely ignored. Quite the opposite. The incommensurability here seems significant, and if the choice criteria in question are also significant, then the incommensurability will have a "swamping" effect. That is, to the extent that our decision problems are best described in the way just outlined, then we must live with widespread indeterminacy or incommensurability in our overall ranking of options.

Notes

1 The reason for this incompleteness of preference or lack of opinion would presumably be that the criterion in question is itself multi-dimensional. For example, the comfort of other riders may depend on both companionship and ease of riding; these may pull in opposite directions, and it may be unclear how they trade-off against each other.
2 Compare with Bykvist and Stefánsson's (2017) similar interpretation of Paul's proposal for making decisions involving transformative experience. (They go on to critique the proposal.)
3 Note that Rabinowicz (2021) rejects the state-wise competitiveness principles stated here (in line with Hare), but both regard the principles to be at least plausible, if ultimately incorrect.
4 Typically, sets of expected utility representations are thought to be convex, but the appeal to discrete, non-convex sets does not compromise the analysis here.
5 See, for instance, Keeney and Raiffa (1976) for a representation theorem supporting this choice rule, at least when the preference orderings concerned are complete. In our examples, some preference orderings for criteria are assumed incomplete. But Keeney and Raiffa's theorem may nonetheless be taken to provide support for the weighted average choice rule, even a modified version that yields sets of weighted averages. And we can suppose that there is independent support for the representation of the preferences under each criterion as given in the tables.
6 Several reasons are given against endorsing weak (state-wise) competitiveness. Hare (2010) notes that there is a consideration in favor of Box 2, the dollar, but none in favor of Box 1. Moreover, Hare appeals to an argument from transitivity that is made explicit by Rabinowicz (2021): if one switches the prizes for Box 2 between the columns in Table 11.5, this option dominates and is thus better than Box 1. Surely one should be indifferent between ordinary Box 2 (as it appears in Table 11.5) and the switched Box 2. But then, by transitivity, ordinary Box 2 is better than Box 1. To deny this, and thus maintain that Box 1 is not worse than ordinary Box 2 and thus a permissible choice, one would have to resist the indifference of the ordinary and switched Box 2 or else admit a violation of transitivity.
7 The term "complementarity" used here is indeed taken from Joyce, although he uses it in a narrower sense and with respect to candidate probability functions.

8 The phrase "(more or less) indifferent" is intended to account for cases of "minor incommensurabilty" (as discussed earlier and illustrated in Table 11.2), as well as cases of genuine indifference.

References

Bales, A., Cohen, D. and Handfield, T. (2014), 'Decision Theory for Agents with Incomplete Preferences', *Australasian Journal of Philosophy* 92/3: 453–470.

Bykvist, K. and Stefánsson, H.O. (2017), 'Epistemic Transformation and Rational Choice', *Economics and Philosophy* 33: 125–138.

Greaves, H. (2016), 'Cluelessness', *Proceedings of the Aristotelian Society*, 116/3: 311–339.

Hare, C. (2010), 'Take the Sugar', *Analysis* 70/2: 237–247.

Joyce, J.M. (2010), 'A Defense of Imprecise Credences in Inference and Decision Making', *Philosophical Perspectives* 24: 281–323.

Keeney, R.L. and Raiffa, H. (1976), *Decisions with Multiple Objectives: Preferences and Value Tradeoffs* (Wiley).

Lenman, J. (2000), 'Consequentialism and Cluelessness', *Philosophy & Public Affairs* 29/4: 342–370.

Paul, L.A. (2014), *Transformative Experience* (Oxford University Press).

Rabinowicz, W. (2021), 'Incommensurability Meets Risk', in H. Andersson and A. Herlitz (eds.), *Value Incommensurability: Ethics, Risk and Decision-Making* (Routledge).

12 Hard Choices Made Harder

Ryan Doody

1 Introduction

Suppose you are deciding between law school and pursuing a PhD in philosophy. This is a hard choice. The life of a philosophy professor – which is what you hope would result from pursuing the degree in philosophy – and the life of a successful lawyer – which is what you hope would result from going to law school – manifest significantly different qualities of what you value in a career. Becoming a successful lawyer promises greater financial security, holidays in Majorca, and a revered social status. Becoming a philosopher, however, promises greater intellectual satisfaction, lazy mornings, and a great deal of professional autonomy. When it comes to these two careers, you're *ambivalent*: you don't prefer one to the other, nor are you indifferent between the two.[1] It's hard to choose when you're ambivalent.

But this choice is even harder still. For this is also a choice under uncertainty. Merely choosing to *pursue* a particular career doesn't guarantee success. Going to law school *might* result in you becoming a successful lawyer. But it might not. Instead, it might result in you flunking out and being buried beneath a mountain of student loans. Similarly, pursuing a PhD in philosophy might result in becoming a professor of philosophy, but this is far from guaranteed. Instead, it might result in a very interesting dissertation on the metaphysics of absences and very troubling absences of tenure-track positions in metaphysics. Both options are a *gamble*: how things shake out turn on features of the world you're uncertain about.

When you are uncertain about what will result from your choices, the standard advice is to maximize expected utility (i.e., for each of your options, you should weigh the value of its possible outcomes by your credence in that outcome being the one that would actually result were you to perform it; the weighted average of these values is an option's *expected value*; the option with the highest expected value is the one you ought to perform). But if you're ambivalent between the potential outcomes of your options – as you are in this example – this advice is not helpful. Because

you are ambivalent between being a lawyer and being a philosopher, your preferences cannot be represented with a utility-function, and, thus, there is no well-defined quantity whose expectation you can maximize.

What can be said instead? Are there any general principles – or, better still, a full-fledged decision theory – to guide us when facing decisions of this kind? If ambivalence and uncertainty are ubiquitous, which I contend it's plausible to think they are, this is a pressing question.[2]

To the end of addressing this question, I develop an impossibility result: there are a handful of independently plausible constraints that no such decision theory can jointly satisfy.

2 The Puzzle of Opaque Sweetening

Let's begin by looking at Caspar Hare's puzzle of opaque sweetening (Hare 2010, 2013), which itself underlies an impossibility result.

> **Opaque Sweetening**. There are two opaque boxes: a Larger box (L) and a Regular box (R). A fair coin has been tossed. If the coin landed heads, then a voucher for an all-expenses-paid Alpine ski vacation (A) was placed in the Larger box and a voucher for an all-expenses-paid beach vacation (B) was placed in the Regular box; if the coin landed tails, then B was placed in the Larger box and A was placed in the Regular box. In either case, you don't know which prize is in which box (See Figure 12.1).

$$\textbf{Larger box} = \begin{cases} A & \text{if Heads} \\ B & \text{if Tails.} \end{cases} \qquad \textbf{Regular box} = \begin{cases} B & \text{if Heads} \\ A & \text{if Tails.} \end{cases}$$

Figure 12.1 Two vacations in opaque boxes, distributed by coin-flip.

> Now imagine that \$20 is added to the Larger box. If you choose the Larger box, you will win whichever prize it contains plus \$20. Nothing is added to the Regular box. You are asked to choose one of the two boxes, taking home whichever prize is in the box you choose (See Figure 12.2).

	Heads	Tails
L^+	$A^{+\$20}$	$B^{+\$20}$
R	B	A

Figure 12.2 An example of Opaque Sweetening. The Larger box has been sweetened with \$20.

The puzzle exploits a characteristic feature of being ambivalent – namely, that it, unlike indifference, is *insensitive to mild sweetening*: if you are ambivalent between X and Y, you should likewise be ambivalent

between X and Y⁺, a slightly improved version of Y. Let's suppose that you are ambivalent between the adventurous skiing trip in the Alps and the relaxing trip to the beach. There are various good things to be said in favor of each. You don't prefer either one to the other. Moreover, you don't prefer the beach vacation *plus 20 bucks* (B⁺$²⁰) to the ski vacation, and you don't prefer the ski vacation *plus 20 bucks* (A⁺$²⁰) to the beach vacation. Small improvements aren't enough to resolve your ambivalence.

Is there a box you're rationally required to take? Or is it rationally permissible to take either? The puzzle brings out a tension between two attractive ideas. One of these ideas supports the verdict that one is rationally required to take, or prefer, L^+ over R. And the other supports the verdict that it is rationally permissible for one to take either.

The first idea is that you ought to evaluate an option in terms of its corresponding *prospects*: the probability distribution over its potential outcomes. In this case, L^+'s corresponding prospects (a 50% chance of A⁺$²⁰ and a 50% chance of B⁺$²⁰) are better than R's (a 50% chance of A and a 50% chance of B); and so you ought to prefer, and thus take, L^+ over R.[3]

The other idea is that rationality shouldn't require you to prefer one option to another if there is no way for the one to be better than the other. In this case, even though you don't know which box contains which prize, you know that you don't prefer what's in L^+ to what's in R. If the coin landed heads, then L^+ contains A⁺$²⁰ and R contains B, and you don't prefer A⁺$²⁰ to B; if the coin landed tails, then L^+ contains B⁺$²⁰ and R contains A, and you don't prefer B⁺$²⁰ to A. No matter how the coin landed, you're ambivalent between the prizes in each box. There is no way for L^+ to result in an outcome you prefer to what you would get by choosing R instead. And – so the thought goes – you aren't rationally required to prefer, or take, L^+ over R.

This conflict can be presented as an impossibility result: no decision theory, which allows for ambivalence, can satisfy the following four constraints.

Neutrality
If φ and ψ are probability-preserving permutations of each other, you are rationally required to be indifferent between the two.

Two options are *probability-preserving permutations* of each other just in case every outcome that might result from the one is just as likely to result from the other. Figure 12.3 serves as an example.

	Heads	Tails
L	A	B
R	B	A

Figure 12.3 An example illustrating *Neutrality*. L and R are probability-preserving permutations of each other.

Given that a fair coin is equally likely to land heads as it is to land tails, *L* and *R* are probability-preserving permutations of each other. Every outcome that might result from *L* – namely, A or B – is equally likely to result from *R*. According to *Neutrality*, then, you are rationally required to be indifferent between *L* and *R*.[4]

The motivation for *Neutrality* is that, when evaluating an option, all that should matter is how good its outcomes are and how likely those outcomes are to result. The identity of the state in which the outcome occurs is only relevant insofar as it affects the outcome's value or probability of obtaining. Outcomes are to be individuated in such a way that the states in which they occur don't affect their value. So, options that are probability-preserving permutations of each other are the same in every respect that's relevant to rational choice.

The next constraint is a familiar one. It relies on the notion of *strict dominance*. Let's say that an option φ *strictly dominates* an option ψ just in case, for every state S that might obtain, you prefer φ's to ψ's outcome in S.[5] Informally, we can say that one option strictly dominates another just in case it always does better. (Although we don't need it now, the notion of *weak dominance* will play a role later on. One option *weakly dominates* another just in case, for each state, the former's outcome is at least as good as the latter's, and there is some state in which the former's outcome is better.[6] Informally, we can say that one option weakly dominates another just in case it's always at least as good and sometimes better.)

Strict Dominance Principle
If φ strictly dominates ψ, you're rationally required to prefer φ to ψ.

The motivation for the *Strict Dominance Principle* is that, if φ strictly dominates ψ, φ is guaranteed to be better than ψ. Although you might not know which outcome would result from performing φ, you do know that whichever it is, you prefer it to the outcome that would result from performing ψ instead. You know, then, that φ will net you more value than ψ would. Therefore, you should prefer it. Figure 12.4 is as an example.

	Heads	Tails
L^+	A$^{+\$20}$	B$^{+\$20}$
L	A	B

Figure 12.4 An example of Dominance.

L^+ strictly dominates L. You don't know how the coin has landed, but you do know that, either way, L^+ will net you a better prize than L would. Because L^+ strictly dominates L, you are in a position to know something significant about how the options compare in the actual word: L^+ is better.

Ideally, you'd always prefer the option that promises to be *actually* better. Of course, we don't always know which of our options promises to actually be best – but, when we do know (as we do when one of the options strictly dominates the others), it would be irrational not to prefer it.[7]

The third constraint is also a familiar one: *transitivity*. Because we're concerned with a number of different ordering-relations – e.g., strict preference, indifference, ambivalence – there are a number of different corresponding transitivity claims.

The first concerns only strict preference:

> **Transitivity of Preference**
> If you prefer X to Y and you prefer Y to Z, then you should prefer X to Z.

The next concerns only indifference:

> **Transitivity of Indifference**
> If you're indifferent between X and Y and you're indifferent between Y and Z, then you should be indifferent between X and Z.

And then there are versions of transitivity that mix preference and indifference, the following of which most concerns us here:

> **Sensitivity to Sweetening**
> If you prefer X to Y and are indifferent between Y and Z, then you should prefer X to Z.

These different claims admit different possible motivations. The most famous motivation for *Transitivity of Preference*, for example, is the money pump argument: if you violate it, it's possible to construct a series of trades, each of which (if you act on your preferences) it is rational for you to take but which collectively guarantee a sure-thing loss. Alternatively, you might take transitivity to be partially *constitutive* of the meaning of these notions (Davidson 1976). For example, part of what it *means* to be indifferent between Y and Z is that, if you prefer X to Y, you should then prefer X to Z. This is how I propose we interpret *Sensitivity to Sweetening*.[8]

The final constraint has the flavor of a dominance principle. It relies on the notion of one option being *never worse* than another. Let's say that an option φ is *never worse* than an option ψ just in case, for every state S that might obtain, you don't prefer ψ's outcome to φ's in S.[9]

> **Never Worse Principle**
> If φ is never worse than ψ, then you are not rationally required to prefer ψ to φ.

The motivation for this principle is somewhat similar to one given for the *Strict Dominance Principle*: if φ is never worse than ψ, then you are in a position to know that you don't prefer ψ's actual outcome to φ's – and if you know, as a matter of actual fact, that ψ isn't any better than φ, you aren't rationally required to prefer it.[10] The case of opaque sweetening, which opened this section (see Figure 12.2), serves to illustrate this constraint. In that example, because you don't prefer A[+$20] to B and you don't prefer B[+$20] to A, *R* is never worse than *L*. And so, according to the *Never Worse Principle*, you aren't rationally required to prefer *L⁺* to *R*.

These four constraints – *Neutrality*, the *Strict Dominance Principle*, *Sensitivity to Sweetening*, and the *Never Worse Principle* – are jointly inconsistent, as Figure 12.5 makes clear:

	Heads	Tails
L⁺	A[+$20]	B[+$20]
L	A	B
R	B	A

Figure 12.5 An example illustrating that *Neutrality*, the *Strict Dominance Principle*, *Sensitivity to Sweetening*, and the *Never Worse Principle* are jointly inconsistent.

Given *Neutrality*, you're rationally required to be indifferent between *L* and *R*. From the *Strict Dominance Principle*, you're rationally required to prefer *L⁺* to *L*. If you adopt the attitudes that these principles rationally require you to adopt, then, from *Sensitivity to Sweetening*, you're also rationally required to prefer *L⁺* to *R*, which violates the *Never Worse Principle*. Therefore, there can be no decision theory that both allows for ambivalence and satisfies all four of these constraints.

3 Neutrality and the Never Worse Principle

The puzzle of Opaque Sweetening has sparked a debate, which divides roughly into two camps: those who support (something like) *Neutrality* and those who support (something like) the *Never Worse Principle*.[11] This makes the impossibility result presented above of only limited interest. Both *Neutrality* and the *Never Worse Principle* are – as I will soon argue, needlessly – contentious. The most compelling motivation for the one tells against the truth of the other (and vice versa), so the two aren't really independently plausible.

Consider *Neutrality*, for example. You should only find this principle appealing if you think that the role *states* play in determining what it's rational to do is a derivative one. The states matter only insofar as they serve to determine evaluatively relevant features of the potential outcomes

and those outcomes' probabilities; once the outcomes and their probabilities have been determined, the facts about which outcome would be realized in which state is irrelevant. If two options have the same potential outcomes with the same probabilities, the reasons you have to favor the one perfectly balance the reasons you have to favor the other, and so you ought to be indifferent between them.

But it's precisely this view about states that proponents of the *Never Worse Principle* should deny. On these views, the states matter (over and above their effect on the outcomes and their probabilities) because, ideally, your attitudes should match the value-relations that actually hold between your options – and the facts about which outcomes will be realized in which states are relevant to whether you're in a position to know which value-relation obtains in the actual world. If you know, for example, that φ's outcome is better than ψ's in the actual world, you should prefer φ to ψ; if you know that you're indifferent between φ's and ψ's outcome in the actual world, you should be indifferent between φ and ψ; and if you know – as you do with *L* and *R* – that you are ambivalent between your options' outcomes in the actual world, you should be *ambivalent*, not indifferent, between those options. Permuting states (even if they are equiprobable) might affect what you're in a position to know about how your options compare in the actual world. But, on these views, such knowledge plays an important (perhaps even *decisive*) role in determining what it's rational to do.

Just as the proponents of the *Never Worse Principle* have reason to reject *Neutrality*, the proponents of *Neutrality* have reason to reject the *Never Worse Principle*. They can claim that, while it might often be true that you're not required to prefer one option to another if you know that the latter is guaranteed to not be worse than the latter, this isn't always the case. In many cases, φ is never worse than ψ because φ's outcomes are at least as good as ψ's in each state – in which case, you can't have more reason to ψ than to φ. But, in the example depicted in Figure 12.2, *R* is never worse than *L*⁺ (and *L*⁺ is never worse than *R*), not because *R*'s outcomes are at least as good as *L*⁺'s but because you're ambivalent between their outcomes – and that's compatible with having more reason to take *L*⁺ than *R*.[12] In a similar vein, they could claim that what matters in evaluating an option is not merely how good that option is relative to how things *would've* been had you chosen otherwise but also how good things *could've* been. Were the former all that mattered, then perhaps the *Never Worse Principle* would be correct. But, because *L*⁺ guarantees something better than what you just as well could get from *R*, you are – *contra* the *Never Worse Principle* – required to prefer *L*⁺ to *R*.[13]

Both *Neutrality* and the *Never Worse Principle* are contentious. In the next section, I will show that there exists an impossibility result that follows from less controversial constraints.

4 Less Controversial Constraints

In this section, I'll develop a different impossibility result, which replaces *Neutrality* and the *Never Worse Principle* with less controversial constraints. Recall that *Neutrality* was controversial because it abstracts away from the facts about which outcomes occupy which states. Its critics contend that these facts are important because they're relevant to what you might be in a position to know about how your options compare in the actual world. In short, *Neutrality* effectively allows for cross-state comparisons – which, its critics contend, are illegitimate (at least when there's ambivalence). Also, recall that the *Never Worse Principle* was controversial because the fact that one option is never worse than another is (at least relative to one conception of reasons) nevertheless compatible with having no reason to take the former and some reason to take the latter – as illustrated by the fact that φ can be both never worse than ψ and yet stochastically dominated by it. This is, in part, because the *never worse than* relation isn't asymmetric: it is possible for two options, like L^+ and R, to both be never worse than the other. Thus, that, e.g., R is never worse than L^+ doesn't provide you with a reason to prefer R – or, at least, not one that isn't perfectly counterbalanced by an analogous reason in the other direction. I'll introduce some constraints that avoid these criticisms.

The first is a weakening of the *Never Worse Principle*. Roughly, it says that if one option is never worse than another *and it is likely to be better*, then you are not rationally required to disprefer it. An option, φ, is *likely to be better* than another, ψ, just in case the probability you assign to the disjunction of the states in which you prefer φ's outcome to ψ's is greater than 50%.[14]

> **Never Worse, Likely Better Principle**
> If φ is never worse than ψ and is likely to be better, then you are not rationally required to prefer ψ to φ.

Notice that this principle, unlike the *Never Worse Principle*, makes use of a notion that is asymmetric: if φ is never worse and likely to be better than ψ, then ψ *isn't* never worse and likely to be better than φ. So, if φ is never worse and likely to be better than ψ, you do have a reason – a fairly powerful one, in fact – for taking φ over ψ (and one which isn't perfectly counterbalanced by an analogous reason in the other direction). Given this powerful consideration in φ's favor, it's hard to see how you could nevertheless be rationally required to prefer ψ instead. The *Never Worse, Likely Better Principle* says that you're not. This principle is harder to deny than its cousin, the *Never Worse Principle*.

Figure 12.6 provides an example of the principle in action. Suppose you are deciding between L^+ and R. Before you make your selection, though,

the situation changes. Rather than facing a choice between the Larger box plus 20 bucks (L^+) and the Regular box (R), you are told that, either, you can pay a small fee ($2.50, say) to acquire L^+ outright ($L^±$) or you can opt for a mixed option (M). A fair six-sided die has been rolled. If it landed on 1, you'll receive R; if it landed on anything but 1, you'll receive L^+. Out of these two new options – $L^±$ and M – which should you choose?

Coin:	Heads		Tails	
Die:	1	2–6	1	2–6
$L^±$	A$^{+\$17.50}$	A$^{+\$17.50}$	B$^{+\$17.50}$	B$^{+\$17.50}$
M	B	A$^{+\$20}$	A	B$^{+\$20}$

Figure 12.6 An example illustrating the *Never Worse, Likely Better Principle*.

According to the *Never Worse, Likely Better Principle*, you are not rationally required to prefer paying the $2.50 for L^+ over letting the die decide. M, the mixed option, is five-sixths likely to result in L^+ at no cost, which, no matter which prize it contains, is preferable to paying for it. Furthermore, in the unlikely event that M doesn't result in L^+, it nevertheless promises a prize that isn't worse than the one you'd get by paying to have L^+ outright. There's certainly something to be said in M's favor. And so any view which holds – *contra* the *Never Worse, Likely Better Principle* – that one is rationally required to prefer $L^±$ to M, at the very least owes us an explanation for how that could be.[15]

The next constraint is best introduced with an example (see Figure 12.7). Suppose that you're in possession of R – the box that contains B if the coin landed heads and A if it landed tails – and have yet to open it. Before you do, you are offered a deal: you can either accept a dollar or you can be entered into a drawing for a large sum of money. Which should you choose: the dollar or the drawing? Suppose you do some back-of-the-envelope math and discover that you are completely indifferent between the sure-thing dollar and the drawing; even though the drawing might result in a larger sum, the chance of winning it is too small for it to come out on top. Because this choice has no effect on which prize is in the box, the fact that you've yet to open the box is irrelevant to whether you should choose the dollar or the drawing; if you'd be indifferent between the two *without* R, you should be indifferent between the two *with* R as well.[16]

	($1-p$)		(p)	
Drawing:	Loser		Winner	
Coin:	Heads	Tails	Heads	Tails
$R^{+\$1}$	B$^{+\$1}$	A$^{+\$1}$	B$^{+\$1}$	A$^{+\$1}$
R^{+d}	B	A	B$^{+\$1/p}$	A$^{+\$1/p}$

Figure 12.7 The dollar or the drawing? An example illustrating *Orthogonal Equipoise*.

Notice that, in comparing $R^{+\$1}$ to R^{+d}, you are in a position to know that one of the two is actually better than the other (although you know not which). Unlike the comparison between L^+ and R, where you know you are ambivalent between their respective outcomes, here, you know you're not; it could be (in the very likely event that your ticket isn't a winner) that $R^{+\$1}$ is (slightly) better than R^{+d}, and it could be (in the very unlikely event that your ticket is a winner) that R^{+d} is (significantly) better than $R^{+\$1}$, but it cannot be that you're ambivalent. This is a manifestation of the fact that, although you're ambivalent between A and B, that ambivalence is orthogonal – and, thus, irrelevant – to your choice between the dollar and the drawing. As it happens, the extent to which $R^{+\$1}$ might be better than R^{+d} (+$1), weighted by the chance that it *is* better $(1-p)$ perfectly balances the extent to which R^{+d} might be better than $R^{+\$1}(+\$1/p - 1)$, weighted by the chance that *it* is better (p). So, because the considerations in favor of the one perfectly balance the considerations in favor of the other, you should be indifferent between the two.

In general, the thought is this. Even if you're ambivalent between some of φ's and ψ's outcomes, so long as you're not ambivalent between any of the outcomes *in the same state*, you should be indifferent between φ and ψ if your estimate of the extent to which φ might be better than ψ equals your estimate of the extent to which ψ might be better than φ.

> **Orthogonal Equipoise**
> If, in every state, there's no ambivalence between φ's and ψ's outcomes, then if the probabilistically weighted average of the extent to which φ's outcome might be better than ψ's equals the probabilistically weighted average of the extent to which ψ's outcome might be better than φs, then you should be indifferent between φ and ψ.

This requires some further spelling out, given that, if you're ambivalent between any outcomes whatsoever, it's not straightforward how to measure the extent to which one outcome might be better than another in a way that admits probabilistic-weighting (even if you aren't ambivalent about those two outcomes specifically). There are several ways of making the notion precise. Here's one way. If your preferences could be represented with a utility-function, u, the notion is clear: the probabilistically weighted average of the extent to which φ's outcome might be better than ψ's is $\sum_{S'} Cr(S') \cdot (u(\varphi \wedge S') - u(\psi \wedge S'))$, for all states S' in which $u(\varphi \wedge S') > u(\psi \wedge S')$. We can generalize the idea in the following way. Even if your preference-ordering cannot, due to ambivalence, be represented with a *single* utility-function, it can be represented with a *set* of utility-functions. Consider all of the *coherent extensions* of your preferences: that is, all of the ways of rendering your preferences complete while holding fixed the preferences you do have.[17] We then represent each of these coherent extensions with a utility-function. Let $\mathscr{U}$ be the set of all

these utility-functions. This set represents your preference-ordering. If the probabilistically weighted averages are the same according to *all* of these utility-functions, then, so long as there's no ambivalence between φ's and ψ's outcomes, you should be indifferent.[18]

Orthogonal Equipoise is not logically weaker than *Neutrality*; the two are logically independent. But it is, I contend, a less controversial principle in virtue of the fact that it eschews making the kind of cross-state comparisons that the opponents of *Neutrality* found objectionable. Instead, it compares the differences in value between the outcomes in the same states (assuming that such comparisons can be made). The constraint doesn't presume that it always will be possible to make these comparisons. But, when you can, your ambivalence is, in some sense, *orthogonal* to the choice at hand. It can, then, for the purposes of making the decision, disappear into the background.

The impossibility result, which is presented in the next section, also requires the *Weak Dominance Principle*, as well as *Transitivity*.

5 The Impossibility Result

In this section, I'll show that the *Never Worse, Likely Better Principle*, *Orthogonal Equipoise*, the *Weak Dominance Principle*, *Transitivity of Preference*, *Transitivity of Indifference*, and *Sensitivity to Sweetening* are jointly inconsistent. I'll demonstrate this in a series of steps.

5.1 Opaque Sweetening, Redux

The first step in the argument is to show that *Orthogonal Equipoise*, along with *Dominance* and *Transitivity*, entail that, in the original Opaque Sweetening example (See Figure 12.2), you are rationally required to prefer L^+ to R. The argument goes like this.[19]

Suppose that in front of you is a button that reads "A." You know the following. A fair coin has been tossed. If you push the button, you win prize A if the coin landed heads and nothing if the coin landed tails. If you don't push the button, you win prize A if the coin landed tails and nothing if it landed heads (See Figure 12.8).

	Heads	Tails
Push "A"	A	o
Don't Push	o	A

Figure 12.8 The "A" Button Example.

Because you know that the coin is fair and because you prefer prize A to nothing, *Orthogonal Equipoise* entails that you should be indifferent between pushing the button and not.

Now suppose that in front of you there is a button that reads "B." Again, you know that a fair coin has been tossed. In this case, if you push the button, you win prize B if the coin landed tails and nothing otherwise, and if you don't push the button, you win prize B if the coin landed heads and nothing otherwise (See Figure 12.9).

	Heads	Tails
Push "B"	0	B
Don't Push	B	0

Figure 12.9 The "B" Button Example.

For analogous reasons, *Orthogonal Equipoise* entails that you should be indifferent between pushing the button and not.

Now suppose that both buttons are in front of you (See Figure 12.10). You can push neither, push one but not the other, or push both. Because, for each button, you are indifferent between pushing it and not, by *Transitivity of Indifference*, you should be indifferent between pushing neither and pushing both.

	Heads	Tails
Push Both	A	B
Push Only "A"	A&B	0
Push Neither	B	A

Figure 12.10 The Both Buttons Example.

Because you know the coin is fair and because (assuming you value the prizes independently) the extent to which you prefer (A & B) to B is exactly the same as the extent to which you prefer A to nothing, *Orthogonal Equipoise* entails that you should be indifferent between pushing neither and pushing only "A." For analogous reasons – this time concerning the extent to which you prefer (A & B) to A and the extent to which you prefer B to nothing – *Orthogonal Equipoise* entails that you should be indifferent between pushing only "A" and pushing both. From *Transitivity of Indifference*, then, it follows that you should be indifferent between pushing neither and pushing both.

Now, suppose we sweeten the option *Push Both* with $20 (See Figure 12.11).

	Heads	Tails
Push Both$^+$	A$^{+\$20}$	B$^{+\$20}$
Push Both	A	B

Figure 12.11 Sweetening the Push Both option with $20.

From the *Strict Dominance Principle*, it follows that you should prefer *Push Both⁺* to *Push Both*. And then from *Sensitivity to Sweetening*, it follows that you are rationally required to prefer *Push Both⁺* to *Push Neither*. Because, structurally, this decision-problem is identical to the one between L^+ and R, this completes the first step of the argument.

For reasons that will become clear soon, notice that there's an analogous argument for the conclusion that you are rationally required to prefer $L^{\pm}$ to $R^{+\$17}$ (See Figure 12.12):

	Heads	Tails
$L^{\pm}$	$A^{+\$17.50}$	$B^{+\$17.50}$
$R^{+\$17}$	$B^{+\$17}$	$A^{+\$17}$

Figure 12.12 Another case of Opaque Sweetening.

All one needs to do is to rerun the argument from above with a $17 bonus added to each vacation package.

5.2 *The Dollar and the Drawing, Redux*

The second step of the argument is to show that *Orthogonal Equipoise* and the *Strict Dominance Principle* entail that you should prefer $R^{+\$17}$ to $R*$ (See Figure 12.13).

Coin:	Heads		Tails	
Die:	1	2–6	1	2–6
$R^{+\$17}$	$B^{+\$17}$	$B^{+\$17}$	$A^{+\$17}$	$A^{+\$17}$
$R*$	B	$B^{+\$20.25}$	A	$A^{+\$20.25}$

Figure 12.13 Dollars vs the Drawing: *R* sweetened with $17 versus sweetened with a five-sixth chance of winning $20.25.

The argument is fairly straightforward. First, let's assume (as we implicitly have been) that you value money linearly, that you value the vacations and the money independently, and that you are risk-neutral. Given these assumptions, you should be indifferent between a lottery that affords you a five-sixth chance of winning $20.25 and a sure-thing $16.875. Imagine, as we did earlier, that you have a choice between the two – but that, also, you are in possession of *R*. As we saw, *Orthogonal Equipoise* entails that, because you are indifferent between the lottery and the cash, you should be indifferent between $R*$ and $R^{+\$16.875}$. But because $R^{+\$17}$ strictly dominates $R^{+\$16.875}$, the *Strict Dominance Principle* says that you are required to prefer it. And, thus, from *Sensitivity to Sweetening*, you are also required to prefer it to $R*$. The extent to which the $R^{+\$17}$ might be better than $R*$ is large enough to counteract the more likely, but

less significant, extent to which the latter might be better than the former. This completes the second step of the argument: you are rationally required to prefer $R^{+\$17}$ to R^*.

5.3 The Drawing and the Roll

The third step is to argue that you are required to prefer the drawing (R^*) to the roll (M). Recall that R^* is just R sweetened with a lottery that has a five-sixth chance of paying out \$20.25 and that M is a mixture of L^+ and R, where, if you take it, you have a five-sixth chance of getting L^+ and a one-sixth chance of getting R. Let's suppose that both R^* and M turn on the same roll of the same die, as represented in Figure 12.14.

Coin:	Heads		Tails	
Die:	1	2–6	1	2–6
R^*	B	$B^{+\$20.25}$	A	$A^{+\$20.25}$
M	B	$A^{+\$20}$	A	$B^{+\$20}$

Figure 12.14 The Drawing vs the Roll: R sweetened with a five-sixth chance of winning \$20.25 versus the mixed option.

The argument here is very similar to the one in the first step (see Figures 12.8–12.11). First, we consider a slightly soured version of R^* that promises a potential windfall of only \$20 rather than \$20.25. Then, as in step one, we "disassemble" that choice into two different decision-problems, like so (see Figure 12.15):

Coin:	Heads		Tails	
Die:	1	2–6	1	2–6
R^*_B	B	$B^{+\$20}$	0	0
M_B	B	0	0	$B^{+\$20}$

Figure 12.15 The Drawing vs the Roll, but with only prize B.

By *Orthogonal Equipoise*, you are required to be indifferent between R^*_B and M_B (see Figure 12.16).

Coin:	Heads		Tails	
Die:	1	2–6	1	2–6
R^*_A	0	0	A	$A^{+\$20}$
M_A	0	$A^{+\$20}$	A	0

Figure 12.16 The Drawing vs the Roll, but with only prize A.

For the same reasons, you are required to be indifferent between R^*_A and M_A.

There are four possible combinations of actions you could take: both M_A & M_B, both R^*_A & R^*_B, M_A & R^*_B, or R^*_A & M_B. From *Orthogonal Equipoise* and *Transitivity of Indifference*, it follows that you are required to be indifferent between (M_A & M_B) and (R^*_A & R^*_B). Then, from the *Weak Dominance Principle*, it follows that you are required to prefer R^* to (R^*_A & R^*_B). From *Sensitivity to Sweetening* (and the fact that (M_A & M_B) is extensionally equivalent to M), you are required to prefer R^* to M.

5.4 The Final Step: Pay or Roll

In the first step, we reached the conclusion that you are required to prefer $L^{\pm}$ to $R^{+\$17}$. In the second step, we reached the conclusion that you are required to prefer $R^{+\$17}$ to R^*. In the third step, we concluded that you are required to prefer R^* to M.

Coin:	Heads		Tails	
Die:	1	2–6	1	2–6
$L^{\pm}$	A	A$^{+\$17.50}$	B$^{+\$17.50}$	B$^{+\$17.50}$
$R^{+\$17}$	B$^{+\$17}$	B$^{+\$17}$	A$^{+\$17}$	A$^{+\$17}$
R^*	B	B$^{+\$20.25}$	A	A$^{+\$20.25}$
M	B	A$^{+\$20}$	A	B$^{+\$20}$

Figure 12.17 The Final Step: Pay or Roll.

From the fact that you're required to prefer $L^{\pm}$ to $R^{+\$17}$ and that you're required to prefer $R^{+\$17}$ to R^*, it follows from *Transitivity of Preference* that you are required to prefer $L^{\pm}$ to R^*. From that fact and that you are required to prefer R^* to M, it follows from another application of *Transitivity of Preference* that you are required to prefer $L^{\pm}$ to M (See Figure 12.17).

But as we saw in the previous section, this violates the *Never Worse, Likely Better Principle*. You can be certain that however the world turns out to be, M's outcome won't be worse than $L^{\pm}$'s. Moreover, if the die rolls anything other than a 1, which is fairly likely to happen, $L^{\pm}$'s outcome will be worse than M's. These are some powerful considerations that speak in favor of M over $L^{\pm}$, which is why the *Never Worse, Likely Better Principle* holds that you are, therefore, not required to prefer $L^{\pm}$ to M. And yet, the argument in this section led us to the conclusion that you *are*. And that's a contradiction.

6 Conclusion

It's hard to choose when you're ambivalent about your choices. It's harder still – or so I hope I've shown – to choose when you are both ambivalent and uncertain. I've argued that a number of independently plausible principles governing rational choice in such situations are jointly inconsistent. I've argued that this result is, in some ways, more serious than the Opaque Sweetening puzzle because it relies on principles that are less contentious.

What should we make of this result? One possible reaction is to give up on developing a decision theory for such situations. If you're ambivalent, you should work to resolve it rather than expect decision theory to provide you with any helpful advice.

That might ultimately be correct, but I think that's to give up too early. I think that, as a matter of fact, we are all ambivalent about something at least some of the time. I'm skeptical that one even realistically could resolve all of one's ambivalence. I'm also skeptical that one should – that it is some kind of character flaw or mark of irrationality to be ambivalent about at least some things at least some of the time. In fact, in some cases, ambivalence might be the correct – or most fitting – attitude to have, especially in the face of what might be actual value indeterminacy out in the world. And so it would be a shame – an embarrassment, even – if formal decision theory had absolutely nothing to offer us.

What are our other options? Give up one (or more) of the offending principles. I've attempted to argue that each of the principles used to derive the contradiction is fairly plausible. But plausibility is not infallibility. I think, in terms of rejecting a principle, the most promising routes are, either, to deny the *Never Worse, Likely Better Principle* or to deny *Transitivity*. Proponents of views, like *Prospectism*, which evaluate options solely in terms of the probability distribution over its outcomes, will need to reject the *Never Worse, Likely Better Principle*. But, in doing so, they owe us an explanation for why it seems right – and it's not clear that the extant responses to the *Never Worse Principle* successfully carry over. On the other hand, proponents of dominance-style principles, like the *Never Worse Principle* and the *Never Worse, Likely Better Principle*, should (I think) reject *Transitivity*. But that, too, of course, comes at a cost – even a potentially non-metaphorical one (in the event that one is transformed into a money pump).

I think more work needs to be done in exploring the terrain surrounding hard choices. I hope this chapter has helped impose some structure on that terrain. In the meantime, hard choices will remain as hard as ever.

Acknowledgements

same institution, the Recent Work in Decision Theory and Epistemology workshop at Columbia University, and the Fourth Annual Meeting of the PPE Society. I'd especially like to thank Chrisoula Andreou, D Black, Luc Bovens, Jennifer Carr, Jamie Dreier, Nir Eyal, Melissa Fusco, Ze'ev Goldschdmidt, Simone Gubler, Caspar Hare, Harvey Lederman, Chris Meacham, Kian Mintz-Woo, Agustín Rayo, Ittay Nissan-Rozen, Bob Stalnaker, Una Stojnić, Christian Tarsney, and likely many others. I'd also like to thank the editors of this volume, Henrik Andersson and Anders Herlitz. Any and all mistakes are my own.

Notes

1 In decision theory, this is known as having *incomplete preferences*. And while the Completeness Axiom is a mainstay of traditional axiomatic decision theory, it is often, since the very beginning, assumed for the sake of mathematical convenience rather than justified as a genuine rational requirement. See, for example, (von Neumann and Morgenstern 1944: 630), (Savage 1954: 21), (Chernoff 1954), and (Aumann 1962).

2 Chang (1997, 2002, 2009, 2017) is a notable example of someone who, like me, thinks the phenomenon is ubiquitous. However, I intend "ambivalence" to refer to a subjective, psychological state. The psychological state is conative, not cognitive: ambivalence might correspond to, but is not identical with, a belief about the objective value-relations that hold between the items. Chang's "parity," on the other hand, is a (non-transitive) fourth value-relation; in addition to "better than," "worse than," and "as good as," two comparable alternatives might be "on a par." Discussions of related phenomena can be found in (Griffin 1986: 80-98), (Parfit 1984: 431), and (Broome 2007).

3 Although not always argued for explicitly, there are a number of views committed to this underlying idea: e.g., I. J. Good's *Quantizationism* (Good 1952), Isaac Levi's *V-admissibility* (Levi 1986, 2008), Amartya Sen's *Intersection Maximization* (Sen 2004), and (Weirich 2004). The view is also popular among economists: e.g., (Dubra et al. 2004; Evren and Ok 2011; Galaabaatar and Karni 2013; Ok et al. 2012). I'll treat Hare's *Prospectism* as a representative of this family of views (Hare 2010, 2013). These views, roughly, represent one's preferences with the set of utility-functions that correspond to all of the coherent extensions of one's preference-ordering such that you count as preferring X to Y just in case every function in the set ranks X ahead of Y, you count as being indifferent between X and Y just in case every function in the set ranks them that way, and you lack a preference between the two when there's disagreement among the functions in the set.

4 It might be objected that really the two options have different outcomes because, e.g., in the case of *L*, there is A-when-the-coin-has-landed-heads, whereas, for *R*, there is the outcome A-when-the-coin-has-landed-tails. However, the typical convention in decision theory, which I will be adhering to here, is that outcomes are individuated finely enough to encompass all of the features that matter to the chooser. If the flip of the coin mattered in this way, it would be a mistake to represent, e.g., the outcomes (L ∧ Heads) and (R ∧ Tails) with the same symbol. The fact that we are representing these outcomes with the same symbols should be taken to imply that Heads and Tails are (to borrow a phrase from (Ramsey, 1931, p.177)) *ethically neutral*: you

are indifferent between (L ∧ Heads), A, and (R ∧ Tails), and you are indifferent between (L ∧ Tails), B, and (R ∧ Heads).

5 In symbols: φ *strictly dominates* ψ just in case, for all states S, $(\varphi \wedge S) \succ (\psi \wedge S)$.

6 In symbols: φ *weakly dominates* ψ just in case, for all states S, $(\varphi \wedge S) \succeq (\psi \wedge S)$, and there is some state S^* such that $(\varphi \wedge S^*) \succ (\psi \wedge S^*)$.

7 A quick note of caution. The *Strict Dominance Principle*, like all statewise dominance principles, risks delivering incorrect verdicts when the probability of a state depends on which option is performed (Nozick 1969). It's controversial what notion of dependence matters here – a causal or evidential one. Let's hereby avoid the issue by understanding the dominance principles to be restricted to cases in which the states are both causally and evidentially independent of one's options.

8 I'll mention that the analogous claim that replaces "indifference" with "ambivalence" – that is, the one that says "if you prefer X to Y and are ambivalent between Y and Z, then you should prefer X to Z" – is exactly what was being denied earlier when we said that ambivalence is insensitive to mild sweetening. There might be cases in which, because your preference for X over Y is so strong, you should prefer X to Z. But this principle does not hold in general – and certainly doesn't hold when your preference for X over Y is slight.

9 In symbols: φ is *never worse* than ψ just in case, for all state S, $(\varphi \wedge S) \nprec (\psi \wedge S)$.

10 In the service of defending a similar principle, which they call *Competitiveness*, Bales et al. (2014) explicitly appeal to an analogy with weak dominance. Schoenfield (2014), also defending a similar principle (which she calls LINK because it links facts about expected value to what one knows about value), worries that views that violate it are "imposing requirements that transcend what we actually care about: the achievement of value." See (Doody 2019b) for discussion, and critical assessment, of these arguments.

11 In the former camp, we have: Hare (2010, 2013), who first introduced the puzzle and who (cautiously) endorses a principle (called "Prospects Determine Permissibility") that entails *Neutrality*; Rabinowicz (2021), who argues for a very similar principle (which he calls "Permutation (P)"); and Bader (2018). In the latter camp, there is Bales et al. (2014), who defend a principle (which they call "Competitiveness") that is similar in spirit to the *Never Worse Principle* and Schoenfield (2014), whose previously mentioned "LINK" principle provides a metaethical motivation for our own.

12 Hare (2013: 51) argues along these lines. On his view, you should choose L^+ over R because you have no reason to choose R over L^+ (anything that can be said in favor of doing the former can equally well be said in favor of doing the latter) but you do have a reason to choose L^+ over R (you're sure to get 20 bucks). The fact that R is never worse than L^+ doesn't give you a reason to choose it (given that L^+ is also never worse than R), nor does it remove the reason you have to choose L^+ over R. On this view, rationality is about doing what you have the most reason to do. And, in this case, you have more reason to choose L^+ than you do to choose R.

13 Bader (2018: 505) argues along these lines. In his view, you should prefer L^+ to R because the former stochastically dominates the latter. One option *stochastically dominates* another just in case, for every outcome, the probability of getting something at least as good is at least a great for the former as it is for the latter, and there is some outcome such that the probability of getting something at least as good as it is greater for the former than the latter. The probability of getting something at least as good as A is the same for both L^+ and R: 50%. Likewise for B. However, the probability of getting something at least as great as $A^{+\$20}$ is 50% for L^+ but 0% for R. Likewise for $B^{+\$20}$.

14 In symbols: φ is *likely to be better* than ψ just in case $\sum_S \mathrm{Cr}(S)\cdot\mathbb{1}[(\varphi\wedge S)\succ(\psi\wedge S)]>.5$, where $\mathbb{1}[p]$ is an indicator function that returns *1* when *p* is true and returns *0* when *p* is false. In a similar fashion, we could define notions like "very likely to be better" by replacing ".5" with a higher threshold.

15 For further discussion of this example, see Doody (2019b), who presents it as a counterexample to views like *Prospectism* (Hare 2010, 2013), that recommend evaluating options solely in terms of their prospects.

16 The reasoning here presupposes that you value the prizes in the outcomes *independently* of each other: that having the one does not add or subtract from the value of having the other. If that assumption doesn't hold, substitute in prizes for which it does.

17 In symbols: $\succeq^+$ is a *coherent extension* of your preference-ordering if, and only if, (i) for all X and Y, X $\succeq^+$ Y if you prefer X to Y or are indifferent between them, and (ii) $\succeq^+$ is complete.

18 This strategy makes the principle somewhat similar to Hare's *Prospectism* (Hare 2010, 2013). But because the principle is restricted to cases in which there is no statewise ambivalence, *Orthogonal Equipoise* is significantly weaker. Another way of generalizing the notion takes its inspiration from Hare's *Deferentialism* (Hare 2010). The details are more complicated, but the rough idea is this: for each state *S*, we choose a representative utility-function from $\mathcal{U}$ (subject to certain constraints) and perform the calculations using these "state-dependent" utilities. *Deferentialism* is meant to generalize the intuition underlying the *Never Worse Principle* into a full-fledged decision theory, so if the view also entails *Orthogonal Equipoise*, that's further confirmation of that principle's plausibility.

19 This argument is discussed and presented with greater detail in Doody (2019a).

References

Aumann, R. J. (1962), 'Utility Theory Without the Completeness Axiom', *Econometrica* 30/3: 445–462.

Bader, R. (2018), 'Stochastic Dominance and Opaque Sweetening', *Australasian Journal of Philosophy* 96/3: 498–507.

Bales, A, Cohen, D, and Handfield, T. (2014), 'Decision Theory of Agents with Incomplete Preferences', *Australasian Journal of Philosophy* 92/3: 453–470.

Broome, J. (2007), 'Does Rationality Consist in Responding Correctly to Reasons?', *Journal of Moral Philosophy* 4/3:349–374.

Chang, R. (1997), 'Introduction', in R. Chang (ed.), *Incommensurability, Incomparability, and Practical Reason* (1–34). Harvard University Press.

Chang, R. (2002), 'The Possibility of Parity', *Ethics* 112/4: 659–688.

Chang, R (2009), 'Voluntarist Reasons and the Sources of Normativity', in Sobel, D. and Wall, D. (eds.), *Reasons for Action* (243–271). Cambridge University Press.

Chang, R. (2017), 'Hard Choices', *APA Journal of Philosophy* 3/1: 1–21.

Chernoff, H. (1954), 'Rational Selection of Decision Functions', *Econometrica* 22: 422–443

Davidson, D. (1976), 'Hempel on Explaining Action', *Erkenntnis* 10/3: 239–253.

Doody, R. (2019a), 'Opaque Sweetening and Transitivity', *Australasian Journal of Philosophy* 97/3: 559–571.

Doody, R. (2019b), 'Parity, Prospects, and Predominance', *Philosophical Studies* 176/4: 1077–1095.

Dubra, J., Maccheroni, F. and Ok, E. (2004), 'Expected Utility Theory Without the Completeness Axiom', *Journal of Economic Theory* 115/1:118–133.

Evren, O. and Ok, E. (2011), 'On the Multi-utility Representation of Preference Relations', *Journal of Mathematical Economics* 47: 554–563.

Galaabaatar, T. and Karni, E. (2013), 'Subjective Expected Utility Theory with Incomplete Preferences', *Econometrica* 81/1: 255–284.

Good, I.J. (1952), 'Rational Decisions', *Journal of the Royal Statistical Society B* 14: 107–114.

Griffin, J. (1986), *Well-Being: Its Meaning, Measurement and Importance* (Clarendon Press).

Hare, C. (2010), 'Take the Sugar', *Analysis* 70/2: 237–247.

Hare, C. (2013), *The Limits of Kindness* (Oxford University Press).

Levi, I. (1986), *Hard Choices* (Cambridge University Press).

Levi, I. (2008), 'Why Rational Agents Should Not Be Liberal Maximizers', *Canadian Journal of Philosophy* 38(Supplementary Vol 34): 1–17.

Nozick, R. (1969), 'Newcomb's Problem and Two Principles of Choice', in N. Rescher, et al. (ed), *Essays in Honor of Carl G. Hempel* (Reidel).

Ok, E., Ortoleva, P. and Riella, G. (2012), 'Incomplete Preferences Under Uncertainty: Indecisiveness in Beliefs Versus Tastes', *Econometrica* 80/4: 1791–1808.

Derek Parfit. (1984), *Reasons and Persons* (Oxford University Press).

Rabinowicz, W. (2021), 'Incommensurability Meets Risk', in H. Andersson and A. Herlitz (eds.), *Value Incommensurability: Ethics, Risk and Decision-Making* (Routledge).

Ramsey, F.P. (1931), 'Truth and Probability', in R. B. Braithwaite (ed), *The Foundations of Mathematics and Other Logical Essays* (156–198). Routledge.

Savage, L. J. (1954), *The Foundations of Statistics* (Wiley).

Schoenfield, M. (2014), 'Decision Making in the Face of Parity', *Philosophical Perspectives* 28/1: 263–277.

Sen, A. (2004), 'Incompleteness and Reasoned Choice', *Synthese* 140/(1/2): 43–59.

von Neumann, J. and Morgenstern, O. (1944), *Theory of Games and Economic Behavior* (Princeton University Press).

Weirich, P. (2004), *Realistic Decision Theory: Rules for Nonideal Agents in Nonideal Circumstances* (Oxford University Press).

Index

Note: Page numbers followed by 'n' refer to note numbers