Vagueness in Action

Luke Elson

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Philosophy in the College of Arts and Sciences.

Chapel Hill
2014

Approved by:

Geoffrey Sayre-McCord

Simon Blackburn

Ram Neta

Keith Simmons

Walter Sinnott-Armstrong

ABSTRACT

LUKE ELSON: Vagueness in Action.
(Under the direction of Geoffrey Sayre-McCord)

In the papers which compose this dissertation, I defend a distinctive role for evaluative vagueness: (i) I argue that value incommensurability is best understood as vagueness; (ii) I defend an expectation-maximising account of rational action under indeterminacy; and (iii) I argue that Warren Quinn's notorious *Puzzle of the Self-Torturer* is an instance the Paradox of the Sorites.

I argue that when two options are incommensurate (or incomparable) with respect to some value, it is—depending on one's account of vagueness—either unknown or indeterminate which option is better. I criticise Ruth Chang's 'Chaining Argument' for her parity view of incommensurability, and I argue that John Broome's 'collapsing principle' leads to contradictions.

Ignorance doesn't present such a distinctive challenge, but how can we choose under *indeterminacy*? I develop an account of rational choice in such situations. I vindicate the intuition that in some cases of incommensurability, it is permissible to simply flip a coin. I also argue that—as in some other expectation-maximising contexts—sometimes we ought to choose what is not the best: sometimes, under indeterminacy, we ought choose such that there is determinately a better option.

Finally, I argue that Warren Quinn's Puzzle of the Self-Torturer is a disguised instance of incommensurability and does not, as many have argued, undermine orthodox rational choice theory. Self-Torture can be accounted for in the same manner as other kinds of choice under incommensurability/indeterminacy.

For my wife, Katie, for my parents, Laurence and Siân, and for my pets, Bobo and Tarquin.

ACKNOWLEDGMENTS

Many people have offered me help—and support, argument, encouragement, criticism, and patience—in the writing of this this 'multi-paper' dissertation. I reiterate my gratitude to those mentioned at the beginning of each paper, and I offer my apologies to those I have inevitably forgotten to mention.

I have been extremely lucky to count myself a junior member of the UNC intellectual community. Of course my greatest debt is to my chair and advisor, Geoffrey Sayre-McCord. I remain in awe of his facility with philosophical argument, and indebted to him for trusting me with the freedom to pursue my rather inchoate interests. I'm extremely grateful to the rest of my committee: Simon Blackburn, Ram Neta, Keith Simmons (who, along with Jesse Prinz, introduced me to vagueness), and Walter Sinnott-Armstrong. I hope that this dissertation is worthy of their support.

Many others have provided invaluable comments on drafts, including Patrick Connolly, Matthew Kotzen, Marc Lange, John Lawless, Daniel Layman, Douglas MacLean, Julia Nefsky, John Roberts, Nathaniel Sharadin, and Susan Wolf. Many philosophers outside UNC have taken the time to meet with me and discuss my (often ill-formed) ideas or work. In this connection, I am particularly grateful to John Broome, Ruth Chang, Joseph Raz, Miriam Schoenfield, and Walter Sinnott-Armstrong.

I remain grateful to Antony Eagle for guiding and supporting me through the graduate-school admissions process, and to Marilyn McCord Adams, for her encouragement at a low point.

Now I turn to a pair of incommensurate goods. For funding, I am extremely grateful to the Philosophy Department and Graduate School, and the Richard Brooke Scholarship, all at UNC. For friendship, I am many-times lucky in this department, but I must mention Patrick Connolly, my first and best friend in Chapel Hill and the United States.

Finally, I owe everything to Katherine Meehan, my wife.

TABLE OF CONTENTS

1 Why not Incommensurability as Vagueness?

Abstract. Two options are said to be 'incommensurate' when neither is better than the other, but they are not equally good, with respect to some value. Typically, we will say that one option is better in some ways, and the other in others, but neither is better 'all things considered'. I argue that when two options are incommensurate, it is vague—that is, indeterminate or unknowable—which is better. Vagueness can explain the formal structure of incommensurability, and this presents an unmet explanatory burden for competing views. I resist arguments by Ruth Chang and Joseph Raz that incommensurability is not vagueness, or not entirely vagueness. This burden-shifting strategy avoids the problems that have afflicted previous arguments for incommensurability as vagueness, including a reliance on controversial premises.[1]

Churchill, in Northern Canada, is one of the best places in the world to see polar bears. But—I'll assume that you do not live in or near Churchill—it is an uncomfortable and expensive trip, not least because the roads do not go that far north. A holiday closer to home might be much cheaper, but less exciting and spectacular. If each is better in some ways but worse in others, which is better *all things considered?*

The holidays are *incommensurate* (or 'incomparable' or 'incommensurable') when there is apparently no positive answer to that question. It seems that neither is better, but they are not precisely equally good. The last claim—that they are not equally good—might

[1]For discussion and comments, I am indebted to Dorit Bar-On, Ruth Chang, Ram Neta, John Roberts, Geoffrey Sayre-McCord, Walter Sinnott-Armstrong, Susan Wolf, audiences in Charleston, Chapel Hill, Eindhoven, and Reading, and an anonymous reviewer.

need some justification. If the trip to Churchill and one to a nearby destination were precisely equally good (as a holiday for you), then reducing the price of the Churchill trip by an arbitrarily small amount would make it clearly better for you than the nearby destination. But this seems absurd.[2]

Though no overall comparison seems to apply, the question was not a silly one. We implicitly compare holidays when we choose and book (even in the present case, the incommensurate two are each likely better than many others, such as holidays that are both expensive and uninteresting). Incommensurability is ubiquitous, and threatens to arise in almost any choice situation involving several virtues or 'dimensions' of goodness. Who is more creative, Mozart or Michelangelo? Which is more impressive, Stonehenge or Salisbury Cathedral? Which is a better career, law or music? We may agree with Joseph Raz that incommensurability is one of the 'universal features of human thought'.[3]

Typically, we'll suppress mention of the value or standard in question, unless required for clarity, and talk in terms of 'better' and 'worse'. Incommensurability often arises when each option is better in some ways, and worse in others. Let's say that such options are *Pareto-mixed*. Pareto-mixing is closely tied to incommensurability, but not sufficient for it. For example, suppose that the Churchill trip was only $5 more expensive than another holiday, but markedly superior in every other way; we should not hesitate to say that the Churchill trip is clearly superior.

But incommensurability is also deeply puzzling. When some a and b are incommensurate, none of the 'trichotomous' comparisons obtains: a is not better than b, b is not better than a, and they are not equally good. So how *do* they compare, if at all?

The most popular view—Trichotomous Incomparabilism—is that determinately, none of the trichotomous comparisons holds:

---

[2] This is a 'Small-Improvement Argument'; this feature is also known as known as 'inertia' and 'insensitivity to sweetening'. See De Sousa (1974), Sinnott-Armstrong (1985), Raz (1985-86), Blackburn (2010) and Schoenfield (2012).

[3] Raz (1985-86), p. 133. The examples are due to Broome (1997), Chang (2002), and Raz (1985-86), respectively.

> **Trichotomous Incomparabilism (rough formulation).** When a and b are incommensurate, it is false that a is better than b, false that b is better than a, and false that a and b are equally good. None of the trichotomous comparisons applies.

In this paper, I'll instead argue that incommensurability is vagueness:

> **Incommensurability as Vagueness (rough formulation).** When a and b are incommensurate, it is vague whether a is better than b, or b is better than a, or a and b are equally good. One of the trichotomous comparisons applies, but it is indeterminate or unknowable which.

Rather than a *sui generis* phenomenon affecting evaluative comparisons, incommensurability is simply an evaluative manifestation of the everyday phenomenon of vagueness.

This view has been defended by John Broome, but his argument is unsound: it relies on a controversial 'Collapsing Principle' which is not widely accepted, and which I have argued elsewhere is false. But the thesis that incommensurability is vagueness is not dependent on the Collapsing Principle; if the view is correct, then it should be possible to argue for it in some other way. The Collapsing Principle imports some implausible consequences—that Broome himself finds 'hard to believe'—into the view that incommensurability is vagueness, so by arguing for the view without appeal to that Principle, I aim to reach a more plausible formulation.[4]

Vagueness offers a complete, parsimonious, and explanatorily superior account of incommensurability, or so I'll argue. It can explain incommensurability in terms of a well-known albeit puzzling phenomenon—vagueness—without resorting to the *sui generis*.

Seeing that incommensurability is vagueness sheds light on issues of rational choice in

[4]See Broome (1997) for his argument and the Collapsing Principle. The quote is from p. 85. See Chapter 3 and the references there for criticism. Carlson (2012) discusses the restrictions that the Collapsing Principle places on incommensurability as vagueness.

the face of incommensurability, since it assimilates that phenomenon to one of rational choice in the face of vagueness: when we choose between incommensurate options, it is vague which option is better. I argue elsewhere that standard theories of rational choice can be extended to cover this case.[5]

In the next section, I present more precisely the thesis that incommensurability is vagueness. I'll then argue that evaluative vagueness is pervasive, and that vagueness engenders incommensurability. This presents an explanatory burden for those who defend other views: vagueness should be the 'default' explanation of incommensurability. I'll then consider several influential arguments that vagueness could not ground some or all incommensurability; I argue that such arguments rest on mischaracterisations of either incommensurability or of vagueness. The vagueness explanation is more parsimonious than that provided by trichotomous incomparabilism, so we would need some reason to prefer the latter; but no such reason seems forthcoming. I'll conclude by sketching some implications of the thesis for rational choice between incommensurate options.

1.1 Incommensurability, and the View to be Defended

But why accept the existence of incommensurability at all? One might instead think that if two options are properly compared with respect to some value, then one of the trichotomous comparisons always determinately holds between those options, relative to that value: that all putative cases of incommensurability simply rest on ignorance.[6]

If my arguments below are successful, then they refute this debunking view, since I'll show that evaluative vagueness generates incommensurability. But this does engender a taxonomic complication: what if we think that incommensurability is vagueness, but also think that vagueness is ignorance? I don't think that this should count as a debunking view, since epistemic accounts of vagueness typically posit rather radical sources of ignorance, such as ignorance of the use-meaning relation. To mark the difference, let's say that debunking incommensurability amounts to claiming that one of the trichotomous

---

[5]See Chapter 4.

[6]Regan (1997) holds such a view, at least about the Moorean good.

comparisons always obtains determinately, but it is unknown which. By contrast, epistemic accounts of vagueness will typically hold that the relevant claims are *unknowable*, not merely unknown.

With that in mind, we can formulate a definition of incommensurability:

> **Incommensurability.** Two objects or options a and b are *incommensurate* with respect to some value V if and only if:
> (i) it is not knowably true that a is Ver than b;
> (ii) it is not knowably true that b is Ver than a;
> (iii) it is not knowably true that a and b are equally V;
> (iv) the comparison is not silly or malformed.

Clause (iv) sets aside cases of 'noncomparability', such as that implied by the question "what's louder, my chair or this paragraph?" This definition is neutral between all non-debunking accounts of incommensurability: compatibly with this definition, the trichotomous comparisons might be determinately false, determinately true but unknowable, or indeterminate. Giving a theory of incommensurability amounts to saying what relation (if any) holds between incommensurate objects.

As I have mentioned, the standard view of incommensurability is that each trichotomous comparison is determinately false:

> **Trichotomous Incomparabilism.** In incommensurate cases, it is determinate that none of the trichotomous comparisons applies. If a and b are incommensurate with respect to V, then it is false that a is Ver than b, false that b is Ver than a, and false that a and b are equally V.

Trichotomous Incomparabilism encompasses several quite different views. Joseph Raz argues that *no* comparison applies (in most cases, as we'll see below). Ruth Chang thinks that the trichotomous comparisons are not exhaustive, and that a fourth comparison—parity—holds at least sometimes.[7] These views have radically differing

[7]See Raz (1985-86) and Chang (2002).

implications for rational choice and the nature of comparison, but for our purposes they agree on a crucial point: in at least some cases, when a and b are incommensurate, it is false that a is better than b, false that b is better than a, and false that a and b are equally good.

The main thesis of this paper is that when a and b are incommensurate, it is vague whether a is better than b, or b is better than a, or they are are equally good. To state this more precisely, I'll say a little about vagueness. Vagueness is associated with two main phenomena: borderline-cases and Sorites-embeddability. In borderline-cases, the predicate in question neither clearly applies nor clearly fails to apply. Consider the classic example of a vague predicate ('is a heap'): amongst piles of sand, there are clear heaps, clear non-heaps, and borderline-cases, which seem to be 'neither a heap nor not a heap'.

This predicate is also vulnerable to the paradox of the sorites, or heap, because the *tolerance* principle 'if n grains of sand form a heap, then $(n-1)$ grains of sand form a heap' seems undeniable.[8] Together with the claim that a million grains form a heap, and repeated applications of modus ponens, this tolerance principle implies that three grains form a heap, even in contexts where that is plainly false. Similarly, the vagueness of 'is bald' allows us to argue that if a man with fifty head hairs is bald, then so is a man with fifty thousand. Borderline-cases may arise in other ways (such as, perhaps, the open future), but tolerance is plausibly the distinctive mark of vagueness. I'll argue that incommensurability rests on borderline-cases, but it is plausible that tolerance can explain the puzzling phenomenon of incommensurate *value-pumping*, which I discuss briefly below.

Theories of vagueness come in two main flavours. On indeterminist views, it is indeterminate how many grains are required for a heap, and it is indeterminate whether a given borderline-heap is a heap. This may be put in terms of 'neither true nor false'. In this camp, metaphysical indeterminists hold that there is a real property of heapiness with an indeterminate extension. Semantic indeterminists have it that vagueness is a

---

[8]For more on tolerance, see Wright (1975).

matter of semantic indecision: perhaps meaning supervenes on use, but our use of 'heap' does not determine a precise extension for that term.[9]

Epistemic views deny the existence of such indeterminacy: 'is a heap' has a precise extension—there is a number k such that k grains determinately do not form a heap, and (k+1) grains determinately do—but this extension is unknowable. Epistemicists characterise borderline-cases in terms of ignorance.[10]

But on all mainstream accounts of vagueness, tolerance principles are false; it is simply indeterminate or unknowable *which instances* of a particular tolerance principle are false. It is false that for all n, if n grains of sand form a heap, then $(n-1)$ grains of sand form a heap; similarly, it is true that there is a precise minimum number of grains of sand required for a heap, and that there is a precise minimum number of hairs required to be non-bald. But these numbers are indeterminate or unknowable. From now on, we'll use terms like 'borderline', 'borderline-case', and 'vague' neutrally between the main accounts of vagueness.

What I'll call 'categorical' borderline-cases are familiar. Some pile of sand is neither a clear heap nor a clear non-heap. A 15 year-old is neither a clear child nor a clear non-child. And so on. In such cases, we have a predicate F ('is a child'), and an object a (some 15 year-old) such that the question 'is a F?' has no clear answer. The answer is indeterminate or unknowable.

In *comparative* borderline cases, the relevant question is instead 'is a Fer than b?' If Hank has fewer thin hairs widely distributed over his head, and Henry has more thick hairs concentrated in a ring around his scalp, then it may be indeterminate or unknowable whether Hank is balder, or Henry is balder, or they are precisely equally bald. Hank and Henry form a comparative borderline-case of 'balder'.

As we'll see, comparative borderline-cases are ubiquitous, but their significance has not been appreciated. When theorising about vagueness, they do not represent a

---

[9]For semantic indeterminism, see Fine (1975), Keefe (2000), and Dorr (2003). Metaphysical indeterminism has recently been defended by Barnes (2010).

[10]Epistemicism is famously defended by Sorensen (1988) and Williamson (1994).

substantially distinct phenomenon, since any theory of ordinary or categorical vagueness is likely to extend trivially to the comparative sort. But once we appreciate its source and structure, we'll see that it can ground incommensurability.

Comparative borderlineness arises in two main ways (I can't think of any others). The first is tied to *multidimensionality*, or the dependence of the application of a predicate on multiple contributory factors. Those factors need to be weighed against each other, and it may be vague what the their relative weighting is. Just as 'is bald' does not knowably fix the precise number of hairs required to count as bald, so it does not knowably fix the precise relative weighting of hair numbers and hair distribution. So there are both categorical and comparative borderline-cases of that predicate. Similarly, there are several dimensions relevant to whether a vessel is a ship or a boat, or whether a dwelling is a tent or a teepee, and it is plausible that the relative weights of those dimensions are vague.

When we are required to score something along multiple dimensions, we might assign precise weights to those dimensions for calculative purposes, but this is often an arbitrary exercise. When comparing philosophy papers, should clarity count for twenty or twenty-one percent of a paper's overall quality? There seems to be no knowable answer to this question. Similarly, when grouping men as 'tall' or 'not-tall', we might have to fix an arbitrary threshold.

Second, comparative borderlineness may be parasitic on (or derivative of) categorical borderlineness. Consider the artificial predicate 'is militarily effective', and stipulate that one of its dimensions is being tall: all else being equal, the tall are more militarily effective than the not-tall. Consider two men, Hank and Herbert. The other dimensions—whatever they are—are such that Hank has a very slight advantage. Hank is not tall. So far, we have not introduced any vagueness. Suppose that Herbert is borderline-tall (a case of ordinary categorical vagueness). The other factors tip slightly in Hank's favour, so: if Herbert is tall, then he is more militarily effective than Hank; if Herbert is not tall, then he is less militarily effective than Hank. It is borderline whether whether Herbert is tall, so it is borderline whether Herbert is more or less militarily effective than Hank.

Derivative or parasitic comparative borderlineness can also arise without multidimensionality. Say that a is 'heap-richer' than b if a owns more heaps than b. Sarah has twenty clear heaps, and no other piles of sand. Sally has fifteen clear heaps, ten borderline-heaps, and no other piles of sand. It is vague how many heaps Sally has, and so it is vague which of them is heap-richer: if Sally has less than twenty heaps, then Sarah is heap-richer than Sally; if Sally has twenty heaps, then Sarah and Sally are equally heap-rich; if Sally has more than twenty heaps, then Sally is heap-richer than Sarah. Something similar might apply when measuring the relative economic potential of countries in terms of usable farmland, if 'usable' is vague.

So vague relative weightings are one way that comparative borderlineness can arise, but they are not the only way. Even if the dimensions are precisely weighted, it might be vague how one option 'scores' along one of those dimensions. This can apply even in cases where there is only one dimension. It is thus not tenable to accept categorical vagueness but deny the existence of comparative vagueness, since the former engenders the latter. (As we'll see later, Joseph Raz seems to imply that vagueness is typically associated with the unidimensional).

To recap, comparative borderlineness is the phenomenon whereby for some predicate F and two objects a and b, at least two of 'a is Fer than b', 'b is Fer than a', and 'a and b are equally F' are not all false, but none is knowably true (remaining neutral between epistemicism and indeterminism about vagueness).

Comparative borderlineness does not require that each of the three possible verdicts (Fer, less F, equally F) is vague; one may be known to be false. We saw this in the case of Hank and Herbert, where it is vague who is more militarily effective, but certain that they are not equally so.

We can now state precisely the main thesis of this paper, that incommensurability is vagueness:

> **Incommensurability as comparative borderlineness.** When a and b are incommensurate with respect to V, they are a comparative borderline-case of V-ness. Ascriptions of the three trichotomous comparisons ('a is Ver than b', 'b is Ver than a', 'a and b are equally V') are not all false, but none is knowably

true.

Value incommensurability is simply the evaluative instance of a broader semantic (or metaphysical, or epistemic) phenomenon. In the next sections, I'll argue that there are comparative borderline-cases of evaluative predicates, and that these can explain incommensurability in a parsimonious way. This will present a burden for those who favour trichotomous incomparabilism: why *not* think that incommensurability is vagueness?

1.2 Evaluative Comparative Borderlineness

Consider comparative borderlineness of an evaluative predicate:

> **Evaluative Comparative Borderlineness.** The pair a and b are a case of evaluative comparative borderlineness of V when: (i) a and b are a comparative borderline-case of V; and, (ii) V is an evaluative predicate.

In general, evaluative comparative borderlineness may arise in each of the ways discussed above. Evaluative comparisons are typically made relative to an implicit or explicit background consideration—what Ruth Chang calls a 'covering value'[11]—and if this value is multidimensional, it may be vague what the relative weights of its dimensions are: who would be so bold as to assign a precise numerical weight to insight versus rigour?

It might be thought that evaluative predicates are not susceptible to vagueness. But this is not plausible: just as vagueness is ubiquitous in the non-evaluative, so it seems to be in the evaluative: consider the concept of cowardice.[12] Is it more cowardly to run from a single man armed with a shotgun, or from a dozen men armed with just their fists? What is the precise minimum number of men I must be prepared to fight in order to be a non-coward? Vagueness is typically demonstrated with such a 'know it when I see it' presentation of borderlineness and tolerance. There is no more reason to think that 'is a

[11] See Chang (2004).

[12] I owe the example to Nathaniel Sharadin.

coward' is wholly sharp (lacking vagueness) than there is that any other (evaluative or nonevaluative) concept is.

It is hard to deny that our evaluative concepts have at least some vagueness. The vagueness-threshold is low: according to David Braun and Ted Sider, 'an expression is vague if it can be unclear to a speaker informed of all relevant facts whether the expression correctly applies'.[13] Now, it would clearly be question-begging to argue along these lines that *incommensurability* is vagueness, but I am making a less tendentious point: by the standards that are used in discussions of non-evaluative vagueness, there are clearly instances of evaluative vagueness. This is strong (but defeasible) evidence that there is evaluative vagueness.

Perhaps it might be worried that even if the broadly evaluative is subject to vagueness, specifically *moral* notions must be sharp. But as TK Seung and Daniel Bonevac argue, every plausible moral theory seems to involve some degree of imprecision. Compare even Kant on imperfect duties, such as that to give aid to those in need. There is no precise, determinate minimum provision of aid that is required.[14] I argue elsewhere that 'satisficing' versions of consequentialism are most plausible when the threshold of right action is vague.[15] Morality is not like income tax, with precise thresholds (and even income tax is beset by legal vagueness).

It might be worried in particular that vagueness is incompatible with certain plausible principles concerning the moral and evaluative, such as that 'ought' implies 'can', or that there are no unknowable moral obligations. But such issues are complicated, and do not obviously undermine vagueness. I discuss the application of the ought-reasons platitude—that one ought to do what one has most reason to do—elsewhere.[16]

---

[13]Braun and Sider (2007), p. 133.

[14]Seung and Bonevac (1992), pp. 805-806.

[15]See Elson (Manuscript[b]).

[16]See below and Chapter 4.

Evaluative comparative borderlineness begets incommensurability. Take any evaluative notion V, such as goodness or creativity, which has with comparative borderline-cases. Suppose that a and b are a comparative borderline-case of V. This is a case of evaluative comparative borderlineness. Then comparing definitions of evaluative comparative borderlineness and incommensurability, we see that a and b are thereby incommensurate with respect to V. This is how vagueness engenders incommensurability.

Even a small dose of such vagueness can generate widespread incommensurability:

> **Wealth Comparison.** Sarah has a lot of property (real estate), and not much cash; Sheila has more cash but less property. Thanks to vagueness in how property is to be weighed against cash, it is vague whether Sarah is richer than Sheila, or Sheila is richer than Sarah, or they are equally rich.

Sarah and Sheila are a comparative borderline-case of 'richer'. This borderlineness rests on a relatively small dose of comparative vagueness: how much are cash and property worth, relative to each other? We did not assume any vagueness in what is to count as cash or as property, for example. Nevertheless, this vague weighting will ground many more comparative borderline-cases. Any other pairs with similar ratios of cash to property will also form comparative borderline-cases of 'is richer than' (at least, assuming that this predicate has a relatively simple linear structure). This vagueness will also infect any concept which has comparative richness as a dimension or component such as, plausibly, socio-economic status.

Such vagueness in the nonevaluative predicate 'richer' can quickly generate evaluative vagueness. Many evaluative notions plausibly depend on wealth in some way: elegance, success, being an oligarch, or a role model, and so on. For a moral example, suppose that Sheila and Sarah are relatively rich but misers; neither gives anything to charity. We might judge that the wealthier of the two is more blameworthy for this. Since is is vague who is wealthier, it is vague who is more blameworthy. This line of thought was rather artificial, but it doesn't take a lot of vagueness ground widespread comparative borderlineness, including of evaluative and moral predicates.

I have argued that there is evaluative comparative borderlineness, and such instances meet the formal definition of incommensurability: none of the trichotomous comparisons knowably obtain.

Vagueness naturally and parsimoniously explains incommensurability in terms of an existing phenomenon. Now, I'll consider two influential challenges to the thesis that incommensurability is vagueness. Both seek to undermine the claim that incommensurability could be entirely explained by vagueness. Since my claim is that these challenges fail, it is important to consider them carefully.

1.3 Chang's Challenge

Ruth Chang has argued that incommensurability *could not* be rooted in borderline-cases. If she is correct, then incommensurability is a distinct phenomenon, not explained by vagueness. Chang calls incommensurate comparisons *superhard*. This terminology is in some ways preferable, since it does not stipulatively rule out a debunking account.

She concedes that superhardness and incommensurability share some formal similarities, but argues that they differ in several ways—phenomenologically, in resolution conditions, and particularly in *moral* resolution conditions. I'll reject each of these arguments: none marks a genuine distinction between the vague and the superhard.[17]

Her first argument is that the phenomenology of superhard cases differs from that of borderline-cases:

> In borderline cases, insofar as we are willing to judge that the predicate applies, we are also willing to judge that it does not apply. [...] in a superhard case, insofar as we are willing to judge that 'better than with respect to V' does not apply, we are not also willing to judge that it does apply. In the absence of any explanation for why the phenomenology should be different, there is

[17] Chang (2002), pp. 682ff. I am indebted to Ruth Chang for generous and incisive discussion and comments.

good reason to think that superhard cases are not cases of vagueness.[18]

Chang does not mean that 'in a borderline case one cannot favor one judgement more than another'.[19] Instead, we may illustrate her claim with an example. Consider a borderline-red ball. If we are willing to judge that it is not-red, we are also willing to judge that it is red. This tension seems characteristic of borderlineness: we fluctuate between both verdicts. But incommensurability, it seems, lacks this tension: we judge that some A is *not* better than some B, without also being willing to judge that it *is* better.

But this argument fails in both directions, so to speak. First, the relevant judgements of comparative borderlineness are less provisional than Chang implies. On the view I am defending, to judge that a pair form a superhard case is to judge that the pair is a comparative borderline-case of the relevant value. This is to make a metalinguistic (or epistemic, or metaphysical) judgement, and such judgements do not seem to have the degree of instability that Chang claims. There are paradigm borderline-cases of predicates like 'is red', and 'is more militarily effective'. It is perfectly possible to judge that we are dealing with a borderline-case, without also being willing to judge that it is *not* a borderline-case. Indeed, this grounds the puzzle of higher-order vagueness, which is the phenomenon where some cases seem to be 'clearly borderline' and others 'borderline-borderline'. There do seem to be clear cases of borderlineness.

Second, judgements of superhardness are more provisional than Chang claims—as Simon Blackburn notes, it is very hard to know whether a quandary grounded in incommensurability will be stable in the long-run, and a judgement that it *is* stable may always be defeasible[20]—and often have the same air of paradox and revisability as judgements of borderlineness. Here we are trading intuitions, and I will report my own: judgements that some pair is incommensurate with respect to some value do have a

[18] Chang (2002), p. 682.

[19] Chang (2002), p. 682 fn. 26.

[20] Blackburn (2010), especially pp. 48 and 61. Blackburn discusses the general category of 'quandaries'; his definition of a quandary includes incommensurability, however one understands the latter phenomenon.

phenomenal instability to them, as do judgements that something is a borderline-case of a vague predicate. Phenomenology does not separate the vague from the superhard.

Chang's next argument is more threatening. She argues that vagueness and incommensurability have different *resolution conditions*—that whilst disagreement in the face of the superhard is subject to substantive dispute, disagreement in the face of vagueness can be stipulated away:

> in borderline cases, it is perfectly permissible to resolve the indeterminacy in favour of application or not by arbitrary stipulation, but in superhard cases, resolution of the perplexity in context-free cases is not permissibly given by arbitrary stipulation.[21]

Chang introduces a thought experiment, to abstract from contextual factors. Jack is playing a game, wherein he must sort people into a 'bald' or a 'non-bald' pile. Players of the game are fully informed; Herbert is borderline-bald, and Jack must choose a pile; Jack sorts Herbert into the 'bald' pile. Then:

> What is crucial is that the sorting decision is perfectly arbitrary; he could just as well have flipped a coin to determine how Herbert was to be sorted [. . .] if we add another player, Jill, who happens to sort Herbert into the 'not bald' pile, Jack and Jill have no real disagreement; their 'disagreement' is simply a clash of arbitrary stipulations in the face of indeterminate application. Of course, this is not to say that the answer to the question, Is Herbert bald, is given by a coin flip, for the answer to that question is that it is indeterminate whether he is. The point here is only that a resolution of this indeterminacy can be appropriately given by a coin flip. In general, the resolution of borderline-cases can always be a matter of arbitrary stipulation.[22]

[21]Chang (2002), pp. 679-688.

[22]Chang (2002), p. 684.

I agree with almost everything in this paragraph, provided that 'resolution' is interpreted in terms of deciding what to *do*: if it is indeterminate which pile Herbert must go into, but determinate that he must go into one, then what is there to do, but stipulate arbitrarily? By hypothesis, there is no determinate, knowable reason to choose one over the other.

Note that here 'arbitrary stipulation' need not be limited to random picking, such as coin flips. One can adopt a general policy to deal with borderline cases, such as putting the tall borderline-bald men into one pile, and the not-tall in another. The point is just that the rules of the game do not specify into which pile Herbert should go, so we must appeal to something 'outside the rules', so to speak. This will be arbitrary from the perspective of the rules.

Chang claims that superhard cases lack such arbitrariness:

> The resolution of perplexity in superhard cases is very different in nature. Suppose that Jack is confronted with the superhard case involving Mozart and Michelangelo. The rules of the game require him to put the ordered pair in one pile or the other ... what is crucial is that his decision is not properly arbitrary; it is not true that the could just as well have a flipped a coin to resolve the perplexity. If Jill puts the pair in (the other pile), Jack and Jill have a genuine disagreement over whether Mozart has the greater creativity—this is no clash of arbitrary decisions but a substantive disagreement in which arguments can be brought to bear.[23]

If the Mozart/Michelangelo comparison really is superhard, then this disagreement could not be substantive in the way Chang claims. If Mozart and Michelangelo are incommensurate, then how could a dispute between two fully-informed players about how to classify them be substantive? Since they are fully-informed, Jack and Jill agree that the comparison is superhard: they agree that it is not (knowably) true that Mozart is more creative, or that Michelangelo is more creative, or that they are equally creative.

[23]Chang (2002). pp. 684-5.

Since this is common knowledge between the players, it is hard to see how resolution could be a matter of substantive debate between them. Suppose that Jack thinks that it would be better to choose Mozart, solely on the grounds of creativity: does that not imply that he thinks that Mozart is more creative? It is incoherent to think that Mozart and Michelangelo are paradigmatically incommensurate with respect to creativity, but that on substantive grounds of creativity, Mozart should be put in the 'more creative' pile.

Interestingly Simon Blackburn, who has no stake in the debate over incommensurability as vagueness, agrees that incommensurability often demands arbitrary stipulation, and that 'both common sense and high theory tell us how to handle it. The agent has to *plump* for one alternative.'[24] Blackburn may be wrong about this, of course, but this does show that Chang's claim—that the resolution of superhard cases must be substantive—is at the very least, far from obvious.

But what about a phenomenal claim, that it *feels* substantive? It is true that there is more room for debate about *whether* the Mozart/Michelangelo pair is superhard than there is in the baldness case. There is more at work in creativity than in the largely observational 'is bald'.

As Cristian Constantinescu and others have pointed out, there are many differences between the cases.[25] Here are some. Whereas 'is bald' is a monadic predicate that applies to individuals, 'is more creative than' is a dyadic comparative predicate that applies to pairs. The application of 'is bald' is determined by perhaps a few dimensions—hair number, distribution, colour, thickness—where the first and second are by far the most important; the application of 'is more creative than' is certainly determined by many more dimensions (there may not even be a determinate list, if the predicate has 'open texture'). And whilst 'is bald' is normally a nonevaluative predicate, 'is more creative than' is straightforwardly evaluative.

---

[24] Blackburn (2010), p. 50.

[25] See Constantinescu (2012). But his own preferred explanation—that 'the application of vague predicates can seem less arbitrary not because those predicates are any less vague, but simply because it matter to us more in those contexts how we apply the relevant predicates' (p. 62)—can't be the whole story. Stipulatively, here we are dealing with a game, which rules out contextual factors like importance.

The cases share little in common, other than that Jack and Jill face perplexity about the application of a predicate. This is not a good test for whether superhardness introduces substance. Consider another, which avoids these potential distortions. We can press into service a nice example discussed by Seung and Bonevac: university admissions.[26] In the United States, admission to most undergraduate programmes is determined by several dimensions, including exam scores, essays, money, letters of recommendation, sports and extracurriculars, and ethnic or racial category. Suppose that Jess and John are paid to play a game: they must select candidates for a given university. They must put candidates into one of two piles: accept, or reject.

They have one slot left, and two remaining candidates, a and b. a and b are roughly equal in most respects, but for the following differences: a has test scores about 5% better than b's; b has done several more sporting activities than a. So a is better in one way, and b is better in another.

Consider two ways that Jess and John's game might be run:

> **Admissions as Superhard.** Jess and John are under instructions to 'admit the best candidates'. They are working to decide who is the better candidate, which is a plainly evaluative task. The choice between a and b is superhard.

> **Admissions as Vague.** Jess and John working from an instruction sheet; the new management has excised individual judgement from the process as much as possible. The sheet contains the following vague instruction: 'admit candidates with higher test scores, unless they have done substantially less sport'. The a/b pair is a comparative borderline-case of this criterion: a has better scores, but it is vague whether she has done *substantially* less sport than b. It is vague what the instructions require.

The two cases are exactly parallel, except that one is superhardness, and the other is vagueness. If Chang's argument were correct, we would find the difference she

[26] See Seung and Bonevac (1992), pp. 806-809.

describes—that in one case resolution is arbitrary and stipulative, in the other case substantive—here too. Yet in both versions, arbitrary choice is permissible, even required. What other way is there to proceed? In both cases, the admissions officers Jess and John agree that the rules of the game do not determine which candidate to admit. How are they to fill the slot if they do not go 'beyond the rules' by choosing arbitrarily in some sense?

If there were a striking difference in acceptable resolution methods between the superhard and the borderline, then it would manifest in this case, but it doesn't. The best explanation is that resolution in the Mozart/Michelangelo case must also be arbitrary, and any appearance to the contrary can be explained by special features of that case. Chang has not shown that the arbitrariness implied by vagueness undermines incommensurability as vagueness.

If I am right that in cases of incommensurability, arbitrary stipulation *is* required, then Chang's argument actually strengthens the case for incommensurability as vagueness: Chang is right that vagueness mandates arbitrary stipulation, and if incommensurability is vagueness, then this neatly explains why incommensurability mandates arbitrary stipulation.

Chang's final argument is a moral application of the resolution argument:

> Consider a superhard case involving comparison of a particular act of promise keeping and a particular act of bringing about great happiness with respect to moral goodness [. . .] Could the resolution of the case be an arbitrary matter—could the perplexity concerning which is morally better be answered by the flip of a coin? Clearly, the resolution of this superhard case cannot be a matter of arbitrary stipulation but is a substantive matter concerning which is better.[27]

But again—how could this *not* be a matter of arbitrary stipulation? Remember that *arbitrary stipulation* here might include choice on non-moral grounds. By hypothesis, the case is 'morally incommensurate': neither option is morally better, but nor are they

[27] Chang (2002), p. 685..

morally equipreferred. To talk in terms of moral reasons, there is no overall moral reason to prefer one or the other, and no overall moral reason to equiprefer them. What is there to do but choose arbitrarily (from the point of view of morality)? We are in a similar position to the admissions officers: moral guidance has run out.

If evaluative incommensurability is vagueness, as I argue, then 'moral incommensurabilities' are also vagueness. This is not a bullet to be bitten, but a natural consequence of the view. As we have noted, many plausible normative ethical views incorporate a degree of indeterminacy or vagueness. Here, I simply invite you to share the judgement that in a suitably fleshed-out moral case (along the lines that Chang describes), there may be no alternative to arbitrary choice. This is not to rule out the possibility of moral dilemmas: that arbitrary choice is permitted does not undermine the possibility of conflicting non-overridden moral requirements, since we may be in the unfortunate position of arbitrarily stipulating which conflicting moral obligation to meet.[28]

Now, there might be more 'broadly pragmatic' reasons to be wary of coin-flipping when the moral stakes are so high. Perhaps coin-flipping would fail to show due deference to morality, or in the case of legal vagueness, to the law.[29] But it doesn't follow that there is a non-arbitrary way to proceed. As Blackburn notes, some such arbitrary picking methods—such as cards, or drawing straws—do carry an air of seriousness.[30] We should not confuse the culturally-contingent connotations of some methods of arbitrary stipulation with evidence that arbitrary stipulation *per se* is inappropriate.

Chang has offered three arguments that superhard cases could not be vagueness. But each fails; given the argument of the precious section, we may tentatively conclude that at least some incommensurability is evaluative comparative borderlineness. But we have yet to defend the more tendentious thesis of this paper: that *all* incommensurability is

[28]Sinnott-Armstrong (1988) frames moral dilemmas in terms of such requirements.

[29]See Endicott (2000).

[30]Blackburn (2010), p. 50.

vagueness.

1.4 Against Hybrid Views: Raz's Challenge

A 'hybrid' view holds that though some incommensurability is vagueness, there are also instances of another kind:

> **Hard-incommensurability.** a and b are hard-incommensurate when it is false that a is better than b, and false that b is better than a, and false that a and b are equally good.

Joseph Raz claims that there are widespread instances of hard-incommensurability, and some (less dramatic) instances of incommensurability grounded in vagueness. I'll argue that this view is consistent and coherent, but that it is unmotivated, and faces an unmet explanatory burden.

Raz claims that vagueness is 'a less dramatic case' of incommensurability:

> The most important source of incomparability is 'incomplete' definition of the contribution of the criteria to value. [...] It is possible that our way of weighting the different criteria does not establish a different ranking of all possible combinations. [...] There are at least two other sources of incomparability. First, indeterminacy results from vagueness and the absence of sharp boundaries which infects language generally and therefore apply to value measured by a single criterion as well. These apply even in cases in which a single descriptive criterion determines the value of options...[31]

But as we have seen, this 'incomplete definition' or balancing is easily (and most naturally) characterised as vagueness. The argument might be that since incommensurability arises mainly in cases of multiple balancing considerations, and vagueness is also found in 'value measured by a single criterion', there must be something besides vagueness underlying incommensurability. But then the argument fails, since vagueness can explain why incommensurability typically arises from multiple

[31] Raz (1985-86), p. 122.

balancing considerations: they must be weighted against each other, which is a locus for vagueness in their weights, and thus for widespread comparative vagueness. We illustrated this phenomenon above.

Vagueness can also explain some of the specifics about the connection between multidimensionality and incommensurability. When options are Pareto-mixed, the relative weights of the dimensions come into play: are the ways in which the first option is better weightier than those in which the second is better? In close calls, the answer to this may be vague, generating comparative borderlineness; but as with all vagueness, there are some clear cases. Pareto-mixing does not guarantee incommensurability. Incommensurability as vagueness strikes the right balance: it explains Pareto-mixing's close connection to, but insufficiency for, incommensurability.

Raz seems to imply that because vagueness can occur even in unidimensional cases, this counts against the view that incommensurability is vagueness. We have however seen that this argument fails, though it is initially plausible. Comparative borderlineness is more closely tied to multidimensionality, but it can also (rarely) arise in unidimensional cases. This can be seen most plainly in our discussion of the comparative heap-wealth of Sally and Sarah, above. If it is better to be heap-richer, then there is no principled reason to exclude Sally and Sarah as a case of incommensurability, purely on the grounds of unidimensionality. Incommensurability as vagueness can explain why multidimensionality grounds the central cases of incommensurability.

It might be thought (and Raz seems to argue) that any incommensurability due to vagueness would be essentially marginal, and that it could not be the central phenomenon. But as we have seen, a little comparative vagueness goes a long way.

Hybrid views are coherent; but they must be argued for. As things stand, they have the same status as a view that though there are ordinary borderline-cases of 'is tall', there are also people of whom it is *false* that they are tall, and *false* that they are not tall. The natural response to such a view would be a request for evidence.

1.5 Conclusion

Raz and Chang argue in different ways that incommensurability could not be vagueness, or not entirely so. But they fail to appreciate the resources available to the vagueness view. Vagueness need neither be marginal, nor unidimensional, nor problematically arbitrary.

Incommensurability as vagueness and its rivals can each capture the formal structure of incommensurability: that none of the three trichotomous comparisons is knowably true. So we must then turn to the virtues of the competing explanations. (Which in some cases, though not here, itself generates a kind of explanatory incommensurability.) Incommensurability as vagueness offers a complete, parsimonious explanation of the phenomenon of incommensurability. Hard-incommensurability is explanatorily superfluous; arguments to the contrary mischaracterise either incommensurability or vagueness.

Moreover, there is a general question for the Trichotomous Incomparabilist who nevertheless accepts that there is widespread evaluative comparative borderlineness: where is it all? If Chang is right that it does not manifest as incommensurability, then how does it manifest? And if Raz is right that some incommensurability is grounded in vagueness, but the main part is hard-incommensurability, then why can we not identify clear cases of the two different phenomena? This has proceeded at a very general level, but there is another way of pressing the point: for any case of incommensurability, we can locate a source of vagueness that plausibly generates it. I can only report my own experience, but I cannot think of any instances of incommensurability that cannot be given a plausible analysis in terms of vagueness.

This is a burden-shifting argument.[32] As such, we cannot rule out the possibility of an as-yet unseen argument that some or all incommensurability could not be comparative borderlineness. But the prospects for this are not promising: vagueness seems able to explain the structure of incommensurability, its connection to multidimensionality and

[32] I'm grateful to John Broome for suggesting this way of putting it.

especially Pareto-mixing, and why incommensurability requires arbitrary choice or stipulation.

Incommensurability is a puzzle. But the thesis that incommensurability is vagueness has far-reaching consequences for the nature of rational choice and action. If choice in the face of incommensurability is choice in the face of vagueness, then it is often (at best) borderline whether we choose the most valuable option. I think that this, along with vague credences in epistemology, are manifestations of a broader phenomenon of *vague reasons*, which fail to pick out one class of actions as uniquely justified. As I argue elsewhere, the ought-reasons platitude—that one ought to do what one has most reason to do—must be modified in the face of vagueness, since it cannot cope with the claim that whichever way we plump, we have done as we ought.[33]

Incommensurability as vagueness also promises to explain the troublesome phenomenon of *value-pumping*: by exchanging incommensurate objects—apparently permitted on every account of incommensurability—repeatedly, we find ourselves determinately, clearly worse off. This also arises in (what I argue are) other manifestations of vague reasons, such as Warren Quinn's Puzzle of the Self-Torturer, and sequences of bets that exploit vague credences.[34] Assimilating incommensurability to vague reasons promises an explanation of the impermissibility of some sequential choices, in terms of tolerance. It is far from clear whether other accounts have the resources to do so.[35] But this is an issue for another day, and here I merely offer a promissory note on the value-pumping front.

Today's more modest conclusions are that incommensurability is vagueness—at least, we haven't seen any reason to think that it isn't—and that it is vague which holiday is better.

---

[33]See Chapter 4.

[34]See Quinn (1990), Elga (2010), and Tenenbaum and Raffman (2012).

[35]See Chang (2005), pp. 346ff for a discussion in terms of parity.

2 Heaps and Chains: is the Chaining Argument for Parity a Sorites?

Abstract. I argue that the Ruth Chang's Chaining Argument for her *parity* view of value incomparability trades illicitly on the vagueness of the predicate 'is comparable with'. Chang is alert to this danger, and argues that the predicate is not vague, but this defence does not succeed. The Chaining Argument also faces a dilemma. The predicate is either vague or precise. If it is vague, then the Argument is most plausibly a Sorites. If it is precise, then the Argument is either question-begging or dialectically ineffective. I argue that no Chaining-type argument can succeed.[1]

Ruth Chang defends the strikingly original claim that there is a fourth value relation of *parity*, instantiated in some cases of value incomparability (also called 'incommensurability'). The Parity view is that in at least some evaluative comparisons between two options, neither is better than the other, and they are not equally good, but the options are nevertheless comparable: they are *on a par.* Parity is a fourth comparative relation, besides the 'trichotomous' three of betterness, worseness, and equality.

A central argument for this view is the Chaining Argument.[2] But, I shall argue, it has not been noticed that the Chaining Argument illicitly trades on the vagueness of its key predicate, namely 'is comparable with'. Chang is alert to the danger of vagueness, but her defence against it applies to only one of the two ways vagueness could undermine the Argument, and there is no clear way to generalise this defence.

---

[1] I am indebted to Ruth Chang, Geoffrey Sayre-McCord, Keith Simmons, and three anonymous reviewers for *Ethics*, for extensive discussion and comments. This paper appears in *Ethics* Vol. 124, No. 3 (April 2014), pp. 557–571.

[2] Chang (2002), especially section II. Chang has also defended Parity elsewhere. For discussion, see for example Gert (2004) and Wasserman (2004).

In section 1, I sketch the terrain. In section 2, I describe the Chaining Argument. Section 3 shows how the Chaining Argument is structurally similar to a Sorites fallacy, and describes Chang's defence against this. Section 4 shows that this defence is ineffective against another, overlooked way in which the Chaining Argument could be a Sorites. In section 5, I argue that this possibility explains the intuitive plausibility of the argument's central premise. In section 6, I argue that Chang's defence cannot be extended to cover this possibility without question-begging. Section 7 concludes with an argument that no Chaining-type argument is likely to be effective in defence of Parity.

2.1 The Terrain

In this dialectic, we are concerned with cases of Trichotomous Incomparability:

> **Trichotomous Incomparability.** Two objects A and B are *trichotomously incomparable* just in case it is not true that A is better than B, not true that B is better than A, and not true that A and B are equally good.

This definition of trichotomous incomparability rules out 'epistemic' views, according to which, relative to a given comparison, it is always the case that one option is better or they are equally good, though we are frequently ignorant of which. Chang argues convincingly against such views.[3]

Let's follow Chang in two ways. First, 'good' and 'better' are often placeholders for some more specific evaluative term. We will be especially concerned with the evaluative predicate 'creative' and its comparatives. Second, we use the names 'Mozart' and 'Michelangelo' somewhat stipulatively: (i) they are trichotomously incomparable, with respect to creativity; (ii) they are the most favourable case for the Parity view—if parity exists, then Mozart and Michelangelo are creatively on a par.

Following John Broome,[4] such cases of incomparability are usefully represented as a *standard configuration*. Here is how we might represent the comparison of various sculptors to Mozart, with respect to artistic creativity:

[3] For discussion, see the discussion at Chang, "The Possibility of Parity" 668ff, and the works cited there.

[4] Broome (1997).

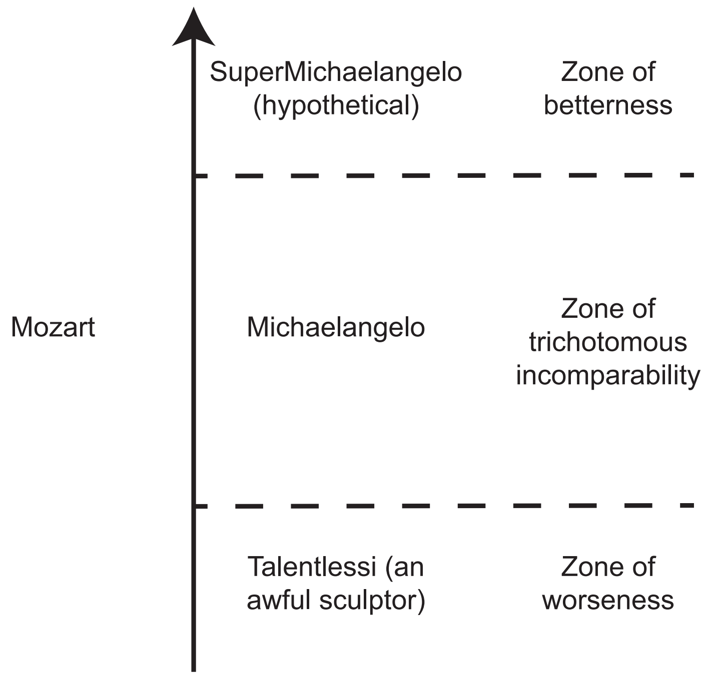

Figure 2.1: A standard configuration with trichotomous incomparability.

Here Mozart is the 'standard' in question, against which the sculptors will be compared. The arrow indicates sculptors of increasing creativity; the dashed lines indicate that the nature of the zonal boundaries is unspecified. For clarity, I have included a region of (perhaps merely possible) sculptors who are more creative than Mozart. Finally, the precise relative positions of Mozart and Michelangelo should be ignored; they are artefacts of diagramming.

The question at hand is, what's going on in the zone of trichotomous incomparability? It is natural to think that if A and B are trichotomously incomparable, then they are incomparable full stop:

> **Hard Incomparabilism.** In cases of Trichotomous Incomparability, it is false that A is better than B, and false that B is better than A, and false that A and B are equally good. In virtue of this, A and B are not evaluatively comparable.[5]

[5]For a classic presentation, see Raz (1985-86).

To say that A and B are evaluatively comparable is to say that there is a true comparison that holds between them.[6] If Hard Incomparabilism is right, then Mozart and Michelangelo are trichotomously incomparable, and incomparable full stop:

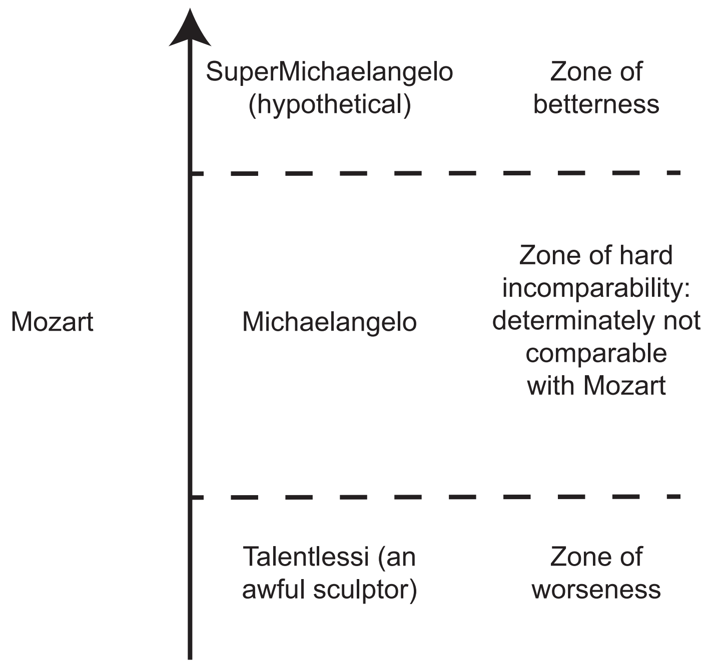

Figure 2.2: Hard incomparabilism.

After all, what other comparison *could* hold? This intuition is what grounds the Trichotomy Thesis:

> **Trichotomy Thesis.** '[If] two items A and B are evaluatively comparable, then A must be better or worse than B, or A and B must be equally good.'[7]

If the Trichotomy Thesis is right, then the only such relations are the classic three. Chang denies this trichotomous orthodoxy:

> **Parity view.** There are some cases of trichotomous incomparability which are nevertheless cases of evaluative comparability. In these cases, the

[6]In particular, 'comparable' does not mean *more or less equal*, as sometimes in everyday use.

[7]Chang (2002), p. 660.

trichotomous relations determinately do not hold, and the fourth relation of parity determinately holds.

On this view, parity is not derivative or otherwise second-class. It is as respectable as betterness, worseness, and equality. So even though Mozart and Michelangelo are trichotomously incomparable, they are comparable. There is a fourth comparative which applies. Mozart and Michelangelo are on a par:

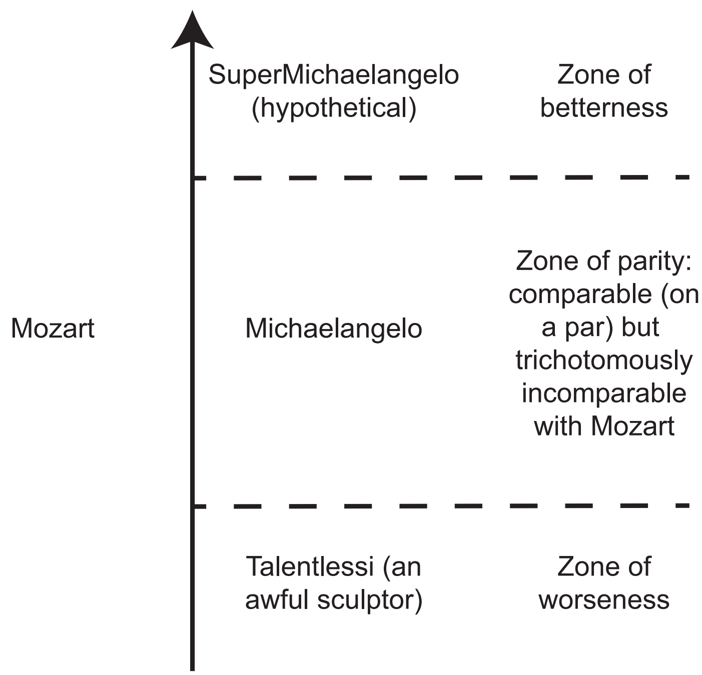

Figure 2.3: Parity

So the locus of disagreement between Hard Incomparabilism and Parity amounts to this: are there such a Mozart and Michelangelo, comparable yet not trichotomously comparable? Hard Incomparabilism stands or falls with the Trichotomy Thesis. Trichotomy is true just in case there is no such pair, and in that case Hard Incomparabilism is the right view.

Ruth Chang offers an existence proof for parity, and thus an argument that Trichotomy is false; in terms of our stipulations, she gives an argument that Mozart and Michelangelo are on a par. This is the Chaining Argument, and it is the topic of this paper.

2.2 The Chaining Argument

The Chaining Argument has two uncontroversial premises, and two crucial and nonobvious premises. The first uncontroversial premise follows from our stipulations:

> (CA1) Mozart and Michelangelo are trichotomously incomparable: it's not true that Mozart is more creative, and not true that Michelangelo is more creative, and not true that they are equally creative.[8]

The second uncontroversial premise is that Mozart is comparable with Talentlessi, an awful sculptor—Mozart is much better.[9] Let's call Talentlessi S0 (sculptor-zero), and Michelangelo Sn (sculptor-n) for reasons that will shortly become clear. Then:

> (CA2) Mozart is comparable with S0.

The nonobvious premises are the *Existence Claim* and the *Small Unidimensional Difference Principle*.

The Existence Claim is that Talentlessi is connected to Michelangelo via a series of 'small unidimensional differences'.[10] By repeatedly improving Talentlessi in single small ways, we could make him as creative as Michelangelo. Put another way, there exists a chain of (in most cases) merely possible sculptors

> S0 (Talentlessi), S1 (a sculptor slightly more creative in one respect than S0), S2, . . . , Sn (Michelangelo)

such that between any two adjacent members of the chain there is only a small unidimensional quality difference. Let's extend our numbering in the obvious way:

> (CA3) **Existence Claim.** The chain S0, S1, . . . Sn exists, and between any two adjacent members there is just a small unidimensional difference.

---

[8] Remember that we are ruling out epistemicism, and using these names somewhat stipulatively.

[9] See Chang (2002), p. 672 fn. 18.

[10] Chang (2002), pp. 673-674.

This premise could be denied, but it is not our focus here. For our purposes, the next premise is crucial. This is the Small Unidimensional Difference Principle, which has it that these small unidimensional differences preserve comparability, at least in this most favourable case:

> '... between two evaluatively very different items, a small unidimensional difference cannot trigger incomparability where before there was comparability. In other words, for items that bear very different respects of the covering consideration, incomparability between them cannot be a matter of some small difference in one of the respects borne such that without this small difference the items would be comparable.'[11]

In this case, the Difference Principle claims that if Mozart is comparable with one member of the chain, then he is comparable with adjacent members of the chain:

> (CA4) **Small Unidimensional Difference Principle.** If Mozart is comparable with some Sk in the chain, then if there are S(k-1) and S(k+1) in the chain, then Mozart is comparable with S(k-1) and S(k+1).

We are ready to state the rest of the Chaining Argument:

> (CA5) Mozart is comparable with Sn (Michelangelo). [CA2, CA3, CA4, and n applications of Modus Ponens.]

> (CA6) Mozart is comparable but not trichotomously comparable with Michelangelo. [CA5 and CA1.]

> (CA7) So the Trichotomy Thesis and Hard Incomparabilism are false, and a fourth comparison—stipulatively named 'parity'—holds between Mozart and Michelangelo. [CA6.]

[11] Chang (2002), p. 674.

The structure of this argument is clear. Since Mozart is comparable with Talentlessi, and there exists a chain of the right sort from Talentlessi to Michelangelo, we may repeatedly apply the Difference Principle and show that Mozart is comparable with Michelangelo. But none of the original three comparative relations applies, so Mozart and Michelangelo are comparable in some fourth way—parity. If this argument is sound, then Hard Incomparabilism, which denies that Mozart and Michelangelo are comparable, is false.

2.3 Is the Chaining Argument a Sorites?

The Chaining Argument resembles 'classic' cases of Sorites reasoning:

> (R1) Someone with £1 is not rich. [Premise.]
>
> (R2) If someone with £k is not rich, then someone with £(k+1) is not rich [Premise].
>
> (R3) So: someone with a million pounds is not rich. [R1; R2; 999,999 instances of Modus Ponens.]

Call this the 'rich man argument'. It is not a sound argument for its conclusion, but is a Sorites fallacy. Given this structural similarity, we should worry that the Chaining Argument is also a Sorites.

Chang rightly notes that not all arguments with the form of a Sorites *are* Sorites. Mathematical induction is valid (one hopes), but it has a similar form to the Sorites. So is the Chaining Argument more like the rich man or more like mathematical induction, in the relevant respects? The structural similarity between the cases introduces a burden that must be discharged: why should we not think that the Chaining Argument is a Sorites?[12]

Sorites fallacies turn on the vagueness of the predicates involved. In particular, it is the vagueness in 'is rich' that renders the rich man argument fallacious, and a lack of corresponding vagueness that renders mathematical induction respectable. So showing that there is no vagueness present is an effective means of discharging this burden.

[12] I'm grateful to an anonymous *Ethics* referee for pointing out several infelicities and unclarities in this part of an earlier version of the paper.

Chang adopts this strategy to mitigate the worry that the arguments for Parity 'look suspiciously like sorites arguments'. In particular, she argues that:

> if [the arguments] did turn on the vagueness of these comparatives, the cases at issue [ie, cases of trichotmous incomparability] would not be examples of some fourth value relation but rather borderline cases of one of the traditional trichotomy of relations.[13]

Chang offers several 'direct' arguments that trichotomous incomparability is not borderlineness, and so the consequent of this implication is false. She rejects the Semantic Indeterminist view of incomparability:

> **Semantic Indeterminism.** It is determinate that either Mozart is better than Michelangelo, or that Michelangelo is better than Mozart, or that they are equally good. However, it is indeterminate *which* comparison applies.[14]

On the Semantic Indeterminist view, 'Michelangelo is incomparable with Mozart' is true just in case it is indeterminate which of the comparisons better/worse/equality applies between Michelangelo and Mozart. Depending on one's account of indeterminacy, it may be determinate that one of them applies, but indeterminate which. To say that they are incomparable is to make a metalinguistic claim:[15] that Michelangelo is a borderline-case of the comparative predicate 'is more creative than Mozart'.[16]

This indeterminacy might be grounded in indeterminacy in how the features relevant to 'is more creative' are to be weighed: on some acceptable weightings, Mozart is better; on some, Michelangelo is better; perhaps on a few, Mozart and Michelangelo are equally

[13] Chang (2002), p. 665.

[14] A neat defence of this view (but which which I think ultimately fails) is Broome (1997). I defend a version of the view in 1.

[15] Or an epistemic one, or a metaphysical one, depending on one's view of vagueness.

[16] And, depending on the case, of its 'less creative' and 'equally creative' counterparts.

good. Then it will be indeterminate which of the comparisons applies:[17]

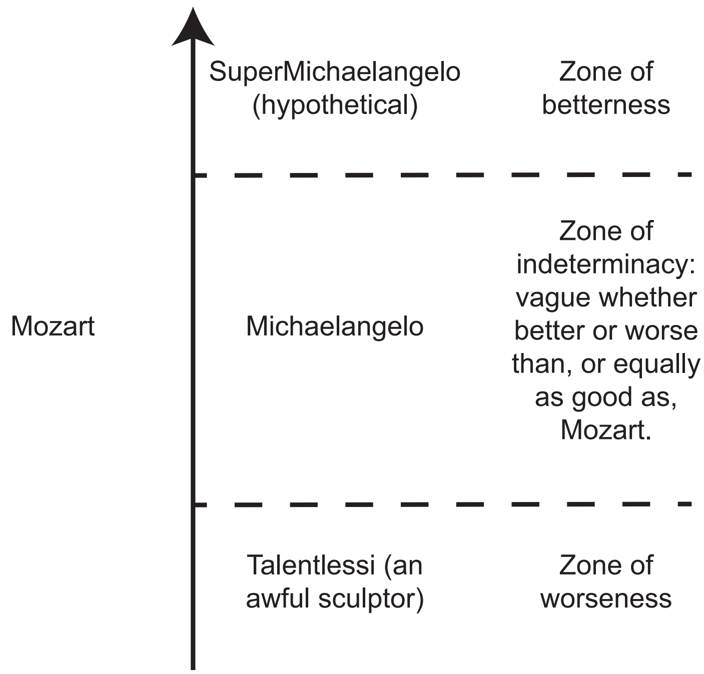

Figure 2.4: Semantic Indeterminism

As I mentioned, Chang offers several arguments against Semantic Indeterminism.[18] The claim is that if we can show that Semantic Indeterminism is false, then the predicates in question are not vague in a way that would undermine the Chaining Argument.

So Chang offers a three-stage response to the worry that the Chaining Argument might turn on vagueness: (1) if the argument is a Sorites, then the predicate involved is vague; (2) if the predicate involved is vague, then Semantic Indeterminism is true; (3) the direct arguments refute Semantic Indeterminism. (1) is unproblematic, and we may accept (3) for the sake of argument. But in the next section, we'll see that (2) is false: Semantic Indeterminism is not the only way that 'is comparable with Mozart' could be

[17] Remember that we are setting aside epistemicism as an account of trichotomous incomparability. Note that one could combine Semantic Indeterminism with an epistemicist view of vagueness; I would also count this as a form of Semantic Indeterminism, albeit a terminologically odd one. The core claim of Semantic Indeterminism is that incomparability is borderlineness; this is compatible with an ignorance analysis of borderline-cases of vague predicates.

[18] Chang (2002), pp. 682ff. These arguments are incisive, and are amongst the central challenges to my own semantic indeterminist view. I respond to them in 1.

vague.

2.4 Fuzzy Hard Incomparabilism

Let's look again at Hard Incomparabilism, which has it that Michelangelo occupies a zone of determinate incomparability with Mozart. That view does not specify the nature of the boundaries of that zone. In particular, that zone could have vague boundaries. Such a view would also posit vagueness, but it would be quite distinct from Semantic Indeterminism:

> **Fuzzy Hard Incomparabilism.** Hard Incomparabilism is correct. There are some sculptors who are determinately incomparable with Mozart, and some who are determinately comparable with Mozart. But the boundary between the comparable and the incomparable is vague. Some sculptors lie on the vague border between being incomparable with Mozart, and worse than Mozart. They are borderline-comparable with Mozart.[19]

On this view, there is no determinate number of distinct small unidimensional improvements required before the improved sculptor is incomparable with Mozart:

---

[19] As two anonymous referees for *Ethics* have rightly pointed out to me, it is slightly tendentious to claim that if the boundary is vague, then there are borderline-cases. Such talk adds clarity, but should officially be interpreted in terms of a vague boundary. My argument does not rely on borderline-cases: it is not borderline-cases but *tolerance* (discussed below) that brings the possibility of Sorites fallacy. As an autobiographical note, I suspect that borderlineness is central to vagueness, but that is admittedly controversial.

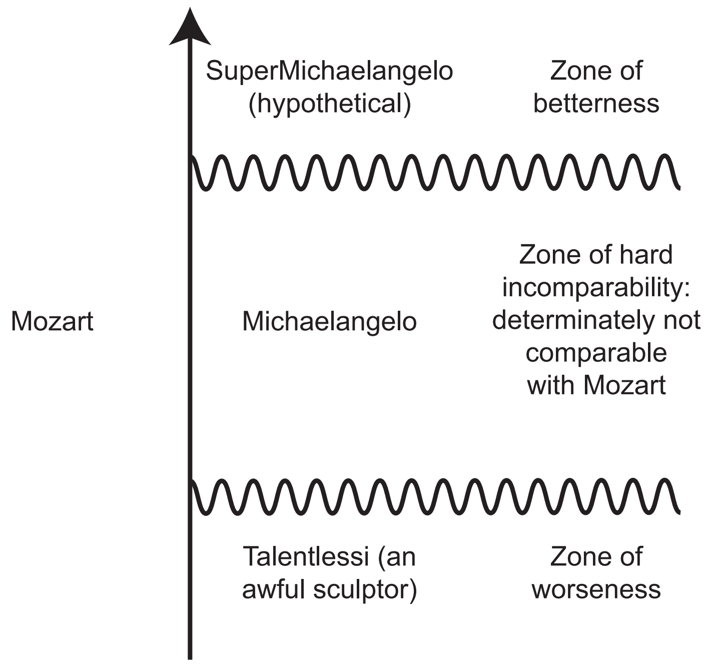

Figure 2.5: Fuzzy Hard Incomparabilism (wavy lines indicate vagueness)

The answer to that question is vague, just like the answer to the question 'how many pounds does a pauper need to become rich?' is:

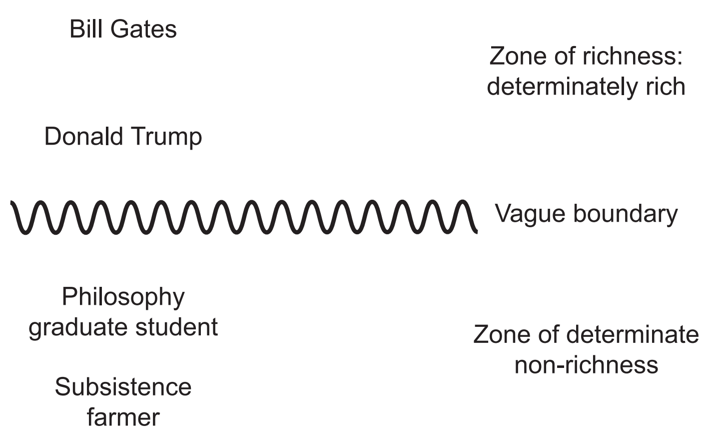

Figure 2.6: The vague boundary of 'is rich' (wavy lines indicate vagueness)

But just as this is compatible with there being many amounts of money n such that having £n *does* make one determinately rich, so there are many sculptors at the 'upper end of the chain' such that having been improved n times *does* make Sn determinately incomparable with Mozart.[20] This shows that (2) is false: even though this view is not Semantic Indeterminism, the crucial predicate 'is comparable with Mozart' is vague.

Even if Chang's direct arguments are effective against Semantic Indeterminism, they do nothing to rule out this view. The target of those arguments is the possibility where cases of trichotomous incomparability 'would not be examples of some fourth value relation but rather borderline cases of one of the traditional trichotomy of relations'.[21] In other words, the arguments are against the claim that cases of trichotomous incomparability are borderline applications of a vague predicate (or several). They aim to show that there are significant differences between cases of incomparability like that of Mozart and Michelangelo with respect to creativity, and cases of vague-borderline cases like those of 'is balder than'. But Fuzzy Hard Incomparabilism does not claim that cases of incomparability are borderline-cases; its central claims are that cases of incomparability are those where the three trichotomous comparisons betterness/worseness/equality determinately fail to apply, and that 'is incomparable with Mozart' has vague boundaries.

In slogan form: Semantic Indeterminism is the claim that incomparability is a *phenomenon of vagueness*—to be incomparable with Mozart is to be a borderline-case of the three trichotomous comparisons. Fuzzy Hard Incomparabilism denies this. On this view, to be incomparable with Mozart is for each of these comparisons to be false. But the predicate 'is incomparable with Mozart' has vague boundaries, just as 'is a heap' and 'is tall' do. This view implies the falsity of Semantic Indeterminism, so arguments against the latter (such as Chang's direct arguments) do nothing against it. Now, we can see how Fuzzy Hard Incomparabilism would render the Chaining Argument a Sorites. The point here is twofold. First, it is admitted by all that the Chaining Argument is structurally similar to a Sorites, and that this introduces a burden of showing that it is not a fallacy. This was to be met by showing that the predicate in question is not vague, but that

[20]It might be worried that the two cases are somewhat disanalogous, because 'is rich' is monadic, whereas 'is comparable with' is relational. But this is not a problem: the predicate in the Chaining Argument is 'is comparable *with Mozart*', which is monadic.

[21]Chang (2002), p. 665.

standard has not been met: vagueness has not been ruled out.

But I also wish to make a stronger claim. The vagueness in Fuzzy Hard Incomparabilism means that the Chaining Argument is relevantly similar to that which grounds other Sorites paradoxes. Sorites fallacies typically trade on the sort of vague boundaries that Fuzzy Hard Incomparabilism posits. As we have seen, the rich man argument illicitly trades on the vagueness of the predicate involved ('is rich'). Consider other classic Sorites arguments, such as those involving 'is tall', 'is a heap', or 'is bald' for example. Each of these is grounded in the fact that the predicate involved has vague boundaries, which is what Fuzzy Hard Incomparabilism claims for 'is comparable with Mozart'.

Let's focus on the rich man. In response to worries that this argument is a Sorites fallacy, we might (correctly) claim that these worries could be dispelled by showing that 'is rich' is not vague. But to carry this plan out, it would not do merely to show that a semantic indeterminist account of richness (according to which to be rich is to be a borderline-case of some other predicate) is false. We would need to show is that 'is rich' does not have vague boundaries. It is the analogue of this latter claim which would undermine Fuzzy Hard Incomparabilism, and which Chang's direct arguments do not establish.

The point is not just that there are two ways that 'is comparable with' could be vague—Semantic Indeterminism and Fuzzy Hard Incomparabilism—and that Chang's direct arguments rule out only one of them. The kind of vagueness which is *not* ruled out (that which is characteristic of Fuzzy Hard Incomparabilism) grounds the most common instances of the Sorites paradox, including the original: 'is a heap' has vague boundaries, but to be a heap is not to be a borderline-case of some other predicate. The direct arguments not only fail to rule out half of the sources of vagueness; they fail to rule out the more threatening half. In the next section, I'll present another reason to think the Chaining Argument is a Sorites—seeing it as a Sorites explains the undeniable plausibility of its key premise.

2.5 An Error Theory for the Difference Principle

Fuzzy Hard Incomparabilism can also explain the plausibility of the Difference Principle. Here are some quotes from Chang's defence of that principle:

> Now we bring to bear a key intuition ... between two evaluatively very different items, a small unidimensional difference cannot trigger incomparability where before there was comparability ... a small unidimensional difference just does not seem powerful enough to effect a switch from two such items being comparable to their being incomparable.[22]
> The Small Unidimensional Difference Principle has deep intuitive appeal. It is just plain hard to believe of two evaluatively very different but by hypothesis comparable items that making a small unidimensional improvement or detraction in one of them can thereby effect a switch from the items' being comparable to their being incomparable.[23]

> ... how can changing one of the careers ever so slightly in a single respect change the relation between the careers such that before there was an evaluative difference and now there is none—not even a zero difference?[24]

This consideration is by no means decisive, but this is noticeably similar to how one might motivate tolerance principles in the case of vague predicates: how could adding just £1 make a non-rich man rich? It's just hard to believe; the addition of just one pound doesn't seem powerful enough to effect such a change. Chang is correct that the Difference Principle has 'deep intuitive appeal', but if there is a boundary between the comparable and the incomparable zones (ie, if some version of Hard Incomparabilism is right), then the Difference Principle can be read as a statement that this boundary has vagueness.

---

[22]Chang (2002), p. 674.

[23]Chang (2002), p. 675.

[24]Chang (2002), p. 676.

I think that the best diagnosis of the Principle's intuitive appeal is that it is a *tolerance principle* for a vague predicate. As Crispin Wright has argued, tolerance ('a notion of a degree of change too small to make any difference, as it were'[25]) is a key characteristic of vague predicates that are vulnerable to the Sorites paradox. One millimetre couldn't make a short man not-short, one hair couldn't make a bald child not-bald, and so on. In the same way as the vagueness of 'rich' can make 'if a man with £n is rich, then a man with £(n+1) is rich' intuitively compelling, so the vagueness of 'comparable' on Fuzzy Hard Incomparabilism would be expected to make 'if Sn is comparable with Mozart, then S(n+1) is comparable with Mozart' compelling.

The key point is that the Principle would appear to be correct, even if it were just an artefact of the vague boundary of 'is comparable with Mozart'. If Fuzzy Hard Incomparabilism were correct, we would expect a tolerance principle identical to the Difference Principle to be intuitively overwhelming. The Difference Principle *is* intuitively plausible. But this intuitive plausibility adds to the suspicion that 'is comparable with Mozart' might have vague boundaries, and that Fuzzy Hard Incomparabilism might be correct.

2.6 Can Chang's Strategy be Generalised?

We've seen that the possibility of Fuzzy Hard Incomparabilism is one way that the Chaining Argument could trade on the vagueness of 'is comparable with Mozart', that this could explain the plausibility of the Difference Principle, and that arguments against Semantic Indeterminism are ineffective against this. But what about an adaptation of Chang's strategy—showing that 'is comparable with Mozart' also lacks *this* sort of vagueness? This would be to acknowledge that there is a burden of showing that 'is comparable with Mozart' does not have vague boundaries, and to try to meet it. This is not promising. It would be to show that the Hard Incomparabilism in question is not Fuzzy, but Sharp:

Sharp Hard Incomparabilism. Hard Incomparabilism is correct.

[25]Wright (1975), p. 333.

> Michelangelo is incomparable with Mozart. Every object is either determinately comparable with Mozart, or determinately incomparable with Mozart. There are no 'borderline-comparable' objects. The predicate 'is comparable with Mozart' does not have vague boundaries.

On this view, there would be some number k of small unidimensional improvements to Talentlessi, such that a sculptor with k improvements is worse than—and therefore comparable with—Mozart, but a sculptor with (k+1) is neither better than, worse than, nor as good as Mozart, and is therefore incomparable with him:

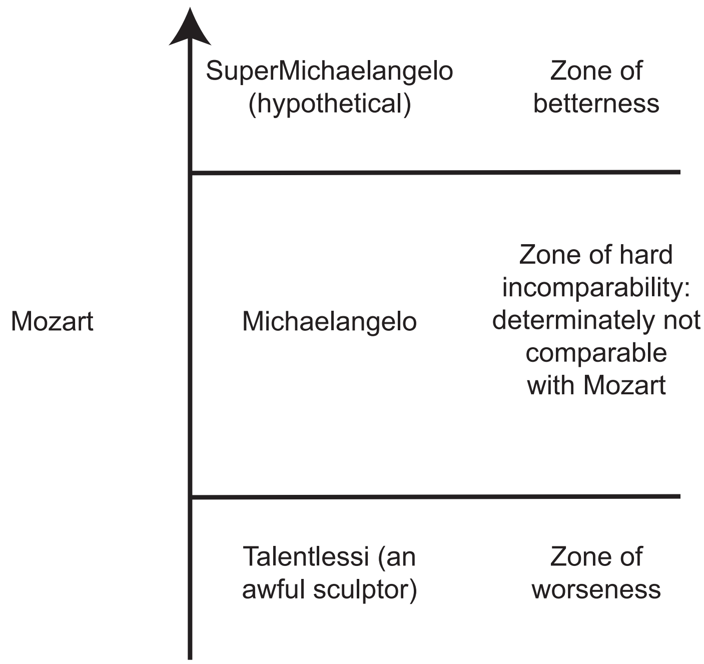

Figure 2.7: Sharp Hard Incomparabilism (unbroken horizontal lines indicate lack of vagueness).

Such a strategy *could* show that the Chaining Argument is valid: the Chaining Argument is certainly valid against Sharp Hard Incomparabilism. But this would be at the cost of completely defanging the Argument. The crucial premise of the Chaining Argument is the Small Unidimensional Difference Principle, which says that for every i, if

Si is comparable with Mozart, then S(i+1) and S(i-1) are comparable with Mozart.[26] The Difference Principle denies that 'is comparable with Mozart' could be sharply bounded. But the claim that there *is* such a sharp boundary is what is absolutely distinctive of Sharp Hard Incomparabilism.

If the Difference Principle is true, then Sharp Hard Incomparabilism is false. The real work of refuting Sharp Hard Incomparabilism would be in establishing the Difference Principle. But once that had been done, the Chaining Argument would be otiose, since it would add no additional weight to a rejection of Sharp Hard Incomparabilism. The Chaining Argument cannot be dialectically effective against Sharp Hard Incomparabilism, since a premise of the Argument—the Difference Principle—is incompatible with that view.

2.7 Conclusion

Hard Incomparabilism, which is the target of the Chaining Argument, must posit either vague or sharp boundaries to the zone of incomparability. If the boundaries are vague (Fuzzy Hard Incomparabilism), then the Chaining Argument has Sorites-structure, and uses the vague predicate 'is comparable with Mozart'. If they are not vague (Sharp Hard Incomparabilism), then the Chaining Argument is not a Sorites, but its key premise—the Difference Principle—immediately falsifies the target view, so the real work is done in defending the Difference Principle.

The apparent soundness of the Chaining Argument relies on eliding the distinction between the two versions of Hard Incomparabilism. But once this distinction is seen, it provides a dilemma for that argument. Either it is a Sorites fallacy (or should be treated as one, because the burden of showing that it is not has not been met), or it contributes nothing, dialectically, beyond the defence of the Difference Principle.

The Chaining Argument cannot fulfill its dialectical role of showing that on premises that everyone should accept—Mozart-Talentlessi comparability, the Existence Claim, the Difference Principle, and the trichotomous incomparability of Mozart and

[26] Officially, we should add an 'if they exist on the chain' qualifier here, to deal with the endpoints of the chain, if there are any.

Michelangelo—there is a good argument for Parity. The class of opponents against whom the Chaining Argument was directed—Hard Incomparabilists—can neatly explain away the apparent plausibility of its most crucial premise, the Difference Principle.

We would accept the Argument's key premise—the Difference Principle—only if we were already convinced of the falsity of both forms of Hard Incomparabilism. If we were not convinced of the falsity of Fuzzy Hard Incomparabilism, we would worry that the Difference Principle might be an artefact of the vagueness of 'is incomparable with Mozart'. If we were not convinced of the falsity of Sharp Hard Incomparabilism, we would worry that the Difference Principle is question-begging. So the Chaining Argument cannot offer any *additional* reasons to deny Hard Incomparabilism.

For all that, I have not shown that Parity is false, or even that the Chaining Argument is definitely an instance of the paradox of the Sorites. To show the latter, one needs the equivalent of the claim that '*of course* someone with £1m is rich, so the argument must have gone wrong somewhere'. Here that would be a claim that '*of course* Mozart and Michelangelo are not comparable ...', which is obviously not available in the current dialectic.

But I do not think the Chaining Argument can be rescued. Chang's paper is admirably 'modular': she carves up the space of views, and rejects each alternative, so that only Parity is left standing. The Chaining Argument's job is to refute the Trichotomy Thesis, and thereby Hard Incomparabilism. But the present discussion shows that the Chaining Argument cannot do this job. This job could be done by another argument, which would, together with the arguments against other views, also provide an existence proof for Parity. But it would not be the Chaining Argument.

3 Borderline Cases and the Collapsing Principle

Abstract. John Broome has argued that value incommensurability is vagueness, by appeal to a controversial 'collapsing principle' about comparative indeterminacy. I offer a new counterexample to the collapsing principle. That principle allows us to derive an outright contradiction from the claim that some object is a borderline-case of some predicate. But if there are no borderline-cases, then the principle is empty. The collapsing principle is either false or empty.[1]

When two options are such that neither is better than the other, and they are not equally good, and the comparison is not irrelevant or nonsensical, they are incommensurate:

Incommensurability. a and b are incommensurate (or incommensurable or incomparable) if it is not true that a is better than b, not true that b is better than a, and not true that a and b are equally good.

Often this is in terms of some specified standard or variety of goodness, such as artistic goodness or goodness as a career. I shan't argue that there are cases of incommensurability, but will simply assume it.[2] John Broome has influentially argued that incommensurability is vagueness: that in such cases, it is indeterminate whether a is

[1]I am indebted to Henrik Andersson, Simon Blackburn, John Broome, Erik Carlson, Patrick Connolly, Ram Neta, Wlodek Rabinowicz, Geoffrey Sayre-McCord, Keith Simmons, Walter Sinnott-Armstrong, Susan Wolf, and participants in the UNC dissertation seminar for discussion of this material and its ancestors. This paper is published in *Utilitas* (2014): volume 26, issue 01, pp. 51–60. I am grateful to Cambridge University Press for permission to reproduce this material here.

[2]This assumption—though plausible given classic 'Mozart/Michelangelo' cases—is not undeniable. For an excellent overview, see Chang (1997).

better, or b is better, or they are equally good. Then each of these possible comparative judgements is indeterminate. On this view, in cases of incommensurability, it is neither true nor false that a is better than b, neither true nor false that b is better than a, and neither true nor false that a and b are equally good.

Broome's argument relies on his 'collapsing principle' for comparative predicates:

> **Collapsing Principle (special version).** For any x and y, if it is false that y is Fer than x and not false that x is Fer than y, then it is true that x is Fer than y.[3]

Here 'Fer than' is a comparative predicate like 'better than', 'tastier than', or 'more impressive than'. The collapsing principle is intended to be a general truth about all such comparatives.

Erik Carlson has proposed several putative counterexamples to the collapsing principle, but Broome has argued that the principle itself rules those counterexamples incoherent.[4] Another line of response is proposed by Cristian Constantinescu, who has recently argued that the scope of the collapsing principle can be restricted, avoiding these counterexamples.[5]

In this paper, I present a new type of counterexample which is not vulnerable to either response. I show that the collapsing principle is either empty, or entails outright contradictions.

3.1 Broome's argument and Carlson's counterexamples

Incommensurability as vagueness is a rival to hard incomparabilism. On this view, if a and b are incommensurate, then it is false that a is better than b, false that b is better than a, and false that a and b are equally good.

Hard incomparabilism is defended most prominently by Joseph Raz, though Ruth

[3] Broome (1997), p. 74.

[4] See Carlson (2004). For Broome's reply, see Broome (2009), pp. 416-7.

[5] Constantinescu (2012).

Chang's 'parity' view also meets this definition.[6] Broome's argument for incommensurability as vagueness—and against hard Incomparabilism—rests on an argument that vagueness and hard incomparabilism are not compossible. Since it is not plausible, even for the hard incomparabilist, that there is no vagueness, this implies that there is vagueness but not hard incomparabilism.

I'll now sketch how the argument goes.[7] Suppose that hard incomparabilism is true, and we are in a rather fortunate choice situation:

> **Jobhunting.** We are comparing jobs, and in particular we are comparing a given philosophical job against a range of different banking jobs. The comparative predicate in question is 'better as a career', or simply better. Our preferences and other relevant factors are such that the philosophy job is clearly better than the banking job that pays $10,000 per year, and clearly worse than the banking job that pays $1m per year. But in this range are many banking jobs that are incommensurate with the philosophy job. Call this collection of jobs the Zone of Incommensurability (or just the Zone).

Assuming hard incomparabilism, for each banking job in the Zone, it will be false that the philosophy job is better than the banking job, and false that the banking job is better than the philosophy job, and false that they are equally good.

Now, suppose that the Zone has vague boundaries: in terms I develop elsewhere, assume that Fuzzy Hard Incomparabilism is true.[8] Pick a banking job in the top, vague, boundary region of the Zone. Call it the x-banking job. It will be indeterminate whether the x-banking job is in the Zone, and thus incommensurate with the philosophy job, or

[6]See Raz (1985-86) and Chang (2002).

[7]For the canonical version, See Broome (1997).

[8]See Chapter 2 for the distinction between fuzzy and sharp hard incomparabilism. The present discussion assumes that vagueness entails the presence of borderline-cases (at least that it does in the cases under consideration), and that some non-epistemic account of vagueness is correct. Neither assumption is problematic in this dialectic. The core of the incommensurability as vagueness view is that incommensurate options are comparative borderline-cases. It's hard to see how this could be sustained without the notion of a borderline-case. And if epistemicism is right, then the collapsing principle as presently formulated is empty or trivial.

‘above’ the Zone, and thus better than the philosophy job. So it is neither true nor false that the x-banking job is better than the philosophy job.

But is the x-banking job worse than the philosophy job? Clearly not: by stipulation, it is indeterminately either incommensurate or better. So it is false that the x-banking job is worse than the philosophy job.

Applying the collapsing principle, we conclude that it is true that the x-banking job is better than the banking job. If the collapsing principle is true, then the asymmetry introduced by hard incomparabilism crowds out vagueness. Hard incomparabilism and vague boundaries are not compossible.

Thus we must choose between (i) hard incomparabilism without vagueness (sharp hard incomparabilism) which posits completely determinate boundaries between the better and the incommensurate, and elsewhere, and (ii) a view on which there is vagueness but no hard incomparability: incommensurability as vagueness. Since, as Broome claims (and I agree), it is implausible that there is no vagueness whatsoever at the edge of the Zone, we should conclude that incommensurability is vagueness.

Clearly, this argument relies on the collapsing principle. Erik Carlson has offered several putative counterexamples to the principle.[9] These counterexamples rely on indeterminately relevant properties, where it is indeterminate whether a given property or its absence is relevant to the application of a predicate.

Here is the first. Suppose that we are comparing Alf and Beth with respect to philosophical ability. They are evenly matched in all relevant respects, except that Alf is slightly better rhetorically. It is clear that Alf is not worse than Beth – he can match her, virtue for virtue – so it is false that Beth is better than Alf. But is it true that Alf is better than Beth? Carlson claims that it might be indeterminate:

> perhaps our concept of a good philosopher is such that it is indeterminate whether rhetorical skill contributes positively to this species of goodness.[10]

---

[9]Carlson (2004).

[10]Carlson (2004), p. 223.

So, Carlson claims, it is possible that it is neither true nor false that Alf is better than Beth, but false that Beth is better than Alf, thanks to the indeterminately relevant property of rhetorical skill. If rhetorical skill is relevant, then Alf is better than Beth. If it is not, then they are evenly matched. Since it is indeterminate whether rhetorical skill is relevant, the situation is indeterminate between these two cases.

This – if coherent – is clearly a counterexample to the collapsing principle. Since it is false that Beth is better than Alf, according to the Principle it cannot be neither true nor false that Alf is better than Beth. The case relies on our judgement that it clearly is (or could be) indeterminate whether Alf is a better philosopher, thanks to the indeterminate relevance of rhetorical skill. The other example – is a waterproof alarm clock better? – has a similar structure.

We might say that the indeterminate relevance cases pit our semantic competence against the collapsing principle. We judge that the application of the predicate is indeterminate in this nonsymmetrical way (one option indeterminately better, but definitely not worse, than another), and the collapsing principle says it cannot be.

There are two main lines of defence for the collapsing principle. Broome has adopted a strategy of intransigence, arguing that the collapsing principle itself rules out indeterminately relevant properties:

> But the question is whether this sort of indefiniteness is really possible; the collapsing principle rules it out, and we are trying to assess the truth of the collapsing principle. [. . .] I see no reason to think there can be that sort of indefiniteness in the facts. At any rate, the example does not demonstrate there is; it assumes it.[11]

Alternatively, Cristian Constantinescu has suggested that the principle can be modified so that it only applies in the absence of indeterminate relevance, thereby dodging the counterexamples.[12]

[11] Broome (2009), p. 417.

[12] Constantinescu (2012), pp. 67ff.

These are both consistent, principled ways to defend the collapsing principle. Both rely on the fact that the counterexamples rely on indeterminate relevance: if that feature can somehow be ruled out, the principle is safe (albeit perhaps at some cost in either plausibility or scope).

In the next section, I'll describe a new kind of counterexample that does not depend on indeterminate relevance. As we'll see, this means that it is immune to both of these responses, leaving the defender of the principle in a far worse position.

3.2 A new counterexample to the collapsing principle

Consider the following comparative predicate on sets of men: set X is settaller than set Y just in case set X contains more tall men than set Y. Now, suppose that there are three sets of men:

Set A contains 10 tall men, and nothing else. Set B contains 10 tall men, 1 borderline-tall man ('the eleventh man'), and nothing else. Set C contains 11 tall men, and nothing else.

Now, compare A and B. It is false that A is settaller than B, since B contains at least as many tall men as does A. But it is not false that B is settaller than A, since it is indeterminate whether B has 11 tall men or 10. It is indeterminate whether (neither true nor false that) B is settaller than A.

Applying the collapsing principle, it is true that B is settaller than A. So far, so good. But let's apply the definition of settaller: it must be true that the eleventh man in B is tall.

But there was nothing special about the eleventh man. All we are given is that he is borderline-tall, that he is a member of set B, and that a comparison is made on sets including B. And via the collapsing principle, we conclude that he is tall. Being a member of a set, upon which a comparison is made, couldn't affect whether or not he is borderline-tall. So he must be tall, not borderline tall. So if the collapsing principle is true, there cannot be any borderline-tall men.

Next, we can repeat the recipe and compare set B and set C with respect to settaller. It will be false that B is settaller than C, and not false that C is settaller than B. By the collapsing principle, it is true that C is settaller than B, and so false that the eleventh man is tall. This is a contradiction: it is both true and false that the eleventh man is tall.

So the collapsing principle has allowed us to draw an outright contradiction from the claim that the eleventh man is borderline-tall. Now, this might not be so bad: there are those who have argued that borderlineness is a contradictory notion.

But this cannot be a good line of defence for the collapsing principle. The collapsing principle is concerned with restricting the forms that indeterminacy can take. In particular, the collapsing principle limits the possibilities for indeterminacy in comparative predicates. But comparative indeterminacy involves borderline-cases of comparative predicates. If there are no borderline-cases, then the principle is empty. Moreover, the project of showing that incommensurability is grounded in such borderline-cases is sunk: how could incommensurability be vagueness, if there is no vagueness?[13]

So this counterexample presents the defender of the collapsing principle with an unpleasant dilemma. Either the notion of a borderline-case is coherent, or it is not.

One one horn – borderlineness is coherent – the collapsing principle implies that there cannot be borderline-cases, and allows us to draw contradictions from the claim that there are. So the collapsing principle lets us draw outright contradictions from a coherent notion, and it must be false.

On the other horn – borderlineness is incoherent – the collapsing principle is empty, and the project it is used in support of cannot succeed.

3.3 Defence of the counterexample

Now I'll resist some natural objections to the counterexample. The first is that the predicate involved ('settaller') is too artificial. The second is that the inference from the truth of 'B is settaller than A' to the truth of 'the eleventh man is tall' is fallacious.[14]

Is the predicate 'is settaller than' objectionably artificial? This is not a promising line of objection. First, the predicate is not all that outre: there is nothing special about counting

---

[13]Again, I am assuming here that there could not be vagueness without borderline-cases.

[14]I am grateful to an anonymous reviewer for *Utilitas*, for catching a crucial mistake here.

the number of tall men in various sets. Moreover, the collapsing principle is intended to be fully general, and not limited to natural-language plausible predicates.

We can also construct more natural versions of the counterexample, at a cost of some numerical simplicity. Here is one. Suppose that I am comparing places with respect to better as a holiday destination. Amongst my other preferences, I strongly prefer to visit large countries: being a large country is a goodmaking feature of a holiday destination, for me.

We should be clear about two things. First, there is no indeterminate relevance here: whether a country is large or not is determinately relevant to whether I consider it a good holiday destination. It is a consideration that can be outweighted, perhaps, but it is always relevant.

Second, I don't have a general comparative preference for bigger countries: it is not that if country x is bigger than country y, then I necessarily have some (perhaps defeasible) preference to visit country x over country y. Instead, I am a 'country size snob': I prefer large countries to small ones. But within the large category, or the small category, size does not matter to me.[15] Russia and the USA are both big countries, although Russia is far bigger. But their size difference doesn't give me any (even defeasible) preference to visit Russia over the USA.

Suppose that, considering all relevant respects but size, Ireland, France, and China are equally good. Now, Ireland is clearly not large, China is clearly large, and France is borderline-large. Since they are on all fours otherwise, we can say that China is clearly better as a holiday destination than Ireland, China is borderline-better than France, and France is borderline-better than Ireland. Let's apply the collapsing principle.

Round 1. It is false that Ireland is better than France (since 'Ireland is large and France is not' is false), but not false that France is better than Ireland (since 'France is large and Ireland is not' is borderline). By the collapsing principle, it is true that France is better than Ireland. Given my preferences, it must be true that France is a large country. It could

[15] In terms I discuss in Chapter 5, the desire or project involved is wholly *binary*.

not have been borderline-large after all.

Round 2. It is false that France is better than China, and not false that China is better than France. Therefore, it is true that China is better than France. Given my preferences, it must be false that France is a large country. It could not have been borderline-large after all.

Contradiction.

So I don't think that the counterexample can be resisted on grounds of artificiality. What about the claim that if it is true that B is settaller than A, then the eleventh man in B is tall? This is also hard to resist. How could it be (i) true that A contains 10 tall men, (ii) true that B contains 11 men, and (iii) true that B contains more tall men than A, and yet (iv) not true that every man in B is tall? This seems untenable.

To talk in supervaluational terms: if it is (super-) true that B is settaller than A, then on every admissible sharpening of 'is tall', B contains more tall men than A. On every such sharpening, A contains 10 tall men. So on every such sharpening, B contains at least 11 tall men. But there are only 11 men in B. So on every such sharpening, every man in B, including the eleventh man, is tall. So it is (super-) true that the eleventh man is tall, by the supervaluational definition of (super-) truth.[16]

The counterexample seems immune to these two criticisms.

3.4 What about intransigence or scope restriction?

So the coherence of the counterexample shouldn't be doubted. Now let's see how the two strategies that have been used to defend the principle against Carlson's indeterminate-relevance cases – intransigence and restricting the scope of the collapsing principle – cannot work here.

Intransigence can't work, because the present counterexample does not pit one's semantic competence against the verdict of the collapsing principle. As we have seen, it is thereby not open to the defender to dig in and resist the intuition in question, since the 'intuition' is simply that outright contradictions are not permitted. What about claiming

[16] See Fine (1975).

that, just as indeterminate relevance is ruled out as incoherent by the collapsing principle, so are borderline-cases? As we've seen, this would render the collapsing principle empty, and so is not a viable strategy.

Unlike the indeterminate relevance counterexamples, this counterexample exploits a 'two-level' strategy. The top level consists of comparative facts about sets, and the bottom level of categorical (noncomparative) facts about their members. By carefully constructing some sets and a comparative on them, and applying the collapsing principle at the top level, we draw implausible or contradictory consequences at the bottom level.[17]

What about restricting the scope of the principle? To pull this off, we would need to find some way of isolating the cases that ground the counterexamples, and excluding them in a principled way. If the collapsing principle is to be defended by restricting its scope, it would be better if that restriction were not ad hoc. (Though I don't think it succeeds, Constantinescu's restriction of the principle to resist the indeterminate-relevance counterexamples is admirably principled.)

Now, it is true that there will be many cases where the collapsing principle will not lead to contradiction for a given comparison. The weak point is the inference in our present counterexample from 'Set X is Fer than Set Y' to the conclusion that some designated member of X or Y, previously thought to be a borderline-case of some predicate, couldn't be borderline after all. The counterexample relies on these 'designated candidates': the eleventh man, and France, for example.

But there are lots of cases where there will not be such a designated candidate, and thus an explicit counterexample (or contradiction) will not arise:[18]

> **The warships.** Say that a navy is more warlike than another just in case it has more warships than the other. Both the Royal Navy and the Marine Nationale have ten warships. Suppose also that this all the Marine Nationale has, but that the Royal Navy also has three borderline-warships. Then it is false that

[17] I'm grateful to Walter Sinnott-Armstrong for suggesting this way of putting it.

[18] I owe this point to John Broome.

> the Marine Nationale is more warlike than the Royal Navy, but not false (on the face of it, because indeterminate) that the Royal Navy is more warlike than the Marine Nationale.

Here, there is no designated candidate. Applying the collapsing principle, it is true that the Royal Navy is more warlike, so it must be true that the Royal Navy has at least one more warship. At least on many accounts of vagueness, this could be true that without it thereby being true that any particular or determinate member of the 'extra' British three ships is a (non-borderline) warship.

So could the collapsing principle be modified to take advantage of this, by restricting its application just to cases that lack designated candidates? This would be an adaptation of Constantinescu's strategy of restricting the scope of the collapsing principle.

But I don't see how this could be effective here. There doesn't seem to be principled way of fencing off these cases, and since there are so many sets containing more or less any object you care to consider, it's hard to see how the principle could be restricted without making its application very narrow indeed.

More or less anything that exists can be a member of a set, so we can run a similar trick on nearly any object that is a borderline-case of some predicate. Here is a general recipe: given an object v that is a borderline-case of some predicate F, consider three sets: (i) a set containing two clear cases of F, (ii) a set containing v and one clear case of F, and (iii) a set containing just one clear case of F. Now define the comparative 'set-Fer', which is such that a set is set-Fer than another just in case it contains more F objects. Then we apply the collapsing principle to set-Fer in the usual way.

So the collapsing principle will allow us to derive contradictions from very many seemingly innocuous claims that so-and-so is a borderline case of some predicate. The scope of the collapsing principle cannot be usefully restricted: to avoid this recipe, as well as cases like that of the holiday-destination, any scope restriction would have to be extremely severe. But if the collapsing principle has such tightly restricted application, it cannot ground a general conclusion about value incommensurability, which is a very widespread phenomenon.

So it seems that neither digging in, nor restricting the scope of the collapsing principle, can work to defend it against these sorts of counterexample. The collapsing principle is false.

3.5 Conclusion

Broome's collapsing principle grounds a neat argument for incommensurability is vagueness. I have argued that it is vulnerable to a counterexample, and thus that it is false. This also provides independent support for the view that Carlson's examples really are coherent, and that they really are counterexamples to the collapsing principle: the main reason for thinking that there is something unwholesome about indeterminately relevant properties was that the collapsing principle seemed to rule them out.

But what about incommensurability as vagueness? My own view is that it is the most plausible account of incommensurability. However, the argument from the collapsing principle – though strikingly neat and direct – cannot establish its truth. Incommensurability as vagueness must be defended via a less elegant, but (nearly) as decisive, consideration of its overall costs and benefits, a task which I attempted in Chapter 1.

4 'Ought' and Indeterminacy

Abstract. I describe an answer to the question, 'what ought one do when it's indeterminate what one ought do?', at least in one-shot cases (sequential action requires a separate treatment). Unlike other recent accounts of action under indeterminacy, the view I defend is orthodox and maximising: we ought to maximise an indeterminist analogue of expected utility. Such a 'hyperexpected utility' maximising view can be seen as an extension of orthodox utility-maximising to cover indeterminacy, and as a response to recent arguments that no such extension is possible. This view is distinctively tolerant of compromise: in some rare cases, what we ought do is determinately not the best, and determinately what we have most reason *not* to do. I describe cases where this seems inevitable. Finally, this provides a decision theory to accompany the view that incommensurability is vagueness, and an error theory for John Broome's 'collapsing principle', which he defends by appeal to decision-theoretic considerations.[1]

4.1 Introduction

Vagueness manifests as borderline-cases—cases where a vague predicate neither clearly applies nor clearly fails to apply—and this can have puzzling decision-theoretic implications. A simple example will make this clear:

> **Press-gang.** All adult males must join the Navy. Tall men must serve in the surface fleet, and all others must serve on submarines, where space is at a

[1] For comments and discussion, I am particularly grateful to Simon Blackburn, Matthew Kotzen, Ram Neta, L.A. Paul, Susanna Rinard, John Roberts, Geoffrey Sayre-McCord, Miriam Schoenfield, Keith Simmons, Walter Sinnott-Armstrong, Jesse Summers, J.R.G. Williams, and several audiences at UNC Chapel Hill.

premium. If Achmed is tall, then he ought go to the surface fleet; if Achmed is non-tall, then he ought go to the submarine fleet. But 'is tall' is vague, and Achmed is borderline-tall.

The central puzzle is this: what ought Achmed to do? In a more prosaic case, suppose that you are classifying library books, and though it is clear that some tome ought to go in the history of philosophy section, it is vague whether it should be classified as *Medieval* or *Renaissance* philosophy.[2] What ought you to do? Intuitively, if it really is indeterminate which section to put the book in, you ought to make some kind of arbitrary choice.

Similarly, Achmed ought to choose arbitrarily (in some sense). We can immediately note several things about Press-gang. First, there seems to be something distinctive about the call to *action* here: if the question were simply 'is Achmed tall', then one might answer that the answer was vague or indeterminate or borderline, or simply remain silent. But here, there is no action equivalent to withholding judgement: he must act one way or another. This 'must' can arise in two main ways: it might be that all other actions (including inaction) are determinately *worse* than the two he must choose between, or it might be that inaction is itself one of the options open to him.

Second, there is only really a distinctive issue here if epistemicism—the view that vagueness is merely ignorance—is false. If vagueness is ignorance, then Achmed is either determinately tall, or determinately non-tall, though we don't know which. There is no particular puzzle about action under ignorance: in the simplest terms, one ought maximise expected utility. If vagueness is indeterminacy, then giving a decision theory for action under vagueness is a *distinctive* project.

Third, the case under discussion involves 'one-shot' choices: Achmed goes to the surface, or to the submarine. But vagueness also seems to raise problems involving *sequential* choices: suppose that you are responsible for sending men to the various fleets, and you arbitrarily choose to send Achmed to the surface. Now comes Bobby, who is 5mm taller than Achmed, but also borderline-tall. On the face of it, the same

[2]Compare Edgington (1997), p. 310.

considerations apply as did with Achmed, but you seem to be constrained by your earlier decision: though it might be vague whether you ought send Achmed (x cm) to the surface, and vague whether you ought send Bobby ($x + 0.5$ cm) to the surface, it is plain that you ought not send Achmed but not Bobby to the surface. You don't know whether Achmed is tall, or whether Bobby is tall, but you do know that if Achmed is tall, then so is Bobby.

This engenders the question: how can your earlier choice constrain your later choice, though by stipulation it has no effect on Bobby's height, or on the rules under which you are working? Considerations of sequential action in particular have motivated a recent interest in the topic of action under vagueness. But these accounts have tended to be rather revisionary. For example, Quinn (1990) defends intransitive preferences by appeal to vagueness, Tenenbaum and Raffman (2012) argue that vagueness engenders rational permissions and requirements to diverge from momentary expected utility calculations, and Williams (2014) answers a puzzle like that of Achmed and Bobby by arguing that an agent can treat the fact that she chose an option as evidence that *that* option was better, even knowing that her initial choice was arbitrary.[3]

The answer I'll give is far less radical than its competitors: one ought maximise (the analogue of) expected utility. I'll extend orthodox utility maximising accounts of rationality to cover indeterminacy, by developing the notion of *hyperexpected utility*: expected utility calculated not across ignorance, but indeterminacy. I'll show that this view renders the intuitive view about several puzzling cases of action under indeterminacy, including that of Achmed's choice of naval career.

This view has several advantages. First, it is the only theory that is neutral between epistemicism and indeterminism about vagueness. As we saw, epistemicism brings its own (maximising) decision theory, so any indeterminist about vagueness who wishes to capture the verdicts of that view almost certainly must appeal to something like hyperexpectations. (It's not impossible that another indeterminist view with a different

[3]For other recent work in this area, see Broome (1997); Moss (forthcoming[a]); Rinard (forthcoming), and discussions of 'unsharp credences'. See Elga (2010) for more on the latter.

structure will render the same verdicts, but this seems highly unlikely.) This may carry more force for those of us who find the epistemicist verdicts about action in such cases more plausible, but are not epistemicists. If one wishes to capture the epistemic prescriptions for action without joining the epistemicist camp, the hyperexpectation-maximising view seems unavoidable.

Second, the view supplies a decision theory for the view (defended in chapter 1) that value incommensurability is vagueness. The view that cases of incommensurability are borderline-cases of comparative evaluative predicates is somewhat naked without a decision theory, not least because rendering an intuitively-plausible account of rational choice under incommensurability is one of the main goals of a theory of that phenomenon. On the hyperexpectation-maximising account, incommensurability as vagueness can render plausible verdicts whilst remaining neutral about the nature of vagueness: that in general, it is permissible to pick arbitrarily between incommensurate items. As we'll see below, the view also serves a purpose in the incommensurability dialectic: John Broome defends incommensurability as vagueness by appeal to the Collapsing Principle, but I argued in chapter 3 that that principle is false. The present view will allow us to see where his argument for the Collapsing Principle is unsound.

Finally, this non-revisionary view will form the basis for an account of *sequences* of actions under indeterminacy. As I argue elsewhere, once the expectation-maximising framework is established for one-shot actions, it can be naturally extended to the sequential case. The central idea is that it is not normally an indeterminate matter whether a given action maximises hyperexpected utility, so forbidden sequences of actions can be ruled out under indeterminacy in much the same way as they are under ignorance.[4] But first, we need to work out a theory for the one-shot case; to begin, I'll work out the conception of rational action that we'll be assuming for full determinacy.

In this paper, I am concerned with the one-shot case. The sequential puzzle is interesting, but distinct—and it's important to walk before one can run, so to speak. So

[4]See Elson (Manuscript[a]).

focusing on action under indeterminacy in single actions, our question is: in a one-shot action, what ought one to do when it's indeterminate what maximises expected utility? Such indeterminacy can be located in indeterminate preferences, indeterminate credences (thus indeterminate expectations), and ordinary predicate vagueness (such as that of 'is tall').

4.2 The Ought-Reasons Platitude, and Maximising

A good place to start is the *ought-reasons platitude*—that one ought to do what one has most reason to do—which enjoys almost unanimous philosophical assent, and was recently described by Douglas Portmore as 'perhaps the most fundamental and least controversial normative principle concerning action'.[5] Given this framework, the question we'll be be concerned with might put like this: when it's indeterminate what one has most reason to do, how ought one act?

But first, we should explore how the platitude works in cases of full determinacy. The platitude can be restricted to certain sorts of reasons, as we'll see in the following example. Suppose that you are fishing, far from the shore, and you come across two boats in difficulty. They will both sink if you don't help, but you can help only one of them. All else being equal, you have most moral reason to rescue the boat with more people aboard. You can see that one boat (to the West) carries 100 people, but the other boat (to the East) carries just five. It is then intuitively clear that you have most moral reason to rescue the Westerly boat, and that you morally ought to do so. It's hard to see how it could be that you could have more moral reason to do one thing, and yet you morally ought to do another.

We might also say that you prudentially ought to do what you have most prudential reason to do, and so on. For example, if your life-savings were on the Easterly boat, you plausibly have most (self-interested) prudential reason to help the Easterly boat, and that you prudentially ought to help that boat. The exact nature of the limitation to just one sort of reason—is there just one 'ought' that is relativised to different sets of reasons, or

[5]Portmore (2013), p. 437. Some metaethical views distinguish between *there being a reason* and *having a reason*, but this issue is orthogonal to our concerns here.

are there several 'oughts'?—is interesting, but need not detain us. In his recent paper, J.R.G. Williams neatly focuses on prudential concerns by considering a (stipulatively) wholly self-interested agent, making decisions under conditions of indeterminate personal identity.[6]

Questions may also be asked about the status of the platitude: is it merely extensional, or somehow more metaphysically robust? Mark Schroeder, for example, seems to defend something rather strong: 'oughts are analysed in terms of reasons rather than conversely. You ought to do what you have most reason to do.'[7] But once again, this is an issue that we can sidestep, because our puzzle will arise given a merely extensional principle, together with appropriate indeterminacy on one side of the conditional.

In that vein, the platitude may be a biconditional, or a mere material conditional. One direction of the platitude—if A has most reason to ϕ, then A ought to ϕ—seems even more plausible than the other.[8] Though I find the fully general biconditional most plausible, we can nevertheless restrict ourselves to utility-maximising contexts, where the full biconditional is overwhelmingly plausible. Thus we'll consider a gloss on the platitude where 'most reason' is understood as 'maximises utility':

Reasons-Ought. A ought to ϕ if and only if ϕing maximises utility for A.

But in practice, this won't be the version of the platitude that is most useful for our purposes. There are cases—most famously, that of Buridan's Ass—where two options are equally supported by reason or, in other words, have equal utility. In the case of the boats, imagine that each boat carried fifty people. Here we might speak of a 'disjunctive ought'—the Ass ought to eat either bale of hay, and you ought to rescue the westerly boat or the easterly boat—and say that you have done as you ought just in case you rescue one of the boats. But this locution is awkward, because we'll have to say that though if he eats

[6]Williams (2014), especially p. 3.

[7]Schroeder (2007), p. 193.

[8]I'm grateful to Walter Sinnott-Armstrong for discussion of this point.

the first bale of hay, the Ass has done as he ought, but not that *he ought to eat the first bale of hay*, if the latter implies that he ought not do anything else. So I'll largely frame matters in terms of *permissibility*: it is permissible for you to rescue the westerly boat, permissible for you to rescue the easterly boat, and impermissible to do anything else. Buridan's Ass may permissibly eat either bale of hay.[9]

Here we'll use a stipulative definition of 'permissible' as *not-ought-not*. Combining this with Reasons-Ought: it's permissible for A to ϕ if and only if ϕing realises at least as much utility as any other action available to A. That is, if and only if ϕing weakly maximises utility for A.

Now—in our final clarification—we can note that sometimes we must act without knowing all the facts. In such case we distinguish between *objective* 'ought' claims, which take account of the actual facts, and *subjective* 'ought' claims, which take account only of what we believe, or the information available to us. A similar distinction is made between kinds of reasons. Suppose that in the case of the boats, you mistakenly believed that there were one hundred people in the easterly boat, and just one in the westerly. On that basis, you rescued the easterly boat, before realising your mistake: someone had switched the passenger manifests. Though your actions might be grounds for regret, we would not criticise you, unless perhaps the error was due to some culpable epistemic failing. Though objectively you ought have done something else, you did as you subjectively ought, and your action was therefore subjectively permissible.

We'll therefore treat Reasons-Ought as a plausible way of identifying which actions are permissible for an agent, given her reasons. In the maximising contexts we are considering, an action ϕ is objectively permissible if and only if ϕing maximises utility, and subjectively permissible for an agent if and only if ϕing maximises *expected utility* for that agent. In full-information cases, these two notions will of permissibility will run together, since the agent knows all the facts.

[9]For discussion of permissibility and disjunctive 'ought' claims, and help in getting my view straight, I'm grateful to Matthew Kotzen, Daniel Layman, Geoffrey Sayre-McCord, Keith Simmons, Walter Sinnott-Armstrong, and Jesse Summers.

But since we are here focusing on indeterminacy, we won't consider cases that have limited information together with indeterminacy. So we won't have much cause to consider the notion of subjective permissibility, until we see later that hyperexpectation-maximising introduces a similar notion. For now, we are working with a principle of limited scope: the permissible actions for an agent are those (and only those) which bring at least as much utility as any other. As we'll see in the next section, this principle is not plausible in the presence of indeterminacy.

4.3 Indeterminate-permissibility

Before seeing how indeterminacy makes trouble for Reasons-Ought, it will be worth saying something about the kind of indeterminacy in question. I aim to be as theoretically neutral as possible. When it is indeterminate whether P, it is indeterminate which of the *sharpenings* $\{s_1, \ldots, s_n\} = S$ obtains, and on some sharpenings in S, P is true, and on some sharpenings in S, P is false. Depending on one's account of indeterminacy, one may say that P is 'indeterminate', 'neither super-true nor super-false', or even 'neither true nor false'. Sentences which are true on every sharpening $s_k \in S$ are (determinately, or super-) true, and those which are false on every sharpening are (determinately, or super-) false.

Consider our original example of a vague predicate, 'is tall'. On this view of indeterminacy, there is no precise minimum number of millimetres required to count as tall: there are a number of *admissible sharpenings* of the predicate 'is tall', each corresponding to a precise way of drawing the tall/not-tall boundary, consistent with our judgements about clear cases of that predicate. So it is indeterminate whether Achmed is tall, because on some admissible sharpenings (ways making 'is tall' wholly precise, consistent with our judgements about clear cases), he is, and on other sharpenings, he is not. But it is determinately true that Andre the Giant (7ft, 4in, according to Wikipedia) was tall, because his height is tall on all admissible sharpenings of the predicate.

Now, we may be able to 'count' sharpenings: sometimes, we are inclined to say that some sentence is indeterminate, but 'closer to true than to false'. Depending on the notion of indeterminacy in play, this may be because it is true on more sharpenings than it is false on. We have spoken here as if there are always a finite number of sharpenings, but there is no clear reason why this must be so. If there are infinitely many sharpenings, we must not speak of 'counts' but of 'measures'.[10]

Now we can see how indeterminacy interacts with Reasons-Ought: if it is indeterminate whether some ϕ maximises utility, then it is indeterminate whether ϕing is permissible. The platitude implies what whenever it is indeterminate which action one has most reason to do, it is indeterminate which action one ought to do, and indeterminate which actions are permissible. This is because Reasons-Ought is a biconditional, so if one side is indeterminate, then so is the other.

This may not seem particularly troublesome, and may be intuitive, so why not accept it? In other words, why think that the ought-reasons platitude needs to be modified to cope with indeterminacy? We could adopt what I'll call *Ought-Indeterminism*, and argue that any indeterminacy in (most-) reasons generates indeterminate permissions. Susanna Rinard defends such a view.[11] But (I'll argue) Ought-Indeterminism wrongly classifies many actions that are intuitively permissible as indeterminately-permissible, and grounds a class of what I'll call *indeterministic rational dilemmas*, where all actions available to an agent are (at best) indeterminately-permissible.

We can see this if we consider a case of 'flat' or 'symmetrical' indeterminacy, where it is simply indeterminate whether ψ or ϕ maximises expected utility but neither has any sort of truth or determinacy advantage. If we are prepared to talk in terms of degrees of truth, 'ϕ maximises expected utility' and 'ψ maximises expected utility' are each true to degree 0.5: speaking loosely, half of the sharpenings recommend ϕing, and half recommend ψing. Such cases cause trouble for Ought-Indeterminism:

> **Mushy Boats.** You can rescue only one of two boats. All else is equal, so you ought to rescue the one with more people. It is impermissible to rescue neither (there is no personal danger, and no other complicating factors). It is indeterminate which boat you have more moral reason to rescue: on half of

[10] For interesting discussion of degree theories, see for example Williams (2010); Edgington (1997).

[11] Rinard (forthcoming); John Roberts has also argued for such an account, in conversation.

> the sharpenings, the westerly boat carries more people, and on the other half, the easterly boat does.[12]

The intuitive verdict in this case is that it is (morally) permissible for you to rescue either boat, and impermissible for you to do anything else. What more could be morally demanded of you, than that you rescue one of the boats? In particular, from a moral perspective at least, one can select for no reason, 'pick', or 'plump'.[13] Simon Blackburn puts the point particularly neatly:

> [...] both common sense and high theory tell us how to handle it. The agent has to *plump* for one alternative. I say 'plump' deliberately, because saying that you have to choose carries a bad implicature. Choice is a process that invokes reasons.[14]

But, I think, Blackburn is too sanguine about high theory. Let's see how this plays out with Ought-Indeterminism. Once again, it is (objectively) morally permissible only to rescue the more-populated boat. But it is indeterminate which boat is more populated. Whichever of the two boats you rescue, Ought-Indeterminism must say that it is indeterminate whether you have done as you morally ought, and whether you have acted in a morally impermissible way. This is the best that you can do. If you fail to rescue either boat, then you have definitely done something impermissible. There is no option available to you that is determinately morally permissible.

This is a (moral) *dilemma of indeterminate rationality*. It is impossible for you act in a determinately permissible way. But to judge that situations like the mushy boats always lead to dilemmas of indeterminate rationality seems an unduly harsh verdict, at odds with our everyday evaluations. Part of the project of this paper is trying to reconcile the

[12]I am grateful to Miriam Schoenfield and Jesse Summers for discussion of a variant of this case. If you find the indeterminacy in the setup implausible, consider a case where their passenger numbers are determinate, but your *evidence* is indeterminate.

[13]In the language of Ullmann-Margalit and Morgenbesser (1977) and Blackburn (2010), respectively.

[14]Blackburn (2010), p.50.

high theory that is Reasons-Ought with the piece of common sense that when it is indeterminate which boat is more populated, it is permissible to rescue either. If they cannot be reconciled, then Reasons-Ought must be abandoned.

My claim—which may be denied— is not merely that though it is indeterminate whether you acted permissibility, whichever boat you rescued, it would be churlish to blame you. There may be such cases, since impermissibility and blameworthiness can come apart. But here Ought-Indeterminism seems to get wrong the *prior* verdict about (im)permissibility. In cases like this, we don't don't judge that you did something borderline-permissible. As well as failing to capture our intuitions, a theory that makes such dilemmas commonplace fails for a deeper reason: it is at odds with the plausible principle that rationality (at least one kind of rationality) is doing as best you can in the circumstances you face.

Note that there are two separate issues here: first, which actions are permissible? second, what are the admissible *decision procedures*? It is with the first question that I am here concerned. The second question is derivative: it is permissible to pick or plump *because* either option is permissible.

Epistemicism about vagueness will render the commonsense verdict that whichever boat you save, you act permissibly. On that view, you are ignorant of which boat holds more people (there is a determinate fact of the matter), and it is equally likely that each boat is the more populated. So you may permissibly rescue either—either choice is akin to a lottery ticket with a fifty percent likelihood of winning. Here, as often, epistemicism seems to render the right decision-theoretic verdict, and indeterminism the implausible one. Since indeterminism about vagueness is widely considered more plausible than epistemicism, this is an uncomfortable situation.

Not that there's anything incoherent about dilemmas of indeterminate rationality. Perhaps we are just too squeamish and they are, like moral dilemmas, a fact of life.[15] Sarah Moss has recently argued for the existence of 'credal dilemmas' by appeal to their

[15] In conversation, John Roberts has suggested this line of reasoning to me.

moral analogue.[16] It's not my purpose here to criticise Moss's view—her distinctive account is aimed more at the phenomenon of unsharp credences, and has the agent 'identifying' with a particular sharpening of her credences, so it's not clear that such an agent has reasons credences analogous to those under consideration—but dilemmas of indeterminate rationality differ significantly from moral dilemmas, in two ways.

First, in many putative moral dilemmas, our intuitive verdicts form the most compelling evidence that the situation is dilemmatic. Whatever option the agent in question chose, we think, they did something morally wrong: consider Sartre's student who must choose between fighting for France and caring for his mother. Some cases just seem dilemmatic, and this presents an explanatory burden for any denial that dilemmas could exist. But our verdict in cases like the mushy boats is not like this: we are (at least, I am) inclined to say that the agent 'did the best that she could', or some such, and is thus not subject to criticism.

Second, moral dilemmas are *rare*. But if Ought-Indeterminism is correct, then dilemmas of indeterminate rationality are everywhere. It's not obvious that the undeniable plausibility that attaches to particular (putative) moral dilemmas will extend to a commonplace class of rational dilemmas.

In moral cases, the indeterminacy posited by Ought-Indeterminism may also ramify to other agents. Whenever it is indeterminate what an agent has most reason to do, it will be indeterminate what she ought to do, and thus indeterminate what actions are permissible for her. But if it is indeterminate whether you have committed murder, and thus indeterminate whether you have done something morally impermissible, it might be indeterminate whether I have most moral reason to blame you, and indeterminate whether I ought to blame you, and indeterminate what attitude a third person ought towards take to my blaming of you, and so on. This does not falsify the view, but it should be borne in mind.

So Ought-Indeterminism—the ought-reasons platitude applied without modification

[16] Moss (forthcoming[a]).

in cases of indeterminacy—seems to engender counterintuitively widespread dilemmas. Its ‘threshold’ for permissible action seems to be too high, so to speak: in order to act permissibly, it must be determinate that one does what one has most reason to do.

But if those cases involved an excessively demanding threshold, there are also cases where Ought-Indeterminism has a threshold that is too *low*. It judges actions as indeterminately-permissible when, intuitively, they are clearly impermissible.

Though we’ll eventually disagree with him about their upshot (as expressed in the final quoted sentence), John Broome’s intuitions about the following case seem undeniable:

> **Canberra Suburbs.** Suburbs in Canberra are named after great Australians, and each new suburb has to go to the greatest Australian who does not yet have a suburb. Suppose there are two candidates for the next suburb, and you have to decide between them. Suppose that, on investigating their cases, you conclude it is false that Wye is a greater Australian than Exe, but that it is not false that Exe is a greater Australian than Wye. [. . .] you need not hesitate. It would be quite wrong to give the suburb to Wye. Since the prize was for being the greater Australian, it could not be so obvious who should win unless that person was the greater Australian.[17]

In this case, the world is indeterminate between two situations. In other words, there are two sharpenings: (i), Exe and Wye are equally great, this is a Buridan’s Ass situation, and choosing either is permissible; or (ii), Exe is greater than Wye, and you ought to choose Exe. There are no sharpenings on which Why is greater than Exe; it is determinately false that Wye is greater than Exe.

Let’s agree with Broome that you ought to choose Exe, and that choosing Wye is impermissible. But Ought-Indeterminism cannot explain or capture this judgement. It is indeterminate whether there was most overall reason to choose Wye, since it’s

[17]Broome (1997), p. 74. Here ‘not false’ means ‘not determinately false’.

indeterminate whether the Buridan's Ass case (i) obtains. So according to Ought-Indeterminism, choosing Wye is indeterminately-permissible.

Given these problems, I think that Ought-Indeterminism must be abandoned. In the next section, I'll consider a class of natural alternatives.

But note that nothing we have said undermines a weaker version of the platitude:

> **Determinate Reasons-Ought.** If A *determinately* has most reason to ϕ, then A ought to ϕ.

Respecting Determinate Reasons-Ought is a criterion of adequacy on any account of action under indeterminacy. Much of the intuitive support for Reasons-Ought—since that support consists of intuitions in cases of full determinacy—is really intuitive support for Determinate Reasons-Ought. So we shouldn't be too hasty in assuming that the principle supported by that evidence remains true under conditions of indeterminacy; however, any more general view that also applies to indeterminate cases must collapse into Determinate Reasons-Ought in cases of full determinacy. Ought-Indeterminism satisfies this constraint, but it seems to render the wrong verdict in various cases.

4.4 Caprice and Randomize

One popular class of views is what I'll call *capricious* views. Their central claim is that—perhaps subject to certain additional constraints—being favoured on one or more sharpening makes an action rationally eligible in some sense, and thus permissible.

Brian Weatherson defends the eponymous decision rule:

> **Caprice.** An isolated action is currently permissible for you just in case there is some sharpening such that that action has the highest expected utility according to your current utility function and your current conditional credence function, conditional on that sharpening being correct.[18]

Caprice can be put in terms of non-dominance: it is permissible for A to ϕ if and only if there is at least one sharpening on which A has most reason to ϕ, and thus ϕing is not

[18] This formulation is due to Moss (forthcoming[b]), p. 10.

determinately dominated for A. Compare Roger White's candidate rule 'Liberal':

> **Liberal.** Take any set of bets that maximizes expected utility according to *some* credence function in your representor.[19]

These rules seem to capture the sense that in cases like the mushy boats, our reasons and other normative guidance have somehow 'run out', and there is nothing left to do but pick arbitrarily. Rationality merely requires that agents avoid *determinately unjustified* actions.

Caprice scores well on a number of fronts. Like Ought-Indeterminism, Caprice respects Determinate Reasons-Ought: if it is determinate that I have most reason to ϕ, then ϕing maximises utility on every sharpening (by the definition of determinacy), so in particular it is permissible on at least one sharpening, so it is permissible for me to ϕ (by Caprice).

Caprice also renders our target verdict about the mushy boats: choosing either of the designated options is permissible, but choosing any other is impermissible. This is because on half of the sharpenings, you have most reason to rescue the westerly boat; on half of the sharpenings, you have most reason to rescue the easterly; no sharpening recommends any other action.

Caprice combines two claims. First, there is a *sufficiency* claim that if an action is mandated on at least one sharpening, then it is permissible:

> **Capricious Sufficiency.** If it is not determinate that A has most reason not to ϕ, then it is permissible for A to ϕ.

Capricious Sufficiency is what enables Caprice to render the right verdict in the mushy boats. In general, Capricious Sufficiency says that if it is *true that* or *indeterminate whether* an action maximises utility, then that action is permissible.

Capricious Sufficiency can be strengthened in various ways. Most naturally, we might want to account for the case where many more sharpenings mandate one course of action

[19] White (2009), p. 178. Note that this is concerned with unsharp credences, not vagueness *per se*.

than do another. JRG Williams has recently argued that we must 'replace/supplement' Caprice along these lines:

> **Randomize.** Given a decision situation in which one must choose to ϕ or ψ, if k sharpenings recommend ϕing, and $1 - k$ ψing, then one should choose at random – ϕing with chance k and ψing with chance $1 - k$.[20]

This seems a sensible modification to account for any cases where rather more sharpenings support one action than do another. In general, 'the Randomize rule will assign a chance of ϕing equal to the measure of sharpenings that induce mental states that recommend ϕing'.[21]

We must tread carefully here: Randomize is, unlike Caprice, a rule not only about which action is taken, but about the right *decision procedure* for that action. That is why it is not entirely clear whether it is a replacement for, or a supplement to, Caprice.

Though Caprice handles the mushy boats admirably, I think that Randomize renders a curious verdict about this case: that your rescuing one of the boats was morally permissible only if you chose randomly which boat to rescue. But this is, I think, simply an echo of Williams's regarding the question of decision procedure as somehow primary.

But neither Caprice nor Randomize can render the correct verdict about the Canberra Suburbs. Since there is at least one sharpening on which you have most reason to choose Wye, Caprice claims that if you choose Wye, you act permissibly: In fact, there seems to be no way of saying even that choosing Exe is preferred. Randomize will not help us with cases of indeterminate superiority. Randomize says that you act permissibly if you choose which (class of) sharpenings to act on at random, and:

> If Randomize selects a non-singleton set as the operative recommendation, then the agent's options are whatever they would be in the ordinary case — perhaps either act is permissible (I do not want to assume that here the agent

[20] Williams (2014), p. 11.

[21] Williams (2014), p. 11 fn. 22.

needs to randomize — perhaps she can exercise discretion here on a more voluntaristic basis).[22]

Without further modification, Randomize too cannot account for the fact that it is determinately impermissible to choose Wye.

We should be clear here that the problem is Capricious Sufficiency: it is not determinately false that choosing Wye maximises utility, yet it is clearly impermissible to choose Wye. Capricious Sufficiency must somehow be strengthened or replaced to avoid this. I'll now turn to some suggestions for how that might be done.

4.5 Comparative Capricious Views

The problem with the views discussed in the previous sections is that they set incorrect thresholds for permissible action; I think that the relevant constraint is *comparative*: how does the action in question compare to other available actions?

In his recent discussion of the Canberra Suburbs and related cases, Johan E. Gustafsson suggests a comparative constraint:

> **Avoid indeterminate worseness.** If possible, choose an option x such that it is determinate that no option is better than x.[23]

Recast in terms of permissibility: if it is determinate that for every action ψ, ϕing realises at least as much utility as ψ, then one ought to ϕ. This suggestion is, I think, correct: it requires that one not choose an action that is indeterminately what one has most reason to do, when there is another option that is *determinately* supported by reason available. The comparative aspect here is that an action ϕ which is mandated by some but not all sharpenings may be permissible only if there is no other action x that is mandated by *all* sharpenings. This constraint captures what needs to be done to render the right verdict about the Canberra Suburbs.

These comparative considerations suggests that rules like Capricious Sufficiency are

[22] Williams (2014), p. 32.

[23] Gustafsson (2013), p. 438.

on the wrong track: what matters to the permissibility of an action ϕ is not just that there is some sharpening (or some portion of sharpenings) on which ϕ maximises utility; it also matters how ϕ does on the *other* sharpenings. In the case of the Canberra Suburbs, it seems to matter not only are there some sharpenings that mandate choosing Exe, but that even on those sharpenings which mandate choosing Wye, Exe is *no worse* than Wye.

As a first step, consider a view that I used to defend:

> **Sharpening-Maximize.** ϕing is permissible if for any other action ψ, there are at least as many sharpenings that support ϕing, as there sharpenings are that support ψing.

A quick gloss on Sharpening-Maximize is that one ought to take the option ϕ, such that it is 'most true' that one has most reason to ϕ. This is a substantial strengthening of Capricious Sufficiency: many actions which are permissible according to that principle are not permissible according to Sharpening-Maximize. For instance, since there are more sharpenings that mandate choosing Exe than there are that mandate choosing Wye, it is only permissible to choose Exe.

Sharpening-Maximize is obviously a close cousin of Randomize, but it renders a different (and to my mind, correctly so) verdict in cases of indeterminate superiority. In less outré cases, the difference between a Randomizer and a Sharpening-Maximizer will manifest in long-run frequencies. If 60% of sharpenings support ϕing and 40% support ψing, then in the long run the Randomizer will ϕ 60% of the time, whereas the Sharpening-Maximizer will ϕ every time.

But notice that it doesn't rule out a lot of freedom: in cases of flat indeterminacy, there may still be a lot of options permissible according to the Sharpening-Maximizer. In the Mushy Boats case, where the indeterminacy is 'flat', the Randomizer will in the long run save each boat half the time. But we can make no such prediction about the Sharpening-Maximizer's behaviour, even assuming full rationality: we know that every time, she will choose to save one of the boats, but can say no more about the long-run frequencies of her choices than we can about the Ass of Buridan. Thus, surprisingly, though Sharpening-Maximize is a more restrictive rule than Randomize in many one-shot cases, in terms of which actions are permissible, it allows for more 'long-run' freedom.

Each of the views we have seen is non-aggregative, in the sense that all that matters to the permissibility of ϕing is the existence or measure of sharpenings on which ϕing is mandated. Though not strictly entailed by these considerations, we have been implicitly assuming that aggregative considerations do not matter. But the following case should bring this into question:

> **Imbalanced Utilities.** The situation is flatly indeterminate, and there are two sharpenings. On one sharpening, ϕing realises 1 more util than ψing; on the other, ϕing realises 5000 utils less than ψing.

In this case:

- if we ϕ, then it is indeterminate whether we maximise utility or attain just one util short of the maximial utility level;
- if we ψ, then it is indeterminate whether we maximise utility or fall 500 utils short of the maximal utility level.

To make the case vivid, we can use JRG Williams's device of the wholly self-interested individual with indeterminate survival-conditions. Suppose that it is indeterminate (true to degree 0.5) that the person occupying your (present) physical body in one month is you. I offer you the following options: you may ϕ (take five dollars now), or ψ (take five million in a month). What ought you do? It seems plain to me that you ought to take the five million dollars: doing so is indeterminately either vastly better, or slightly worse, than taking the five dollars today.

When the question is posed directly—'assuming flat indeterminacy, should you take an option that is indeterminately vastly better or slightly worse, or an option that is indeterminately either vastly worse or slightly better?'—the answer seems obvious. Once again, epistemicism and its associated expected utility maximisation can render the correct verdict: if you simply attach credence of 0.5 to the proposition that it will be you who receives the five million dollars, and credence close to unity that it will be you who receives the five dollars, then rational choice is clear on how you should pursue your self-interest.

Alternatively, suppose that you are ill. If treatment A is either slightly better or much worse than treatment B, but each has a fifty percent chance of being the better treatment, then we should apply treatment B.[24] Intuitively, at least, it doesn't matter whether 'either' here is given an epistemic or an indeterminist gloss.

Caprice and Sharpening-Maximize will say that either ϕing or ψing is permitted—it is permitted to take either sum of money—because each is mandated on at least one sharpening, and they are supported on an equal measure of sharpenings. For similar reasons, according to Randomize, we must choose randomly between ϕing and ψing. But these are, I think, plainly the wrong permissibility verdicts. Simply counting sharpenings is a step in the right direction, but it is not enough.

4.6 Compromise and Hyperexpectation

Now, we *could* patch the view by considering not only the measure of sharpenings which mandate ϕing, but also how strongly those sharpenings prefer it:

> **Sharpenings-Value-Maximize.** Take that action ϕ which maximises the number of sharpenings that mandate ϕ multiplied by average utility of ϕing on those sharpenings.

Note that one could either maximise or randomise on the resulting metric: either only ϕing is permissible, or one should choose randomly but weighted towards ϕ.

I think this view is the furthest one can go in the no-compromise direction; that is, it's the most plausible view that respects the second component of Caprice—Capricious Requirement:

> **Capricious Requirement.** If it is permissible for A to ϕ, then it is not determinately false that A has most reason to ϕ.

[24]See Jackson (1991). I'm indebted to Miriam Schoenfield pointing this out to me, and for very helpful discussion of this issue.

Alternatively, Capricious Requirement holds that any permissible action will be mandated on at least one sharpening. But on the view I'll now defend, it is it is false: sometimes the right action is an action such that it is *determinately not the best*.

Here is an example of the sort of case I have in mind:

> **Mushy Boats Redux.** You are in (a variant of) Mushy Boats once more. This time, you know that one boat contains one hundred people and the other is empty, but it is indeterminate which is which. On half of the sharpenings, the westerly boat carries 100 passengers; on half, the easterly boat carries 100 passengers. But this time, you have two rescue craft, each capable of carrying 90 people. You have three options:
>
> ϕ: rescue the entire complement of the westerly boat;
>
> ψ: rescue the entire complement of the easterly boat;
>
> χ: send one craft to each boat, rescuing up to ninety people from each boat.

If you ψ or χ, it is indeterminate whether you save a hundred people or zero. But if you χ, then it is determinate that you save ninety people (and that ten die). I think that χing is at least permissible, and that it is plausible that you ought to χ: it seems to me to be a clearly better option than either ϕing or ψing.

But notice that χing is determinately dominated: no sharpening mandates such a division of resources. Each of the views we have seen embraces Capricious Requirement: on those decision rules, it is permissible only to send both rescue craft to one of the boats, and it is impermissible to χ. But intuitively, we ought to compromise, and Capricious Requirement is false.

These considerations show that though Gustafsson's Avoid Indeterminate Worseness is plausible, a strengthening of it is not. The original principle—if possible, one should choose such that determinately, there is no better option—is correct. But it does not follow that we should always *avoid* options such that it is determinate that there is a better option. If determinate first place is unavailable, sometimes determinate second place beats being indeterminate between first and third. It's true that we shouldn't act such that *there is an option that is determinately better*, but it is not true that we shouldn't act such that *there is determinately a better option*.

If it seems crazy that there might be permissible actions which are determinately not what we have most reason to do—in other words, permissible actions which are mandated on *no* sharpening—then perhaps that worry can be alleviated by considering cases of ignorance (but no indeterminacy), where the analogous claim is plausible.

Suppose that treatments A and B each have 50% chance of killing (0 utils) and of wholly curing (100 utils). Treatment C is certain to cure the patient but leave her with a mild headache (95 utils). It seems clear that one ought to take treatment C.

In the presence of ignorance, it is often better to take an option, even when we know that there's a better option available, if we don't know which option is better. One principle—if possible, one should choose such that knowably, there is no better option—is correct. But it does not follow that we should always *avoid* options such that it is knowable that there is a better option. If knowably first place is unavailable, sometimes knowably second place beats being ignorant between first and third. It's true that we shouldn't act such that *there is an option that is knowably better*, but it is not true that we shouldn't act such that *there is knowably a better option*.

So much is plausible in the case of ignorance, but might encounter some resistance under indeterminacy. If you are still hesitant, consider another case, in which there are not two but a hundred sharpenings. We again appeal to J.R.G. Williams's device of the wholly self-interested man:

> **Selfish Fission.** Our hero must go through a fission machine, from which 100 numbered men $m_1, \ldots, m_{100}$ will emerge. It is indeterminate which man he will be: there are one hundred sharpenings $s_1, \ldots, s_{100}$; on each sharpening s_k, he is the man m_k.
>
> The fission clinic offers two pricing plans:
>
> 1. On Plan Basic, the procedure costs £100, and pre-fission, the patient may designate a number k such that m_k gets £150 and a handshake. The other

ninety-nine men are used for extraordinarily painful medical experiments.

2. Plan Excel costs £110, and every man who emerges from the device given £150 and sent on his way.

Capricious Requirement demands that the patient select Plan Basic; each sharpening simply mandates that he chooses a different number. Different views flesh this out in various ways: Randomize, for example, requires that he pick a number at random. But plainly, even on self-interested grounds, he ought choose Plan Excel. And yet if he chooses Plan Excel, there is determinately some instance of Plan Basic that would realise more utility for him (though it is indeterminate which instance that is).

Intuitively, the difference is that each sharpening mandates Plan Basic together with a given number; but what makes Plan Excel required, even though it is mandated on no sharpening, is that each instance of Plan Basic is so horrible on the sharpenings on which it is *not* mandated. We have seen not only the relevance of those sharpenings which mandate an action, but also those sharpenings which condemn it: how many of them are they, and how strongly do they condemn it? These aggregative considerations cannot be ignored.

In what follows, I'll defend a more general decision rule, which implies Gustaffson's constraint as a special case:

> **Hyperexpectation Maximize.** It is permissible for A to ϕ *iff* ϕing maximises hyperexpected utility for A.

'Hyperexpected utility' is just expected utility calculated across sharpenings, in the same manner that expected utility is calculated across possibilities. Suppose that there are n sharpenings, $s_1, \ldots, s_n \in S$. Then for a given action ϕ its hyperexpected utility is:

$$HE(\phi) = \frac{1}{n}(u_{s_1}(\phi) + \ldots + u_{s_n}(\phi)).$$

Let's take the fission case as an example. Suppose that utility is linear in pounds, and that taking part in a medical experiment is worth £1m. Consider the action of taking Plan

Excel: it is determinate that the agent loses 110 utils from the fee, and that on every sharpening, he gains 150 utils. Then his hyperexpected utility is:

$$HE(Excel) = \frac{(-110+150)+...+(-110)(+150)}{100}.$$

Since there are 100 terms in the sum, this equals 40 utils.

On the other hand, the hyperexpected utility for an instance of Plan Basic is:

$$HE(Basic) = \frac{50-99(1{,}000{,}000+100)}{100} = -990098.5\ .$$

So Hyperexpectation-Maximise captures the awfulness of Plan Basic as compared to Plan Excel, though there is determinately a better option than taking Plan Excel.

This is, I think, the most natural way to implement orthodox expectation-maximising in the presence of indeterminacy. If epistemicism is correct, and vagueness is ignorance, then 'hyperexpected value' is just a slightly silly name for expected value calculated over the particular sort of ignorance that accompanies vagueness.

Again parallel to this epistemic case, there are two senses of 'ought' and 'permissible', corresponding to 'objective' and 'subjective' senses in the presence of ordinary ignorance:

> **Sharp-Ought**. The (original) ought-reasons platitude is true of sharp-oughts. You sharp-ought to do what you have most reason to do.

Ought-Indeterminism is the correct account of sharp-oughts. When it is indeterminate what an agent has most reason to do, it is indeterminate what she sharp-ought do. But another platitude is true of the second 'mushy' ought:

> **Mushy-Ought**. It is permissible for A to ϕ *iff* ϕing maximises hyperexpectation for A.

The view's distinctive claim is that this second notion of mushy-ought or mushy-permissibility is required to explain our judgements of permissibility in the face of indeterminacy. Moreover, this second notion is *primary*—it is closely connected to criticism, praise, and blame. We criticise someone based on whether she has done as she mushy-ought.

In many cases, we will not be able to calculate hyperexpectations, but can appeal to a slogan form: one mushy-ought make it determinately true that one does as one sharp-ought. (Compare with the epistemic case, where 'determinate' and its cognates would be replaced by 'knowably' and its cognates.) But we *cannot* say that in general, if the only options are to make it either indeterminate whether one does as one sharp-ought, or determinate that one does not do as one sharp ought, then one mushy-ought do the former. As we have seen, there are some cases where compromise is the only mushy-permissible option. Compromise is required given a sufficiently detailed setup—including, importantly, measures of sharpenings—and is thus likely to be rare, but it is not something that we can ignore.

Unlike Ought-Indeterminism, Caprice, and Randomize, the present view renders the correct verdicts in the cases under consideration. First, the symmetrical cases. In both cases, it is indeterminate which option is sharp-permissible for you. Therefore, rather than an dilemma of indeterminate rationality, the cases are more akin to an indeterministic Buridan's Ass: you may permissibly choose either. Both options are indeterminately sharp-permissible, and it is mushy-permissible to choose either. (We do not need to calculate hyperexpectations, in general, if the situation is flatly indeterminate.)

Now, consider the Canberra Suburbs. It is indeterminately sharp-permissible to choose Wye, but it is determinately true that choosing Exe is sharp-permissible. Thus, you mushy-ought to choose Exe, and it is mushy-impermissible to choose Wye. Since the mushy notions are primary, we may simply say that you ought to choose Exe, and it is impermissible for you to choose Wye.

In general, the hyperexpectation view renders the same verdict as epistemicism about vagueness conjoined with ordinary expectation-maximising accounts of rationality.

Importantly, the present view can explain the following datum about how we think about rational choice under indeterminacy: the questions 'what actions are permissible when it's indeterminate which actions are permissible?' and 'what ought I do when it's indeterminate what I ought do?' are coherent. Both 'permissible' and 'ought' are ambiguous. This bifurcation of these notions may be what Sarah Moss is hinting at in an interesting recent discussion of *sequential* action under indeterminacy:

> If you have already rejected the investment, some theorists may say that rejecting the loan is permissible, while the joint act of rejecting the investment and the loan is not (cf. Caprice in Weatherson 2008 and Sequence in Elga 2010). If that is right, then intuitively agents should be interested in which actions are such that performing them will not make it the case that you have performed an impermissible sequence of actions, and readers may interpret 'permissible' in the text as denoting this property.[25]

Though Moss does not draw out the connection, I think that a similar distinction can explain why a given action considered in isolation may be mushy-permissible, but considered as the end of a sequence of actions, is mushy-impermissible.

In summary, the present view has the following central claims: (1) it can be indeterminate what one has most reason to do; (2) the ought-reasons platitude is true of one (sharp) 'ought', and in such cases, it is indeterminate what one sharp-ought to do; (3) there is another (mushy) 'ought' which is primary; (4) sharp- and mushy-ought are related in much the same way as objective and subjective oughts in the face of ignorance.

4.7 Bleaching Objections

The Hyperexpectation view is an attempt to capture epistemicism's plausible precriptions for action, but within an indeterminist framework. This motivates a class of objections to the view: doesn't it unacceptably 'bleach out' indeterminacy, and treat situations as determinate when they are not?

There are several ways of putting this line of objection, and I consider them in turn.

Does the view ignore angst and regret? The view might seem to allow agents to rest a little too easy with their choices. Buridan's Ass can rest easy, because whichever option she chooses, it is true that she could have done no better. But you, when facing a symmetrical indeterminacy case, ought perhaps not be so complacent. It is, after all,

[25] Moss (forthcoming[b]), p. 11, fn. 8.

indeterminate whether you acted as you had most reason to do. Seung and Bonevac make a similar point:

> the anxiety of choice [. . .] has been a favorite theme of existentialism. No doubt epistemic indeterminacy [i.e., ignorance] is the most familiar source of anxiety. We have to make our important decisions in fear and trembling, if we have to do it without really knowing what they will eventually amount to. But the anxiety of making decisions can go deeper than the level of epistemic uncertainty; it can also be generated by the model of indeterminate ranking [. . .] you may feel a great deal of discomfort with your decision. For there are weighty reasons against your decision, and they cannot simply be ignored.[26]

My response to this involves a certain degree of bullet-biting. The view does often—as here—render a determinate permissibility-verdict even when the balance of reasons is indeterminate. But we should not think that in general just because taking a certain action (rationally) causes angst and regret, that the action was not determinately permissible.

If degrees of truth are admitted, then the objection can be made more pressing. Suppose it is true to degree 0.6 that you sharp-ought to do P, and true to degree 0.4 that you sharp-ought to do Q. Then (on plausible background conditions) according to my view, there is no question about what you mushy-ought to do: you mushy-ought to P. But how could you avoid anxiety about this action? It is only slightly less than halfway true that this was the wrong action.

Once again, the bullet can be bitten: all else being equal, it is plausibly morally required to save the boat with 51 people over the boat with 50 people; this doesn't mean that the agent who makes this choice should rest easy. We shouldn't put too much weight on examples that rely on degrees of truth: if the situation is really so specific as this, then the 'counterintuitive' verdicts of the view (that such a small change makes a difference in how you should act) become much more plausible, as we'll see in a moment.

[26] Seung and Bonevac (1992), p. 808.

Does the view render overly specific judgements? Maybe the problem is that sometimes when it is indeterminate which of X and Y one has most reason to do, the view says only Y is permissible. Then, in that case, there is no decision-theoretic distinction between an action which determinately maximises utility, and one which maximises utility on 6 out of ten sharpenings. This might seem to gloss over (or 'bleach out') the distinction between cases in which it really is indeterminate (albeit in an asymmetric way) what one has most reason to do, and cases in which it is determinate what one has most reason to do.

I do feel the force of this objection: there is no denying that this risks implausibility. But in its defence we can once again say two things: first, such precision is unlikely to arise; second, we say similar things about ordinary expected utilities. The second point implies that if these verdicts are found implausible, then this is a strike against epistemicism about vagueness, since they seem to be implied by epistemicism.

If these considerations are not convincing, we might also say that such apparently overly precise judgements are the price we must pay to accommodate cases such as Imbalanced Utilities. If we wish to say that in cases where on 99 sharpenings ϕing is overwhelmingly mandated, and on the remaining 1 sharpening ψing is slightly superior, we ought to ϕ, then we are committed to treating at least some cases of indeterminacy as if they were determinate: the verdict here is no different to that which would obtain if it were determinate that ψing maximises utility.

Does the view undermine small-improvement arguments? Such arguments are central arguments for the claim that in putative cases of incommensurability, the options are not simply equally good with respect to the value in question: if they were precisely equally good, then very slightly improving one option would make it determinately better than the other. But this doesn't seem to be the case. But on my view, small-improvement arguments will sometimes fail:[27]

[27] See De Sousa (1974), Sinnott-Armstrong (1985) and Raz (1985-86) for classic presentations of small-improvement arguments. I'm indebted to Miriam Schoenfield for pressing this line of objection against my view, in conversation.

> **Job Negotiations.** It is indeterminate whether a certain legal job (Law) or a certain musical job (Clarinet) is better. That there are ten admissible sharpenings of the relevant predicates; on five sharpenings, Law is preferred, and on the other five, Clarinet is preferred. Another musical job (Clarinet+) differs from Clarinet only in that Clarinet+ pays a little more, but this is enough to say that Clarinet+ is preferred to Law on six sharpenings, and Law to Clarinet+ on four.

My view will then say that Law and Clarinet are incommensurate, and that either may permissibly be chosen. But since it is truer that one sharp-ought to take Clarinet+ than that one sharp-ought to take Law, my view will say that one mushy-ought to take Clarinet+ over Law. Since mushy-ought is primary, we may simply say that one ought take Clarinet+ over the Law.

But then the small improvement (from Clarinet to Clarinet+) did make a difference in what one ought to do: the case was sensitive to sweetening. But wasn't one-sided improvement, or insensitivity to sweetening, supposed to be the mark of incommensurability, a mark that incommensurability as vagueness is particularly apt to accommodate? The decision theory on offer for vagueness in general—and incommensurability as vagueness in particular—risks undermining one of the motivations for incommensurability as vagueness, as well as rendering independently implausible verdicts.

I think that this is a bullet that can be bitten. My view does render the suggested verdicts, and the case is sensitive to sweetening. But this trades on rather special features of the case, which appeals to a degree-theoretic notion of indeterminacy in the form of 'counting' sharpenings, and a finely-balanced count at that. It also stipulates that there is no higher-order indeterminacy, which would manifest as indeterminacy in the sharpening-count.

Once this is seen, the verdict rendered by the generalised ought-reasons platitude is no longer quite so strange. We might say that it is akin to a lottery case with ignorance but no indeterminacy: one may permissibly pick between two tickets for a lottery that each have a 0.5 likelihood of winning; but one is required to pick the ticket with a 0.6 likelihood of winning over that with a 0.4 likelihood.

Any account of incommensurability faces two points at which small improvements indeed make a difference: the small improvement between a musical job that is worse than Law and one that is incommensurate with Law, and that between a musical job that is incommensurate with Law and one that is better than Law. The present view does imply that there is a third such point. Suppose a conception of truth and indeterminacy sufficiently fine-grained to say (i) that it is equally true that Law is better than Clarinet as vice versa; and (ii) that it is truer that Clarinet+ is better than Law than vice versa. Then the verdict of the view—that it is permissible to choose Clarinet over Law, but impermissible to choose Clarinet+ over law—rests on so much being packed into the hypothetical, that it is not an unacceptable cost. Certainly, we would be dealing with an unusual case.

4.8 Incommensurability

I will conclude with a discussion of incommensurability. Rational choice under indeterminacy is relevant to incommensurability as vagueness in two ways. First, the Canberra Suburbs case forms the backbone of John Broome's argument for his 'Collapsing Principle', which restricts the forms of indeterminacy in comparative predicates:

> **Collapsing Principle.** For any x and y, if it is false that y is Fer than x and not false that x is Fer than y, then it is true that x is Fer than y.[28]

Here 'Fer than' is a comparative predicate like 'better than', 'tastier than', or 'more impressive than'. The Collapsing Principle is intended to be a general truth about all such predicates; in effect, it rules out indeterminate superiority cases as incoherent. The Collapsing Principle is the central premise in Broome's argument for incommensurability as vagueness; but it is false—it is vulnerable to various counterexamples.[29] Since it is

[28]Broome (1997), p. 74. This is the less general version of the principle.

[29]See Chapter 3 and Carlson (2004).

false, Broome's argument for it must be unsound, but it is not easy to see where it goes wrong:

> When it is false that y is Fer than x but not false that x is Fer than y, then if you had to award a prize for Fness, it is plain that you should give the prize to x. But it would not be so plain unless x actually was Fer than y. Therefore, x is Fer than y. This must be so whether you actually have to give a prize or not, since whether you have to give a prize cannot affect whether or not x is Fer than y.[30]

Compare this with the final sentence of the Canberra Suburbs: 'Since the prize was for being the greater Australian, it could not be so obvious who should win unless that person was the greater Australian.'

But we can see now that this is an unrestricted application of Reasons-Ought in the presence of indeterminacy: the claim that 'it would not be so plain [that you ought to give the prize for Fness to x] unless x actually was Fer than y' is false. In our terms, this inference assumes Ought-Indeterminism. But as we have seen, that view is false: it can be true that an agent ought to ϕ, without it being true that the agent determinately has most reason to ϕ. We simply require that ϕing maximise hyperexpectation, as it does in this case.

So the current view—in virtue of rendering the correct verdict in cases of indeterminate superiority like the Canberra Suburbs—allows us to identify the false premise in Broome's argument. Besides clearing up that puzzle, the present view can also explain why in (one-shot) choices between incommensurate options, if incommensurability is vagueness, it is generally permissible to choose either option:

> **Incommensurate Cars.** You are choosing between two cars. The sports car is better in terms of style and handling, and the family car is better in terms of room and safety. Suppose that all other cars are worse: these are your

[30] Broome (1997), pp. 74-5.

short-listed two, and you are choosing between them. You have overall reason to choose the better car.

If incommensurability is vagueness, and vagueness is indeterminacy, then in this case it is indeterminate which of the sports car and the family car is better, and thus indeterminate which of them you have overall reason to choose. It is determinate that all other cars are worse than both.

To speak in terms of rankings, it is indeterminate whether the sports car ranks top and the family car second, or the family car top and the sports car second. However, it is determinate that all other cars—which are perhaps neither safe nor stylish—rank third and below. So you definitely have most reason not to choose any of these other cars; applying the platitude, you ought not to choose any of the other cars, and it is impermissible to choose any of them.

But what about the two cars in question? It is indeterminate which of the sports car and the family car you have most reason to choose, and so it is indeterminate (according to the view under consideration) which you ought to choose. Whichever of the two you choose, it is indeterminate whether you choose sharp-permissibly. If neither has any sort of truth or determinacy advantage, then it is mushy-permissible to choose either.

Now, the view does imply that in some cases of incommensurability, it may nevertheless be the case that one mushy-ought to choose one incommensurate item over another. I don't find this especially implausible, since such cases are rare.

4.9 Conclusion

The central message of this paper is this: when you are confronted with a decision under indeterminacy, treat it as if it were ignorance. It is distinctive of vagueness and ignorance that it may be knowably true that some k in K has a given property, and yet for each k, it is, indeterminate or unknowable whether *that* k has the property. The present view exploits this structure.

The strategy of this paper has been to present several puzzling cases of practical choice, which hang on vagueness or indeterminacy in some way, and to argue that Hyperexpectation renders the correct verdicts in those cases. If this account is correct, then it can explain why in cases of incommensurability, it is typically permissible to choose either incommensurate option. It fits this judgement into a general framework, where not all indeterminacy in reasons is so ‘symmetrical’.

There is a delicate issue about whether this is best seen as the ought-reasons platitude being ‘false’. Whichever way the predicates involved are made wholly sharp, the platitude will be true. Thus it is ‘supertrue’. Similarly, claims like ‘there is a precise minimum number of grains required to form a heap’, and ‘there is a precise minimum number of hairs required to count as non-bald’ are supertrue (fixing the relevant contextual elements). But they attribute precision where there is vagueness; whatever status they have, so has the ought-reasons platitude.

In his excellent recent discussion of the ought-reasons platitude, Douglas Portmore gives a strikingly similar maximising account, in response to a different phenomenon, concerning the open future. He discusses the possibility that we cannot guarantee that we will perform some action (for example by forming an intention now), but can merely make it more or less likely that we will do so:

> So if it is indeterminate whether I would do x at t′ even if I were to intend at t to do x at t′, then my doing x at t′ is not an option for me at t. [. . .] although my doing x at t′ is not an option, my trying to do x at t′ is. And it seems to me that how much reason I have to perform this or any other option is just a function of the expected *deontic value* of my j-ing (where j is something whose performance I am able to see to) and how that compares to the expected deontic values of all my other options. [. . .][31]

This is a phenomenon not of indeterminacy in what one has reason to do, but of indeterminacy in one’s future actions. (I’m not suggesting that a bright line can easily be drawn, here.) This is more evidence for the claim that expectation-maximising can go further than one might think. Certainly, as we have seen, we can construct an

[31] Portmore (2013), fn. 34.

expectation-maximising account of action under indeterminacy that captures the verdicts rendered by epistemicism.

5 The Self-Torturer is a Sorites

Abstract. I argue that Warren Quinn's 'Puzzle of the Self-Torturer' is merely an instance of the paradox of the sorites, or heap, and that it thus presents no challenge to orthodox accounts of rational choice. Once we see the essentially comparative nature of the Self-Torturer's preferences—he wishes to maximise money and minimise pain—the soritical nature of the puzzle becomes clear.[1]

5.1 Introduction

Should you eat this piece of cake? Perhaps not, if you are getting married in two months and don't want to gain a significant amount of weight. But this piece of cake will bring you a clear, nonzero amount of pleasure now, and will make no appreciable difference to your weight at the wedding. In terms of utility, even taking account of the upcoming wedding, more is gained from the eating of the cake than is lost from any weight gain. Since you should maximize expected utility, you should eat the piece of cake.

But if you keep thinking along those lines, you will keep eating cake, and you will gain weight—you will fail in your project of maintaining you figure until you get married. The utility calculation seems right every time you eat some cake, yet continuing to eat will foreseeably lead to a dispreferred outcome. We often put value on the completion of *vague projects* such as maintaining one's weight for a wedding, which—like 'is a heap' or 'is bald'—do not have precise satisfaction conditions, and the issue seems to be rooted in this vagueness.

One more calorie won't make an appreciable difference to whether you gain weight.

---

[1]For comments and discussion, I am indebted to John Broome, Ryan Doody, Thomas Hofweber, Douglas MacLean, Katherine Meehan, Julia Nefsky, Ram Neta, C.D.C. Reeve, John Roberts, Geoffrey Sayre-McCord, Keith Simmons, Susan Wolf, audiences at Chapel Hill and the North Carolina Philosophical Society, and an anonymous reviewer.

But many will. You cannot abstain from food entirely—you would starve—and even some cake is perfectly permissible. If a theory of rational choice cannot account for the weighting of such simple goals, and explain why we ought eat, but not too much, then it is in trouble.

A vivid instance of this challenge is Warren Quinn's notorious *Puzzle of the Self-Torturer*. The eponymous victim ('ST') is repeatedly offered a lot of money to accept an imperceptible, but permanent, increase in electric current running through his body. If it is always rational to maximize expected utility, then it seems always rational for ST to accept this deal: doesn't $10,000 outweigh an imperceptible increase in current, in utility terms? And yet, foreseeably, if he takes the deal repeatedly, he will end up in agony: many imperceptible increases in current amount to unbearable pain.[2]

It is natural to think that such issues are what we'll call 'Soritical'—grounded in the paradox of the sorites. In that paradox, it is compelling that removing one grain of sand from a heap couldn't make it a non-heap: that if n grains form a heap, then $(n-1)$ grains form a heap. But if 10,000 grains form a heap, this implies (by repeated applications of modus ponens) that five grains form a heap. This is a sorites on 'is a heap': it is vague where we tip from a heap into a non-heap.

As we'll see, many philosophers have followed Quinn in denying that the Puzzle of the Self-Torturer is Soritical. They instead argue that the puzzle is a counterexample to orthodox utility-maximizing accounts of rational choice. In this paper, I have two main goals. First, to argue that the natural thought is indeed correct: puzzles like the Self-Torturer are instances of the sorites paradox. Second, I show that once the Soritical structure of the puzzle is seen, the immediate threat to orthodox rational choice theory dissolves.

The Self-Torturer is a sorites on 'is the level of current that maximises utility', and other more substantive predicates. Just as it is vague where we tip from a non-heap into a heap, it is vague at what point the less-pain project outweighs the more-money project.

[2] Quinn (1990). Quinn uses male pronouns for ST; Tenenbaum and Raffman use female; I use an inconsistent mixture.

There is a setting on the torture device that optimally trades-off pain and money—a setting from which you should advance no further—but it is vague what that setting is. There is a number of pieces of cake eaten at which your expected utility is maximized—a number at which you should eat no more—but it is vague what that number is. Like sharp (that is, non-vague) projects, vague projects can have various structures, and this diversity has masked their shared Soritical structure.

On every mainstream account of the sorites, claims like 'removing one grain doesn't turn a heap into a non-heap' are false: they have false instances. In the sorites, there are clear heaps, borderline-heaps, and clear non-heaps. There is a minimum number of grains required to form a heap, but it is indeterminate or unknowable what that number is.[3] In cases like the Self-Torturer, there is a range of borderline-utility-maximizing points, at which it is vague whether *that* point maximises expected utility. You ought to maximize expected utility, but it is vague how to do so. The central premise of the challenge to orthodox rational choice theory—that ST is always required by that theory to turn the dial on the torture device—is false.

In the next section, I'll explain the precise challenge that vague projects present to the orthodox view of rational choice.

5.2 The Puzzle for Orthodox Rational Choice Theory

Consider again your wedding. You don't have a specific goal in mind, but you would prefer not to gain more than a few pounds. This is not unusual, since picking a precise target weight is an artificial—though often psychologically useful—exercise. Let's suppose that between now and the wedding, you will be offered many cakes. Your only relevant preferences are for food and for no significant weight-loss. How many ought you to eat?

We are not here worried about *extremal* choices, to eat all or none of the cakes: you might not like cake, or you might be devoted to it. There could be substantive questions about the justification of such preferences, but our puzzle is different. How can accepting

[3] As discussed below, I use 'vague' and 'borderline' neutrally between theories of vagueness.

one or more cakes, but refusing others—a *non-extremal* choice—be justified?

Suppose that you make a non-extremal choice, to eat k cakes. Since the wedding is far away, you could have eaten $(k+1)$, and gained some decent amount of gustatory pleasure at a negligible cost. As we said, just one piece of cake will make barely any difference to your physique. The marginal utility of that piece of cake certainly appears to outweigh the marginal disutility of such a small weight gain. If rationality is maximizing expected utility, then stopping at k cakes seems unjustified. You ought to have another. By parity of reasoning, the same could be said about *any* non-extremal number of cakes. Utility maximization seems to require that if you eat one cake, you eat them all.

But you would far prefer to eat ten pieces of cake and lose weight, than eat one hundred and gain a substantial amount of weight. Utility maximization has led you astray. This seems to show that you do better by deviating from a central tenet of the orthodox view:

> **Tenet.** At every choice point, act to maximize expected utility.

Note that Tenet does not forbid taking necessary means to an extended project. If Jessica has a project of completing her doctoral degree, then this figures into the utility calculation when she chooses whether to sit down for her (required) comprehensive exams. But abstaining from cake seemingly cannot be justified on such grounds. Failing to sit her exams would prevent the completion of Jessica's doctoral degree project, but if one more piece of cake will make no difference to the outcome of your project of maintaining your physique, then Tenet demands that you eat it. But this is the path to weight gain. In order to avoid such an undesirable outcome, Tenet must be disobeyed.

And if you do better by deviating from some account of rational choice, then how can that account be correct? This is analogous to 'Dutch Book' arguments that our subjective credences must be probability functions: if they are not, we will do worse.[4]

The classic instance of this puzzle is Warren Quinn's *Self-Torturer*:

---

[4]Dutch Book arguments show that if an agent's credences do not conform to standard probability axioms, she can be exploited by unscrupulous bookmakers and inevitably lose money.

> Suppose that there is a medical device that enables doctors to apply electric current to the body in increments so tiny that the patient cannot feel them. The device has 1001 settings: 0 (off) and 1 … 1000. Suppose someone (call him the self-torturer) agrees to have the device, in some conveniently portable form, attached to him [...] The device is initially set to 0. At the start of each week he is allowed a period of free experimentation in which he may try out and compare different settings, after which the dial is returned to its previous position. At any other time, he has only two options – to stay put or to advance the dial one setting. But he may advance only one step each week, and he may *never* retreat. *At each advance he gets $10,000.*
>
> Since the self-torturer cannot feel any difference in comfort between adjacent settings, he appears to have a clear and repeatable reason to increase the voltage each week. The trouble is that there *are* noticeable differences in comfort between settings that are sufficiently far apart. Indeed, he can see that if he keeps advancing, he will eventually reach settings that will be so painful that he would then gladly relinquish his fortune and return to 0.[5]

The Self-Torturer is certainly vivid, but it conflates three questions. First: how could *imperceptible* differences in electrical current amount to very severe differences in pain? Second: do vague projects undermine orthodox utility-maximisation? Third: when ought ST stop advancing, and why?

These questions are separate. The question of imperceptibility has important implications for claims such as 'luminosity'—that one can always know whether one is in pain—but a puzzle akin to the self-torturer can arise even in the absence of imperceptibility. if we have the right preferences, the puzzle will also arise if each cake produces a small but perceptible expansion of one's waistline. All that is required is that such effects seem to be *outweighed*. Even the Self-Torturer can be cast in such terms: we

[5] Quinn (1990), p. 79. Emphasis in original.

might modify the case so that each increment adds a small but perceptible amount of pain, one that seems to be clearly outweighed by $10,000.

This allows us to focus on the second question, which can be cast as: in the face of vague projects—whether involving imperceptibility or not—does it remain true that one ought to maximise expected utility? Or are these cases counterexamples to utility-maximisation, as has been argued by, for example, Tenenbaum and Raffman (2012)?

This again is distinct from our third question. Suppose that—as I'll argue—in cases like this, it is true that one ought to maximise expected utility, but that it is vague at what setting the Self-Torturer's utility is maximised. This is not yet to give an answer to how one ought to *act* in the face of such vagueness: in chapter 4 I offered an answer to this question in the synchronic case, but the Self-Torturer is faced with a diachronic puzzle. The question of how to act with indeterminate *credences* has spawned a considerable literature. But the second question is more akin to that of whether a situation really is represented as trading on indeterminate credences.

Despite these complications, the Self-Torturer has structured much subsequent discussion of these issues. So it is in terms of that puzzle that I'll proceed. In this paper, I focus solely on the second question: I'll argue that it is true that the Self-Torturer ought to stop at a setting that maximises expected utility, and that there is a setting that maximises expected utility (though it is vague what that setting is). I shall set aside the first question, concerning imperceptibility.

It might be worried that this is a cheat: isn't imperceptibility central to the puzzle as Quinn presents it? This is not entirely clear: as we'll see below, he claims several times that the puzzle hangs on vagueness. But insofar as we can model the scenario by appeal only to vagueness, this is evidence that imperceptibility is a secondary concern. First, let's see what the Puzzle of the Self Torturer is *not*.

The issue is not one of of procrastination or addiction, or of a slippery slope. The procrastinator knows that she is rationally required to work on her paper now rather than watch more television, yet fails to do so. On slippery slopes, taking one step causally or rationally alters my judgements about the wisdom of further steps. I might now judge that two graduate courses in metaphysics is the optimum number, but also know that taking one will cause me to judge that four is optimal, and so on. Both phenomena have a two-stage structure: (1) the judgement that it is rational to perform a certain action; (2) a deviation from, or change to, that judgement. The self-torturer puzzle kicks in earlier: we are trying to give a justification for the initial judgement, akin to (1), that some non-extremal stopping point is justified.

The puzzle is also not dependent on factors such as the diminishing marginal utility of money (or indeed cake). If that is taken to be relevant, the amount of money on offer at each point may be increased to compensate for the diminishing marginal utility of each dollar.

That is the Puzzle of the Self-Torturer. ST seems always required by Tenet to advance one more stage. But she can see that if she continues to do so, she will eventually disprefer the outcome. Many philosophers have taken this to mandate a *heterodox* view of rational choice (one which denies Tenet).

Quinn himself argued that ST exhibits intransitive preferences: she prefers each setting to the previous one, yet prefers earlier settings to later. Such preferences are not compatible with utility-maximization, since utilities (or even an ordinal preference ranking) cannot consistently be assigned. Suppose that ST prefers some option B to A, and C to B, but prefers A to C. 'Has higher utility than' is transitive, so there is no coherent answer to the question of whether A has higher utility than C for ST. In the presence of intransitive preferences, Tenet must sometimes be rejected: it seems to lead ST to the dispreferred end of the sequence.

More recently, Sergio Tenenbaum and Diana Raffman have defended another heterodox view, according to which Tenet does not apply to vague projects:

> We propose that a vague project issues in a requirement and a set of permissions. The requirement is just an instance of the instrumental requirement: insofar as one is rational one must adopt (what one believes to be) the means (including constitutive means) necessary to execute one's

> project. The permissions are permissions to execute the project in some momentary actions rather than simply maximizing utility in light of one's preferences for momentary actions considered in isolation.[6]

They argue that a refusal by ST to turn the dial cannot be justified on the grounds of expected utility, and instead 'the pain-free life project issues permission to stop turning the dial, independently of what maximises utility in light of ST's momentary preferences', so even though utility may be maximized by advancing one stage, ST has permission to refuse.[7] This is the sense in which their view is heterodox: sometimes, even though it would maximize utility for ST to turn the dial now, she is rationally permitted (perhaps required) to refrain. Expected utilities do not serve as a guide to life:

> By reflecting on the nature of vague projects, we learn that in such cases we cannot simply plug weights in to various ends to generate a preference-ordering; rationality is not always purely calculative.[8]

Both of these views are ingenious and subtle, but I'll argue that they are not supported by the Self-Torturer. Such heterodox views rely on the claim that eating one more cake or turning the dial is preferred on utility grounds. In the rest of this paper, I'll argue that the puzzle is a Sorites, so this heterodox challenge dissolves. It is not true that eating one more cake or turning the dial is always preferred on such grounds, any more than it is true that removing one grain of sand from a heap always leaves a heap.

5.3 Arguments that the Self-Torturer is not a Sorites

Quinn's only explicit mention of the Sorites paradox is as a possible objection to his own view:

> Here we find a kind of *sorites* puzzle. Some of [ST's] clear and immediate judgments about his comparative comfort are true only if others are false.

[6] Tenenbaum and Raffman (2012); p. 102.

[7] Tenenbaum and Raffman (2012), p. 106.

[8] Tenenbaum and Raffman (2012), p. 111.

> Surely preferences based on such paradoxical discriminations are not to be taken seriously. The self-torturer must either change his preferences or give up the mix of vague and precise terms that generates the puzzle.[9]

Quinn rightly rejects this possibility. But such eliminativism is not the only or most plausible route; we do not think that because redness or heapiness can generate sorites, we should give up those concepts.[10] Tenenbaum and Raffman also argue that the puzzle is not Soritical:

> Readers familiar with the sorites paradox may wonder whether the self-torturer puzzle is just an especially picturesque instance of it: perhaps ST is proceeding along a sorites series of pains from a clearly bearable one to a clearly unbearable one, attempting to decide where the bearable ones end and the unbearable begin. However, this way of thinking about ST overlooks a crucial element of her situation: at each step of the way she is also trying to decide whether a certain incremental difference in pain can be compensated by $10,000 at that point in the spectrum of her pain. The latter task is what appears to put pressure on her rationality and is, at bottom, the source of the puzzle.[11]

They do not explicitly deny that the puzzle is an instance of the sorites, but they do deny that the sorites is 'the source of the puzzle'—that the puzzle is Soritical, in our terms. Their argument has some force. Consider again the classic paradox of the heap: 1,000,000 grains form a heap; if n grains form a heap, then $(n-1)$ grains form a heap; so, 5 grains form a heap. The Self-Torturer does not seem to be like this. What predicate corresponds to 'is a heap'? Tenenbaum and Raffman imply that it would be something like 'is a bearable level of pain'. A sorites on *this* predicate would, as they rightly say,

[9] Quinn (1990), p. 82.

[10] Below, with some machinery in hand, we'll consider a more threatening argument due to Quinn.

[11] Tenenbaum and Raffman (2012), p. 88 fn. 3.

misdescribe the puzzle. ST is making a comparative judgement—is this additional pain worth $10,000?—which doesn't seem reducible to a fruitless search for the point where a vague predicate runs out.

But it is a mistake to think that a sorites account of the Self-Torturer must look like this (and it is not quite explicit that Tenenbaum and Raffman do think this, but if they do not, then the quoted argument is manifestly ineffective). ST is clearly not trying to find where bearable pains end, and unbearable pains begin. The best place to stop is *long before* the pain is unbearable; if the puzzle is a sorites, it is not one on a predicate like 'is a bearable level of pain', but on something more *comparative*: is this pain increment worth this money increment? In focusing on non-comparative predicates, they are not considering the Soritical view its strongest form.

Before presenting a more general response to their argument, I'll show that there are other vague projects that do not have this comparative structure, and which can be seen as soritical. Later, I'll argue that the Self-Torturer too is Soritical.

5.4 An Example of a Vague Project that Clearly is a Sorites

Utility maximization also seems to lead to disaster in an example due to Richard Tuck:

> **The cairn-builder.** He could be a shepherd who wishes to build a cairn of stones by himself to guide him in the hills. On setting out in the morning, he can reason as follows. If I work all day, I will have a suitable pile of stones by nightfall. But one stone added to a collection of other stones makes a negligible difference – it can never be enough to tip it over the edge and into a heap. It takes a certain amount of time and effort to find a spare stone. If I do not start immediately, I will still have a heap of stones at nightfall, since the stone I could have picked up in the next few minutes would have made no difference to the outcome. But the same applies to the next stone, and the next: there is no point in ever beginning. Moreover, at some time in the day it will be clear that I have passed the stage where I will have enough time to build a cairn, and after that point there is certainly no benefit to be gained by piling

up stones.[12]

I'll show how this example can be seen as Soritical. To begin, let's consider an artificially precise version of Tuck's scenario:

> **The precise cairn-builder.** Another shepherd, John, will get 200 utils from building a cairn by nightfall. A cairn consists of 15 stones; once he starts building, he cannot stop. Each stone takes 30 minutes to move, and costs 4 utils (he will miss one episode of his favourite television show). He has ten hours before nightfall; the cairn will be useless if not completed by then.

We have made it artificially precise how many stones are required to build a cairn, how long it takes to carry each stone, the value of leisure, and the value of a cairn; moreover, we stipulate that there is no relevant ignorance. Intuitively, it is clear what John should do. He ought to enjoy 2.5 hours of television, and then build a cairn consisting of precisely 15 stones before nightfall. He will then be up 140 utils: 200 utils for the cairn, minus 60 utils from moving 15 stones (and missing 15 episodes).[13]

If John starts building the cairn either earlier or later, then he is leaving money on the table, as they say on Wall Street. If he starts early and builds a cairn with more than 15 stones, then he loses 4 utils for each stone above 15, but gains no utility from his bigger cairn. But starting late and building a partial cairn is by far the worst strategy. With only 7 hours left, he can no longer get the 200 utils from building a cairn. If he starts building then, he will be *down* 56 utils: zero utils for the cairn, minus 56 utils from moving 14 stones. The best he can do is to not move any stones, and come out even: no utils for the cairn, minus 0 utils from moving no stones.

Tenet can explain each of these judgements. Every thirty minutes, John must choose whether to start moving stones, or to wait. Suppose that he has already waited k stages of

---

[12]See Tuck (1979), p. 154. A cairn is a pile of rocks or stones, often used to mark mountain trails.

[13]We set aside questions such as whether it is really rational for him to leave it to the last minute like this, and not set aside some time for unexpected events, which form a different set of issues.

thirty minutes. Is it better to wait another thirty minutes for stage $(k+1)$, or to start building? Applying Tenet, he ought to wait for stage $(k+1)$ just in case expected utility is maximized by doing so.

In terms of expected utility, stage $(k+1)$ differs from k in just two relevant ways. There is more leisure time (that is, television), and one less stone on the cairn. So:

> Expected utility is maximized by advancing to $(k+1)$ from k
>
> *iff*
>
> expected utility from cairn-building at $(k+1)$ *minus* that from
>
> leisure at $(k+1)$
>
> *exceeds*
>
> expected utility from cairn-building at k *minus* that from leisure at k.

$(k+1)$ always involves more leisure time, so it beats k on that score. Since the cairn-utility depends only on whether a cairn is built, $(k+1)$ is worse in cairn-terms only if the cairn can be built at k, but not at $(k+1)$. So expected utility is maximized by waiting for $(k+1)$ rather than starting to build at k *unless* the cairn is buildable at k but not at $(k+1)$.[14]

The clause 'the cairn is buildable at k but not at $(k+1)$' is false at all points except one. It is true just in case there is enough time left to carry 15 stones and no more. When 'is a cairn' is purged of vagueness, orthodox rational choice theory correctly requires that John should start work at the last point when he can do so, and finish the cairn by nightfall.

But what about the original, vague case? Tuck is right that this case is clearly parasitic on the vagueness of 'is a cairn'. In the vague version, it is indeterminate how many stones form a cairn. The predicate is *sorites-embeddable*:

> 20 stones form a cairn;
>
> if 20 stones form a cairn, then 19 stones form a cairn;

[14] I'm assuming here that the following situation does not obtain: stage 2 is worse than stage 1, but stage 3 is better than both stage 2 and stage 1. Such 'darker before the dawn' cases are a little more complicated, but not fundamentally different.

if 19 stones form a cairn, then 18 stones form a cairn;

. . .

if 6 stones form a cairn, then 5 stones form a cairn;

But, 5 stones do not form a cairn.

Sorites paradoxes hang on *tolerance* principles such as ‘one grain couldn’t be the difference between a heap and a non-heap’. We can see that this principle, combined with the claims that 1,000 grains form a heap and 4 grains do not, leads to inconsistency. This is a sorites on ‘is a heap’.

Accounts of vagueness fall into two broad camps. ‘Indeterminist’ views have it that there is a false instance of the relevant tolerance principle, but that it is indeterminate which instance is false. ‘Epistemic’ views claim that there is a determinately false instance of the tolerance principle, but we don’t (perhaps can’t) know which it is.[15]

A tolerance principle like ‘if n stones form a cairn, then (n-1) stones form a cairn’—or, in Tuck’s words, ‘one stone can never be enough to tip it over the edge and into a heap’—seems to be true, but it is not. Repeated applications of the tolerance principle take us from truth to falsity.

Since ‘it is better to proceed to the next stage’ is false if and only if ‘this is the last point at which we can build a cairn by nightfall’ is true, there is a sorites on ‘better’:

It is better for John to build with 19 stones than 20;

If 19 stones are better than 20, then 18 are better than 19;

If 18 stones are better than 19, then 17 are better than 18;

. . .

If 6 stones are better than 7, then 5 are better than 6.

---

[15]For indeterminism, see Fine (1975), Keefe (2000), Dorr (2003), and Barnes (2010). For epistemicism, see Sorensen (1988) and Williamson (1994). The ‘tolerance’ terminology is due to Wright (1975). Tolerance principles are false on nearly all accounts of vagueness, but this could be denied. I argue that seeing the Puzzle as a sorites undermines the challenge it presents to orthodox rational choice theory; this could be seen as conditional on the success of such a ‘tolerance-denying’ account of vagueness. I’m grateful to Julia Nefsky for discussion of this point.

By the transitivity of 'better' (on an utility-maximizing characterisation, at least) we Soritically reach the false conclusion that it is better for John to build with 5 stones than with 20. This is how the cairn-builder is caught in a sorites series.

A Soritical analysis of the puzzle of the cairn-builder classifies instances of the tolerance principle into three groups. First, there are its clearly true instances, corresponding to clear cairns. Eventually, we reach the third group, where the principle is again clearly true, corresponding to clear non-cairns. Here, if the principle is put as 'if n stones form a cairn, then $(n-1)$ stones form a cairn', it has a false antecedent.

In between lies the second group, where instances of the principle are *borderline*, corresponding to borderline-cairns. We use 'borderline' neutrally between indeterminism and epistemicism. An ecumenical characterisation—that not every instance of the tolerance principle in the borderline zone is true—is enough to proceed. It is not true that the shepherd always maximises utility by waiting another thirty minutes. The claim that he does depends on the Soritical tolerance principle, which we know to be false. We saw earlier that orthodox rational choice theory has no problem with necessary means to desired ends; in this case, it is vague what the necessary means are to build a cairn.

Tuck argues that the Soritical nature of these cases means that they should be left to the logicians. But the argument just given is also an argument against such quietism: even on a 'logically neutral' ecumenical characterisation of the sorites, we can draw some implications for rational choice, since tolerance principles are, in general, false.

Tenenbaum and Raffman appeal to a variety of vague projects to motivate a putatively general problem for the orthodox view of rational choice. Yet many of the projects they discuss—such as writing a book—are akin to the cairn-builder. If the cairn-builder is Soritical, so are many other Vague Projects. But to stage a comprehensive defence of orthodoxy, we need to show that the Self-Torturer is also Soritical.

To do this, we'll need to introduce a distinction between two different kinds of desires, ends, and projects.

5.5 Binary and Essentially Comparative Desires

I wish to introduce a distinction between what I'll call 'binary' and 'essentially comparative' desires, ends, or projects. Here is the first:

> **Binary desire.** A *binary* desire that a is F is unsatisfied if a is not F, and satisfied if a is F.

Binary desires divide possible worlds or states of affairs into those where the desire is satisfied, and those where it isn't. Many of our desires are binary. I want to go to Churchill, MB, to see the polar bears. That desire is satisfied in worlds where I go, and unsatisfied in worlds where I do not. The desire to have a cairn is similarly binary. But some desires are not like this:

> **Essentially comparative desire.** An essentially comparative desire is for things that are more F rather than less F. There is no 'Fness threshold' beyond which adding Fness is not desired.

Essentially comparative desires do not divide worlds or situations into two classes; they *rank* them. Instead of desiring to see a polar bear, I might prefer to see larger mammals over smaller ones. I will prefer seeing an elephant to seeing a polar bear, and seeing a polar bear to seeing a Scottish Wildcat. It doesn't make sense to ask outright whether my essentially comparative desire to see a larger animal has been satisfied—satisfied compared to what?

The cairn-builder's desire to watch more television rather than less is of this form. So are ST's desires for more money rather than less, and less pain rather than more, and your desire for more cakes, rather than less.

Notice that we could see binary desires as a degenerate case of the essentially comparative: rank all situations in which a is F above those in which a is not F, and make no other distinctions. But what is crucial is that in the essentially comparative case, given a (local) situation x, there is a more preferred option that is Fer, and a less preferred option that is less F. The 'local' constraint here is intended to allow for the (plausibly actual) possibility that for suitably high or suitably low amounts of money, for example, there is no preferential difference: one might be genuinely indifferent between one trillion and two trillion dollars. In actual cases, the distinction between the binary and the essentially comparative marks a significant difference.

There is a structural difference between binary and essentially comparative cases. Once the cairn-builder has gone too far, so to speak, he ought to press on with television, and at least salvage something of the day (assuming, as we are, that he gets no utility from a partial cairn). The self-torturer is different: we would not say that once ST 'has gone too far' and now regrets turning the dial so many times, he ought to press on until the end. That would just make his situation worse. Differences between the binary and essentially comparative cases are reflected both in how we judge that someone ought act, and in the structure of the relevant sorites.

With this distinction in hand, we can see that the projects of ST are not binary, but comparative. Such comparative projects can also be seen as Soritical.

5.6 The Self-Torturer as a Sorites

To keep things simple, let's assume that ST's only relevant preferences are for more money and less pain, and that he has none of either before the experiment begins. His net utility at setting zero is zero.

As before, let's imagine the scenario made fully precise: the *Precise Self-Torturer*. In this case, too, his net utility is just the sum of the utility of the money and the disutility of the pain. We may thus derive some structural constraints on the precise puzzle.

There are four zones. At early stages (zone 1), the marginal utility of taking the deal is positive: the utility from the extra money outweighs the disutility of the extra pain. We know this since ST's accumulated net utility rises through the initial stages, for normal preferences: someone who takes the deal five times, and then stops, is well ahead of the game. Here, the gradient of the money line must be greater than the gradient of the pain line, on an imaginary graph with torture-stages along the x-axis and (dis-) utility along the y-axis.

Eventually, the advantage of the accumulated money 'tops out' (zone 2). There will be a point where the difference between the accumulated money-utility and pain-utility is maximized, and the relative gradients change. This is the highest point of the utility graph—the top of the mountain—and advancing further decreases net utility.

This is to enter zone 3. At first (zone 3a), the accumulated net utility is still positive: ST is still ahead of where he started. But we know that at some point the utility is zero: eventually, ST would willingly return to setting zero both in pain and in money terms, and setting zero has zero net utility. We may then reach settings where the net utility is negative (zone 3b). These are settings where ST is worse off than when he began. It is the prospect of entering zone 3b that makes the puzzle so troublesome.

Zone	Status
1	Marginal utility of taking the deal is positive.
2	Net utility is maximised.
3a	Marginal utility is negative; net utility is positive.
3b	Marginal utility is negative; net utility is negative.

Table 5.1: Precise self-torturer zones

Clearly, the place to stop is at the top of the mountain, at zone 2. Suppose that this is at setting ten. It might be objected that $100,000 couldn't justify *any* pain for the rest of one's life, because lifelong pain is too serious. I disagree, but it doesn't matter: simply imagine that the device will be attached for only a month. The puzzle recurs, but it avoids such objections.

Let the net utility top out at stage m. Stage $(m+1)$ has a lower net utility than stage m, so the marginal disutility of the extra pain at $(m+1)$ over m must outweigh the marginal utility of the extra money at $(m+1)$ over m. In this precise version, ST should, according to both intuition and Tenet, stop at stage m, where net utility is maximized.

Isn't there a worry that we are grossly misrepresenting the puzzle here? No: Quinn himself discusses 'filtered series', which are akin to the precise Self-Torturer. In filtered series, only every tenth (or hundredth, or …) stage is considered.[16] The thought is that by zooming out, so to speak, we can ignore the pairwise intransitive preferences, and focus

[16]Quinn (1990), p. 86.

only on the transitive ones at longer ranges. Filtered series share an 'initial rise followed by decline (or slump)' structure with the Soritical analysis. Seeing the puzzle as a sorites can explain why filtering works: it is akin to considering claims like 'adding 100 hairs couldn't make a bald man not-bald', which are much more plausibly false than 'adding one hair couldn't make a bald man not bald'.

Significantly, neither the precise filtered series pose any threat to orthodox rational choice theory. The threat hung on the claim that there is clear *and repeatable* reason to advance each week, but in these cases, the reason is not repeatable. Expected utility is maximized by pressing on until the point at which the total utility accumulated is maximized, and no further. This is the point at which advancing further will reduce the total utility, since the extra pain will outweigh the extra money. If everything is made artificially precise, then zone 2 is just a point—the point at which the pain gradient first exceeds the money gradient. (It is unlikely but possible that zone 2 is a group of points carrying precisely equal net utility.)

But real preferences are vague. In particular, it is vague where the pain gradient begins to exceed the money gradient. Zone 2 is 'smeared' into a region, just as the zone of where a cairn becomes a non-cairn is smeared.

In this case, we get a sorites on 'maximises utility': for low settings (zone 1), it is false that utility is maximized; in a borderline-zone (zone 2), it is vague whether utility is maximized; for high settings (zone 3), it is again false. The tolerance principle is:

> **Torturer-Tolerance.** If setting k does not maximize utility, then $(k+1)$ does not maximize utility.

As a tolerance principle, this claim is false somewhere in the borderline-zone, indeterminately or unknowably where. The central challenge of the puzzle of the Self-Torturer was to explain how the requirement to stop before the end could be reconciled with the apparent requirement to press on at every point. We now see that there is no such requirement to press on. At each point in the borderline-zone, it is vague whether expected utility will be maximized by advancing, and thus vague whether ST is required to advance. Similarly, it is vague at some points whether the shepherd can wait another thirty minutes, and whether you should have another piece of cake.

This sorites has a little more structure. Consider the predicate Better(m,n), which compares the net utility of two stages of the Self-Torturer:

> Better(m, n) *iff* stage m has higher utility than stage n.

In zone 1, it is *knowably true* that Better$(k+1, k)$. In zones 2 and 3, Better$(k+1, k)$ is borderline. But this vagueness has a special structure: for m sufficiently greater than n, Better(m, n) is *knowably false*, and Better(n, m) is knowably true. That is, in zones 2 and 3, each pairwise comparison is vague; but for any given stage in these zones, there is a further stage that is knowably worse.

That the predication 'Better$(k+1, k)$' is vague in zone 3 is part of what distinguishes the comparative case from the binary: even once ST has proceeded too far, into zone 3, the utilities of the options open to her at the next choice are vague. But repeatedly advancing is knowably, determinately disadvantageous. This is why the Self-Torture equivalent of 'it's too late to build a cairn, so we might as well enjoy television'—which would be 'we're already in pain, might as well go all the way to the end'—is such a mistake. This would (eventually) just make the torture situation worse.

5.7 Objections Perceptibility, and Second-Order Vagueness

Now, I'll discuss two objections. Quinn briefly considers an account similar to mine, which he calls *preference reversal*:

> The self-torturer starts out preferring early settings to 0, but ends up preferring 0 to late settings. At some intermediate setting his preferences must reverse. There must be some *s* such that he prefers *s* to 0 but 0 to *s+1*. And this must mean that the added money at *s+1* isn't worth the extra discomfort.[17]

This is misleading: it is indeterminate whether the extra money at *s+1* is not worth the extra pain; but by then this will have been indeterminate for some time. By the time ST

[17] Quinn (1990), pp. 81-82.

reaches this point—where he is indifferent between his pain and fortune on the one hand, and no pain and no money on the other—he will have gone much too far. Modulo this adjustment, this is the core of my view. Quinn rejects this proposal:

> This argument goes wrong, I think, in presupposing that for any positive setting s, the self-torturer (counting both pain and gain) determinately prefers s to 0 or 0 to s. But, empirically speaking, his preferences as between s and 0 can exhibit various kinds of *indeterminacy*. Not only is there no empirically determinable *first* setting that he disprefers to 0, there is no empirically determinable *first* setting at which these preferences become indeterminate. There is simply nothing in the way a single increment of current affects him to warrant such precise line drawing. [. . .] A first setting at which his overall preferences (as between 0 and s) reversed or went indeterminate, would have to be a setting at which his comfort (relative to 0) suddenly declined. There could be no other explanation.[18]

But there is no reason to think that preference-reversal requires this sort of determinacy. Ironically, Quinn took it that preference-reversal requires precision, and that this undermines the response, but the Soritical view I am defending can be seen as combining preference-reversal with vagueness: ST's preferences do reverse at some s, but that s is indeterminate or unknowable.

However, the quoted paragraph does suggest two more substantial objections to the Soritical view: an appeal to the imperceptibility of each stage in the Self-Torturer series, and to higher-order vagueness. I consider these in turn.

As discussed above, imperceptibility is not the issue under consideration in this paper: all that is needed to get the puzzle off the ground is that the pain at each stage seems to be *clearly outweighed* by the money offered. But it might be worried that even if my account can explain why the small-but-perceptible pain increase case is Soritical, the

[18] Quinn (1990), p. 82.

imperceptible case seems tougher. How can it be rational by any lights to turn down $10,000 for an imperceptible increase in electric current?[19]

But this is simply the difference between standard and *phenomenal* sorites (where the pairwise change between stages is too small to perceive). Phenomenal sorites can be constructed for all kinds of predicates, with 'is loud', 'is red', and 'is cold' being particularly common examples. Soritical solutions to the cairn-builder, the non-phenomenal Self-Torturer, and (depending on portion-sizes) the cake eater constitute a big weight of evidence that the solution is similar for the phenomenal Self-Torturer.

The vagueness that grounds the cases under consideration generally arises from either vague preferences—such as vagueness in the relative strengths of ST's preferences for more money and less pain—or from projects with vague completion-conditions, such as building cairns. The phenomenal sorites adds a layer of vagueness, grounded in perceptual thresholds, but it does not fundamentally change the issues. Even in the phenomenal case, the relevant tolerance principle is *not true*, defanging the puzzle as a quick argument against orthodoxy.

Second-order vagueness is that phenomenon where 'is a borderline-case' itself has borderline-cases. Second-order vagueness would mean that the boundaries between zones 1, 2, and 3 are themselves indeterminate or unknowable. The problem for my view is that second-order vagueness might ground another sorites *within* the borderline-zone, with a tolerance principle of 'if setting n is in the borderline-zone, then setting (n+1) is in the borderline zone'. So my solution might seem to just push the puzzle back a level.[20]

But once again, second-order vagueness does not undermine the core claim that the (first-order) tolerance principle is false; it simply complicates the issue of where and how it has false instances.

As a dialectical point, since it is admitted by all that the puzzle is somehow grounded in vagueness, general features of vagueness—such as its higher-order incarnations, and

[19]John Roberts and Ram Neta separately pressed this line of objection on me.

[20]Thomas Hofweber pressed this line of criticism on me.

connection to perceptual thresholds—are not likely to present a particular problem for the Sorites view. We may claim companions in guilt. For example, Quinn admits that 'there there is an (indeterminately limited) range of step-sizes for which the self-torturer's preferences are neither determinately transitive nor intransitive.'[21] Since his heterodox solution hangs on identifying the smallest step-size at which the victim's preferences are transitive in a filtered series, we can see that higher-order vagueness is a problem for the intransitive view, too: it will be indifferent which is the correct filtered series to use.

Having argued that the Self-Torturer is a Sorites, we can now see that this undermines the heterodox views of rational choice canvassed above.

5.8 How the Sorites Undermines Heterodox Views

If the Self-Torturer is a sorites on something like 'maximises utility', then according to indeterminism, it is indeterminate what course will maximize utility; according to epistemicism, we simply don't know what course will maximize utility, and the puzzle is a relatively standard one of action in the face of ignorance.

The main point is this: whatever the right account of vagueness, the relevant tolerance principle is *determinately, knowably false*: it is false that utility is always maximized by advancing, and thus false that Tenet requires that ST always advance. But the latter claim was the central plank of the challenge from vague projects against orthodoxy.

More specifically, it is not true (*pace* Quinn) that ST prefers each setting to the previous one. But it was the combination of these pairwise preferences with those at longer ranges that generated intransitivity. Since the puzzle is a sorites, Quinn's view can't be right. And there's no need for the heterodox permissions of Tenenbaum and Raffman to diverge from expected utility calculations: expected utility maximization does not demand always turning the dial, so no permission is required to stop.

The heterodox views are rather far-reaching. Quinn claims that there are intransitive preferences, and Tenenbaum and Raffman argue that 'the axioms of orthodox RCT do not

[21] Quinn (1990), p. 90 fn. 11.

apply to vague projects or ends', and 'rationality is not always purely calculative'.[22] But these far-reaching claims depend on the puzzle not being a sorities. If it is Soritical, then preferences are transitive, and rationality is purely calculative. At least, vague projects do not show that these claims are false. Instead, vague projects show that utilities are often indeterminate or unknowable. This is a less surprising result, and on the face of it, is compatible with orthodoxy.

To push an analogy: the heterodox views claim that orthodox rational choice theory renders the *wrong* verdict and must be disobeyed or modified; the Soritical view I am defending instead claims that in the penumbra, the orthodox verdict will be indeterminate or unknowable. Whereas the heterodox views claim that rational agents may or must 'disobey' Tenet, the Soritical view simply has it that Tenet may not be maximally informative, and may need to be supplemented with a decision theory for action in the face of indeterminacy.

We can also push this a little further: the orthodox utility-maximising view can explain why ST is determinately irrational if she stops outside the borderline-zone. At this point, I'll assume that some version of indeterminism about vagueness is correct—that vagueness is not simply ignorance—not because I think epistemicism is ruled out, but because epistemicists can easily explain the issue: it is simply one of action in the face of (perhaps necessary) ignorance.[23]

The borderline-zone (zone 2) is the only rationally permissible place for ST to stop advancing, but no point therein is determinately preferable to any other. At each point in zone 2, it is indeterminate whether *that* point is the point that maximises utility. In zone 2, there is no 'clear and repeatable reason' to advance, and no parity of reasoning. The borderline-zone is that zone where, somewhere but indeterminately where, the gradient of the pain graph begins to exceed that of the money graph, and the net utility graph

[22] Tenenbaum and Raffman (2012), pp. 112 and 111.

[23] In this section, I'll assume that this is not a case where an aggregative *compromise* is required; this is plausible since the compromise cases seem to require large differences in utility on different sharpenings. See chapter 4 for more.

turns down. At each stage in that zone, it is indeterminate whether this turning point has not yet been reached (in which case ST should advance), or that point is the present one (in which case ST should stop), or whether it has already been passed (in which case ST should stop, and retreat, if that option is available). At each stage in the zone, it is *not false* that that stage is the optimum trade-off of pain and money. If ST stops at any such point, it is indeterminate whether she maximises utility.

But if she stops before or after the borderline-zone, then it is *false* that she maximises utility. If she stops before the zone, then it is determinate that she could have done better by advancing further. Similarly, if she stops after the zone, in zone 3, then it is determinate that she has gone too far, and has not maximized utility.

No point in the borderline-zone is determinately superior to any other: each is borderline-optimal. But for any point in the borderline-zone, there is determinately both a point before and a point after (outside that zone) that is worse.

This class of stopping points is privileged: it is required to stop somewhere in that class; at each point, it is indeterminate whether *that* is the place to stop, so there is no determinate reason to prefer one point in the class to any other.

If that is right, then the immediate challenge that the Self-Torturer and other vague projects pose for orthodox rational choice theory dissolves. As we saw, that challenge hung on the claim that according to the orthodox theory, it is always rationally permitted (or required) to advance from a non-extremal stopping point. But that claim relies on a tolerance principle for a vague predicate, and tolerance principles have false instances. Our justification of ST stopping in zone 2 was given in entirely orthodox terms: if she does not stop there, it will be knowably false that she has maximized utility. If she does stop there, it is indeterminate or unknowable whether she has so maximized.

This does give rise to a second issue—what does rationality demand when some options determinately fail to maximise expected utility, but others are such that it is indeterminate whether they maximise? But this is the third of our three questions, and seeing the puzzle as a Sorites does does not hang on such details: there is a first range of settings which determinately do not maximise utility, a second range where it is indeterminate whether each setting does, and a third range which again determinately fail to maximise.

Since there is no determinate reason to choose one over another in the second range, the case may be compared to an indeterminist Ass of Buridan: since the best he can do is make it indeterminate whether he has maximised utility, that is what he should do, and the Self-Torturer ought to arbitrarily pick some point in the second zone. On an epistemic characterisation, it is certain that the optimum point lies somewhere in the second zone, but unknowable precisely where. He should maximise expected utility by making the likelihood that he maximises utility nonzero, which he does by stopping at any point in the designated zone.

5.9 Conclusion

This paper has made two main claims. The apparent challenge to orthodox rational choice theory posed by vague projects like the Puzzle of the Self-Torturer is an instance of the paradox of the sorites. And once this is seen, the immediate force of the challenge is blunted: utility maximisation does not always require ST turn the dial, or that you eat another piece of cake.

But there still remains a puzzle: on the view I have been defending, it is indeterminate whether each dial-turn in the borderline-zone has net positive or negative marginal utility. Yet it is determinate that if ST advances all the way through zone 2 and into zone 3, then this sequence of actions has negative marginal utility. This raises the question: how can it be that each action in a sequence is indeterminately-permissible, yet the sequence as a whole is impermissible?

This is a characteristic phenomenon of vagueness, and also arises in the incommensurability of value (which I argue is just a species of vagueness) as 'value-pumping'. Adam Elga has argued that no sense can be made of this phenomenon; in other work, I defend the application of such 'sequence' rules, in epistemology and elsewhere.[24] In particular, I argue that the Soritical structure of the Self-Torturer provides

[24]See Elga (2010) and Elson (Manuscript[a]).

more evidence that such rules must be coherent: they are needed to explain why ST must stop turning the dial, amongst other things.

But that is beyond the scope of the present paper. Here I have just been concerned to argue that the Puzzle of the Self-Torturer is a rather special instance of the sorites paradox. Moreover, it can be described entirely in the terms of orthodox rational choice theory. The Self-Torturer is a sorites on 'maximises utility'.

REFERENCES

Barnes, Elizabeth (2010). “Arguments Against Metaphysical Indeterminacy and Vagueness”. In: *Philosophy Compass* 5.11, pp. 953–964.

Blackburn, Simon (2010). “Dilemmas: Dithering, Plumping, and Grief”. In: *Practical Tortoise Raising*. Ed. by Simon Blackburn. Oxford: Oxford University Press. Chap. 3, pp. 47–63. DOI: 10.1093/acprof:oso/9780199548057.001.0001.

Braun, David and Theodore Sider (2007). “Vague, So Untrue”. In: *Nos* 41.2, pp. 133–156. DOI: 10.1111/j.1468-0068.2007.00641.x.

Broome, John (1997). “Is incommensurability vagueness?” In: *Incommensurability, incomparability, and practical reason*. Ed. by Ruth Chang. Harvard University Press, pp. 67–89.

— (2009). “Reply to Rabinowicz”. In: *Philosophical Issues* 19.1, pp. 412–417. DOI: 10.1111/j.1533-6077.2009.00175.x.

Carlson, Erik (2004). “Broome’s argument against value incomparability”. In: *Utilitas* 16, pp. 220–224.

— (2012). “Vagueness, Incomparability, and the Collapsing Principle”. English. In: *Ethical Theory and Moral Practice*, pp. 1–15. DOI: 10.1007/s10677-012-9352-9.

Chang, Ruth (1997). “Introduction”. In: *Incommensurability, Incomparability, and Practical Reason*. Ed. by Ruth Chang. Cambridge, MA: Harvard. Chap. 1.

— (2002). “The Possibility of Parity”. English. In: *Ethics* 112.4, pp.659–688. DOI: 10.1086/339673.

— (2004). “All things considered”. In: *Philosophical Perspectives* 18.1, pp. 1–22.

— (2005). “Parity, Interval Value, and Choice”. In: *Ethics* 115.2, pp. 331–350. DOI: 10.1086/426307.

Constantinescu, Cristian (2012). “Value Incomparability and Indeterminacy”. In: *Ethical Theory and Moral Practice* 15.1, pp. 57–70. DOI: 10.1007/s10677-011-9269-8.

De Sousa, Ronald B (1974). “The good and the true”. In: *Mind* 83.332, pp. 534–551.

Dorr, Cian (2003). “Vagueness without Ignorance”. In: *Philosophical Perspectives* 17, pp. 83–113.

Edgington, Dorothy (1997). “Vagueness by Degrees”. In: *Vagueness: A Reader*. Ed. by Rosanna Keefe and Peter Smith. MIT Press.

Elga, Adam (May 2010). “Subjective Probabilities should be Sharp”. In: *Philosophers’ Imprint* 10.5.

Elson, Luke (Manuscript[a]). “Sequential Action for Mushy Reasons”.

Elson, Luke (Manuscript[b]). “Vague Satisficing Consequentialism”.

Endicott, Timothy (2000). *Vagueness in Law*. Oxford University Press.

Fine, Kit (1975). “Vagueness, truth and logic”. In: *Synthese* 30.3, pp. 265–300.

Gert, Joshua (2004). “Value and Parity”. English. In: *Ethics* 114.3, pp.492–510. DOI: 10.1086/381697.

Gustafsson, Johan E. (Dec. 2013). “Indeterminacy and the Small-Improvement Argument”. In: *Utilitas* 25.4, pp. 433–445. DOI: 10.1017/S0953820813000034.

Jackson, Frank (1991). “Decision-Theoretic Consequentialism and the Nearest and Dearest Objection”. English. In: *Ethics* 101.3, pp.461–482. ISSN: 00141704. URL: http://www.jstor.org/stable/2381464.

Keefe, Rosanna (2000). *Theories of Vagueness*. Cambridge University Press.

Moss, Sarah (forthcoming[a]). “Credal Dilemmas”. In: *Nos*.

— (forthcoming[b]). “Time-Slice Epistemology and Action Under Indeterminacy”. In: *Oxford Studies in Epistemology*.

Portmore, Douglas (Aug. 2013). “Perform your best option”. In: *Journal of Philosophy* CX.8, pp. 436–459.

Quinn, Warren S. (1990). “The Puzzle of the Self-Torturer”. English. In: *Philosophical Studies* 59 (1), pp. 79–90. DOI: 10.1007/BF00368392.

Raz, Joseph (1985-86). “Value incommensurability: some preliminaries”. In: *Proceedings of the Aristotelian Society, New Series* 86, pp. 117–134.

Regan, Donald (1997). “Value, Comparability, and Choice”. In: *Incommensurability, Incomparability and Practical Reason*. Harvard University Press.

Rinard, Susanna (forthcoming). “A Supervaluationist Decision Theory for Imprecise Credences”. In: *Philosophers’ Imprint*.

Schoenfield, Miriam (2012). “Chilling out on epistemic rationality”. English. In: *Philosophical Studies* 158.2, pp. 197–219. DOI: 10.1007/s11098-012-9886-7.

Schroeder, Mark (2007). *Slaves of the Passions*. Oxford University Press. ISBN: 0199299501.

Seung, T. K. and Daniel Bonevac (July 1992). “Plural Values and Indeterminate Rankings”. In: *Ethics* 102.4, pp. 799–813. URL: http://www.jstor.org/stable/2381555.

Sinnott-Armstrong, Walter (Oct. 1985). “Moral Dilemmas and Incomparability”. In: *American Philosophical Quarterly* 22.4, pp. 321–329.

— (1988). *Moral Dilemmas*. Oxford: Basil Blackwell. ISBN: 9780631157083.

Sorensen, R A (1988). *Blindspots*. Oxford University Press.

Tenenbaum, Sergio and Diana Raffman (Oct. 2012). “Vague Projects and the Puzzle of the Self-Torturer”. In: *Ethics* 123.1, pp. 86–112. DOI: 10.1086/667836.

Tuck, Richard (1979). “Is there a free-rider problem and if so what is it?” In: *Rational Action*. Ed. by Ross Harrison. Cambridge University Press, pp. 147–156.

Ullmann-Margalit, Edna and Sidney Morgenbesser (1977). “Picking and Choosing”. In: *Social Research* 44.4, pp. 757–785.

Wasserman, Ryan (2004). “Indeterminacy, Ignorance and the Possibility of Parity”. In: *Philosophical Perspectives* 18.1, pp. 391–403. DOI: 10.1111/j.1520-8583.2004.00034.x.

White, Roger (2009). “Evidential Symmetry and Mushy Credence”. In: *Oxford Studies in Epistemology*. Vol. 3. Oxford University Press, pp. 161–186.

Williams, J. Robert G. (2010). “Degree Supervaluational Logic”. In: *The Review of Symbolic Logic* 4.01, pp. 130–149. URL: http://www.journals.cambridge.org/abstract_S1755020310000237.

— (Mar. 2014). “Decision-Making Under Indeterminacy”. In: *Philosophers' Imprint* 14.4, pp. 1–34.

Williamson, Timothy (1994). *Vagueness*.

Wright, Crispin (1975). “On the coherence of vague predicates”. In: *Synthese* 30, pp. 325–65.