ARTICLE

Noûs

Degrees of commensurability and the repugnant conclusion

Alan Hájek | **Wlodek Rabinowicz**

Australian National University and Lund University

Correspondence
Alan Hájek, Australian National University, Canberra, Australia
Email: hajek.alan@gmail.com

Abstract

Two objects of valuation are said to be *incommensurable* if neither is better than the other, nor are they equally good. This negative, coarse-grained characterization fails to capture the nuanced structure of incommensurability. We argue that our evaluative resources are far richer than orthodoxy recognizes. We model value comparisons with the corresponding class of permissible preference orderings. Then, making use of our model, we introduce a potentially infinite set of *degrees* of approximation to *better*, *worse*, and *equally good*, which we interpret as *degrees of commensurability*.

One payoff is the solution our approach provides to a paradox in population ethics, generated by Parfit's "Continuum Argument". Parfit imagines a sequence of populations, starting with one consisting of excellent lives and, by a sequence of apparent improvements, reaching a much larger population of lives barely worth living. What he dubs "the Repugnant Conclusion" is that the final population is better than the first. Developing Parfit's response, we argue that some of the populations in the sequence are merely *almost better* than their immediate predecessors. *Almost better* is not transitive (unlike *better*). We offer analogies to other 'spectrum arguments', Condorcet's paradox, and to developments in formal epistemology.

1 | INTRODUCTION

The value relations of *better*, *worse*, and *equally good* have been discussed for millennia; but systematic treatments of 'incommensurability' are rather more recent.[1] Two objects of valuation are *incommensurable* when neither is better than the other, nor are they equally good. As such, their values cannot be represented on the same scale, not even on the same ordinal scale.[2] Like the *via negativa* approach to speaking about God, incommensurability is thus characterized by what it is *not*. But so characterized, it is a shapeless catch-all—it has no internal structure, admitting of no finer-grained distinctions.

We can do better. We can characterize incommensurability more positively and give it a far more nuanced treatment. Moreover, revealing the fine structure of incommensurability can shed light on various ethical puzzles and puzzles regarding rationality. In this paper we focus on one such puzzle, generated by a notorious argument in population ethics. We offer a novel account of incommensurability, showcasing its fruitfulness with the solution that it offers to that puzzle.

The puzzle arises from what Parfit (2016) dubs "the Continuum Argument", leading to "the Repugnant Conclusion".[3] The argument has us imagine an initial population of people with excellent lives, and then a finite sequence of populations, each with a slightly lower life quality than its predecessor, but much larger and therefore putatively better on the whole. Eventually we reach a huge population of people with lives barely worth living. The conclusion is that this population is better than the initial one.

Many authors agree with Parfit that this conclusion is repugnant and that the Continuum Argument must be resisted. The lines of resistance are diverse. Some of the argument's critics retain the traditional menu of value relations: *better*, *worse*, and *equally good*. Others do not. Parfit's own response appeals to a further such relation, *imprecisely equal*, which we interpret to mean *incommensurable*. But this is still a widely-recognized value relation.

We submit that our evaluative resources are far richer than have been appreciated. We model value comparisons with a corresponding class of permissible preference orderings. We introduce a potentially infinite set of *degrees* of approximation to *better*, *worse*, and *equally good*, which we interpret as *degrees of commensurability*. In particular, for items *A* and *B*, the higher the ratio of orderings in which *A* is ranked above *B*, the closer *A* is to being better than *B*. If that ratio is close to but below 1, *A* is *almost better* than *B*. We show how to generalize this account to models with infinite domains and infinitely many permissible preference orderings. We argue that degrees of commensurability are independently motivated and, indeed, required to do justice to various axiological intuitions. We deploy these new value relations to provide a response to the Continuum Argument in the spirit of Parfit's, but more nuanced. We argue that some of the populations in the sequence are merely *almost better* than their immediate predecessors. Unlike *better*, *almost better* is not transitive and allows cycles. On this account, the argument leading to the Repugnant Conclusion is structurally similar to (a strengthened version of) Condorcet's Paradox. With an expanded toolkit of value relations, we do more justice to the intuitions that drive the argument and explain why it might seem compelling. So this paper provides both a contribution to our understanding of value relations and a resolution of a disturbing puzzle in moral philosophy. This resolution can also be applied to other 'spectrum arguments'.

2 | DEGREES OF COMMENSURABILITY

We propose that *incommensurability comes in degrees*: Incommensurable items can be more or less close to being commensurable. While incommensurability is usually regarded as an all-or-nothing relation, we contend that it is gradable.

Our intuitions about various cases seem to have more refined structure than a merely binary notion allows. Sometimes, when attempting to compare two alternatives, we are totally flummoxed, regarding them as not really comparable at all. In other cases, we are more inclined to form a preference one way or another, or to regard them with indifference, but we do so with some hesitancy. And in many of these cases, the hesitancy comes in degrees because incommensurability comes in degrees. Typically, in comparing alternatives with each other, we need to consider multiple relevant respects of comparison—multiple criteria of evaluation. Paradigm cases of incommensurability arise when the items we compare do unequally well on different criteria and it is not determined how these criteria should be weighed against each other: there are no fixed 'exchange rates' between them.

Who was more of a genius: Einstein or Bach? Plausibly, they are incommensurable—one was a great scientist, the other a great composer. How about Einstein or Chopin? Plausibly, they are still incommensurable, but perhaps it is easier to favor Einstein: while Chopin was undoubtedly a genius of piano composition, he arguably did not quite have Bach's range. How about Einstein or Schumann? This comparison is arguably easier again—while brilliant, Schumann was not quite as original as Chopin, let alone Bach. How about Einstein or Salieri, the mediocre composer made famous by *Amadeus*? That's easy—Einstein was the greater genius, period. We have proceeded by steps to closer and closer approximations to the 'better' relation with regard to genius.[4]

To be sure, even if there are multiple criteria or dimensions of evaluation, incommensurability might still be avoided if the 'exchange rates' between the criteria can be uncontroversially fixed—if it is determined how the different criteria should be weighed against each other.[5] But this is often not the case. Typically, there will be a range of different admissible assignments of relative weights to the relevant criteria. Each such weight assignment will give rise to a permissible all-things-considered preference ordering of the alternatives.[6]

In this paper, we focus on *value* comparisons. We appeal to the influential *fitting-attitude* analysis of value and value relations. (See especially Ewing, 1947. Cf. also Brentano, 1969[1889] and Scanlon, 1998.) Values and value relations are determined by the pro- and con-attitudes that are fitting or warranted regarding potential value bearers. Fittingness is understood as a requiring notion: fitting attitudes are ones we ought to have.[7] Thus, in particular, two items *A* and *B* are *equally good* iff they ought to be equi-preferred—i.e. it is fitting to be indifferent between them. *A* is *better than B* iff *A* ought to be preferred to *B*—i.e. it is fitting to have this preference.

Following Rabinowicz (2008; 2012), we may posit a class of permissible preference orderings of the domain of items under consideration. A preference is *permissible* if it is *not unfitting*—i.e., if it is not one we ought not to have. 'Unfitting' is a contrary of 'fitting'; they are related as 'forbidden' is related to 'obligatory'. Permissibility is thus the dual of fittingness. Our reading of 'permissible' will matter in what follows; on more common readings, unfitting attitudes might well be legally, socially, or even morally permissible.

We may now reformulate the value relations' definitions in terms of the permissible orderings. *A* and *B* are equally good iff they are equal-ranked in all permissible preference orderings—they ought to be equi-preferred. *A* is better than *B* iff *A* is ranked above *B* in all permissible preference

orderings—*A* ought to be preferred to *B*. However, if permissible preference orderings disagree in their ranking of *A* and *B*, these items are mutually *incommensurable*.[8]

So far, incommensurability, and hence commensurability, are *on-off* relations between two items. We now add that the *degree* of commensurability can be higher or lower depending on the extent to which different permissible orderings agree or disagree in their ranking of the items. If in nearly all permissible orderings *A* and *B* are ranked in the same way, their degree of commensurability is very high—for example, if *A* is almost always ranked above *B*, or they are almost always equal-ranked. But if there is more divergence in how *A* and *B* are ranked, their degree of commensurability is lower. (Equivalently, their degree of incommensurability is higher.)

We may measure the extent of agreement among the orderings. An especially simple measure is *the proportion* of the orderings that deliver a given relative ranking. (Here we assume for simplicity that there are just finitely many orderings, which follows if the domain of items is finite. We will relax this assumption in section 10.) If *almost all* rankings favor *A* over *B*—that is, if the proportion of such rankings in the set of all permissible rankings is close to 1—we will say that *A* is *almost better* than *B*.[9] Then *A* and *B* are incommensurable, but only to a small degree. Equivalently, they are commensurable to a high degree, though not fully so. Similarly, if *almost all* rankings favor *B* over *A*, we will say that *A* is *almost worse* than *B*. If *almost all* rankings treat *A* and *B* equally, we will say that *A* and *B* are *almost equally good*.

More generally, we can define entire spectra of degrees of approximation to betterness, worseness, and equal goodness. Unanimity among the permissible rankings corresponds to maximal commensurability. The closer to unanimity, the greater the commensurability.

Note that we are *not* starting with some intuitive, folk notion of 'almost better' (or 'almost worse' or 'almost equally good') and explicating that. Rather, we are explicating a broad spectrum of value relations; among them, we identify those that we call 'almost better' (close approximations to 'better'). When an item in this sense is almost better than another, it is ranked higher on nearly all permissible preference orderings. We can explain why agents confronting such items might be strongly inclined to prefer the former item to the latter, but perhaps with slight hesitancy. The hesitancy reflects acknowledging that one could reasonably have other preferences—typically due to other ways of reasonably aggregating dimensions of evaluation.

The literature on evaluative relations traditionally trafficked mostly in the relations of *better*, *worse*, and *equally good*. Examples of such relations are harder to come by than we ordinarily think, since many cases that we consider are tainted by some measure of incommensurability— some latitude in permissible preference. That latitude is often so small that we do not notice it or we ignore it—we could easily conflate 'almost betterness' with 'betterness'. Or even when we pay attention to it, we may feel forced to pigeon-hole our judgments into one of the traditional categories nonetheless. And even when incommensurability is explicitly acknowledged, it is a very broad type of a value relation and it cries out for distinctions. It is important to recognize that incommensurability comes in degrees.

This is the set-up for our solution to the Continuum Argument's paradox. We initially discuss our solution more informally (section 5); then we present it more rigorously with a formal model (section 6). But first, we examine the paradox more closely (section 3), and then rehearse two putative solutions that set the stage for ours (section 4). We suggest an intuitive interpretation of our solution (section 7), draw analogies to other 'spectrum arguments' and to Condorcet's Paradox (section 8), contrast our solution with the vagueness approach to the Continuum Argument (section 9), and describe how our modelling of degrees of commensurability can be generalized (section 10). Section 11 concludes.

3 | THE CONTINUUM ARGUMENT FOR THE REPUGNANT CONCLUSION

In more detail, the Continuum Argument runs as follows. Consider hypothetical populations in which everyone has a life worth living and the life quality is the same for each population member. Assume that decreases in a population's life quality are always *pro tanto* worsenings, while increases in its size are always *pro tanto* improvements (if everyone's life is worth living). Now, start with a fairly large population P_1 of people who all live excellent lives. There exists a better possible population P_2 with slightly worse lives: the loss in quality is more than compensated by a sufficient gain in size. Moreover, there exists a still better possible population P_3 with slightly worse lives than P_2, but sufficiently larger to compensate for this. And so it goes. Repeatedly decreasing the quality of lives but sufficiently increasing the population size, we proceed in finitely many steps to a huge population of drab lives that are barely worth living. We arrive there by a sequence of improvements, so this final population is better than the first population—or so the Continuum Argument concludes. But Parfit maintains that this conclusion is *repugnant*—indeed, that the final population is *worse* than the first one—so this argument must be resisted. We have a paradox: how can a sequence of ostensible improvements lead to a final population that is apparently worse than the first one?

We propose a novel solution to the paradox. It builds on Parfit's suggestion, which we present in the next section, that (at least some of) the ostensible improvements are really cases of incommensurability. But it is based on our novel view that incommensurability comes in many degrees, which approximate the *better* relation to greater or lesser extents. In particular, there is an *almost better* relation, which closely approximates the *better* relation—so closely that we might conflate them.[10] At least some of the populations in the sequence are merely *almost better* than their immediate predecessors. Then the last population need not be better than the first—indeed, it may be worse. Our solution avoids the Repugnant Conclusion and yet diagnoses why the argument for it is so seductive. More generally, we offer a new way of thinking about incommensurability, thus refining the space of value relations far beyond their orthodox treatment.

4 | TWO PUTATIVE SOLUTIONS

We turn to a critical discussion of two putative solutions to the puzzle generated by the Continuum Argument. While other putative solutions have also been proposed, considering these two will help to set up our own solution. We do not insist that ours is the only viable one; indeed, it may well complement or refine other solutions. This is true especially of Parfit's proposal, which we will discuss below.

4.1 | Temkin's solution

Temkin's (1996; 2012) solution, motivated by several structurally similar *spectrum arguments*, is to deny that *better* must be transitive if it is understood as an "essentially comparative" concept—i.e., if judgments of betterness are not derived from independent evaluations of each item. In a spectrum argument, two competing variables contribute to the value (disvalue) of a state of affairs. We are supposed to imagine a sequence of cases in which one variable is slightly decreased at

each step, while the other is significantly increased, so that overall things get *better* (worse). Yet eventually we reach a case that is *worse* (better) than where we started. It might, for example, be a sequence of lives, each followed by one that has slightly lower quality but is much longer; or a sequence of pleasures, each followed by one that is slightly less intense but lasts much longer. We are supposed to intuit that things get better at each stage, but by the end they are worse than they were at the outset. Or we might replace pleasures with pains of slightly decreasing intensity but greatly increasing duration, as in the Rachels-Temkin spectrum argument, reversing these verdicts. (Cf. Rachels, 1998. See also Quinn, 1990.)

Applied to the Continuum Argument, a failure of transitivity means that even if every population in the sequence is better than its predecessor, it does not follow that the last population is better than the first. Indeed, we can argue that we have a betterness cycle: each population is better than its predecessor, but the last population is worse than the first.

4.2 | Problems with Temkin's solution

Some authors find this solution more repugnant than the Repugnant Conclusion itself. A number of authors criticize directly Temkin's spectrum argument (e.g. Nebel, 2018; Voorhoeve & Binmore, 2006; Handfield, 2014; Handfield & Rabinowicz, 2018). The transitivity of *better* is commonly considered to be sacrosanct—see e.g., Broome (2004), Dreier (2019), and Huemer's (2008) debunking of non-transitivity intuitions specifically regarding the Continuum Argument. Handfield (2016) argues that viewing *better* as an essentially comparative concept does not threaten its transitivity.

Since we identify *better* with unanimity among all permissible preference orderings, the transitivity of *better* follows from the transitivity of permissible preferences. That well-behaved preferences are transitive is taken as axiomatic in the leading decision theories (see Ramsey, 1926; von Neumann & Morgenstern, 1944/2007; Savage, 1954; Jeffrey, 1983; and even in heterodox theories such as that of Buchak, 2013). It is also defended in various ways. There are arguments from money pumps (Gustafsson, 2010; Rabinowicz, 2000; Dougherty, 2014; Gustafsson & Rabinowicz, 2020), from non-transitivity leading to necessary violations of one's own preferences (Tullock, 1964; Fishburn, 1988[11]; Gustafsson, 2013), and from consequentialist foundations for expected utility (Hammond, 1988). Another line of argument instead treats transitivity as a necessary feature of preferences rather than a normative requirement. Understanding 'preferring' as 'favoring more' (or 'disfavoring less'), rather than as a choice disposition, makes transitivity analytic in virtue of the meaning of 'more'—see Rabinowicz (2012).

We will assume from now on that permissible preferences are transitive and that betterness consequently is a transitive relation (while granting that this remains a lively area of debate). In the Continuum Argument (and indeed everywhere), *better* is transitive, and yet it might *appear* that it fails to be so. Our solution will uphold this transitivity, while explaining the appearance of its failure.

4.3 | Parfit's solution

Parfit (2016) attempts[12] to block the Continuum Argument by appealing to "imprecise equality" in value. By this he seems to mean something very close to incommensurability; in any case, imprecise equality is meant to entail incommensurability.

> *Precisely equal* is a transitive relation. [...] But if *X* and *Y* are *imprecisely* equally good, so that neither is worse than the other, these imprecise relations are *not* transitive. [...] Two things are imprecisely equally good if it is true that, though neither thing is better than the other, there could be some third thing which was better or worse than one of these things, though not better or worse than the other. (Parfit, 2016: 14f)

Is the comparison with "some third thing" intended to be a part of the definition of imprecise equality? Or is it rather meant to be a useful (partial) test that allows us to distinguish this relation from precise equality? Here, we will assume that it is such a test, and that imprecise equality is the same thing as incommensurability.

Parfit denies that small quality losses can always be compensated by increases in quantity. But he also denies that a small quality loss can make a population worse than its predecessor irrespective of how much larger it is. Instead, he suggests that at some, if not all, points in the Continuum Argument's population sequence we will encounter an incommensurability ("imprecise equality") between adjacent populations.[13] This allows him to reject the Repugnant Conclusion—indeed, to claim that the final population is worse than the first.

It won't suffice for Parfit to bring in incommensurability at just one point in the sequence; Handfield (2014) shows that this leads to inconsistency, assuming the transitivity of betterness. However, he also shows that this inconsistency is avoided if there are at least *two* points of incommensurability. In fact, Parfit thinks that incommensurability ("imprecise equality") might well come in at *every* point:

> [Continuum] arguments assume that [...] any slight loss of quality could be outweighed by a sufficient gain in quantity. [...] But we should deny that such truths would be precise. We should then claim that *no* slight loss in quality would either be outweighed by, or outweigh, any such gain in quantity. It would not be better if there existed many more people whose quality of life would all be lower, since two such worlds would at most be imprecisely equally good. (120, our emphasis)[14]

Parfit's suggestion that *no* slight loss in quality can be outweighed by any gain in quantity is radical. It rules out all instances of improvement gained from tiny decreases in quality and huge increases in population size—even reducing each person's excellent life by one second of pleasure while increasing the population by millions. And indeed this suggestion is unnecessarily radical. It is not needed to block the Continuum Argument. It is enough if incommensurability intervenes at more than one point.

4.4 │ Unresolved issues with Parfit's solution

We believe that Parfit's solution is essentially correct, at least on its less radical version. However, as it stands, it leaves some issues unresolved.

For starters, the incommensurabilities themselves need to be explained—why they occur at the steps at which they occur. Moreover, we need to understand why the premises of the Continuum Argument are prima facie plausible: why we apparently intuit at each step along the sequence that it involves an *improvement*. Why don't we recognize the incommensurabilities as such? Or why aren't we instead flummoxed by these comparisons? Those would be more natural responses to incommensurability, one might think. And why, as we move along the population sequence, do

we uniformly make mistakes in the same direction, always judging what are in fact incommensurabilities to be *improvements*? Why don't we instead sometimes mistake them for worsenings? And why don't some of us make mistakes in one direction and others in the other direction? In typical cases of incommensurability, such as comparing the genius of Bach and Einstein, we either feel no pull either way, or we feel pulled both ways. If forced to choose, an individual may feel inclined to choose differently on different occasions, and different people will choose differently from others. But this is not the pattern of responses to the comparisons in the Continuum Argument. Parfit's solution cries out for an error theory.

Temkin attributes a major error to those of us who are seduced by the argument. On his view, we take *better* to be transitive, when in fact it is not; we are mistaken about the very logic of this relation. Parfit attributes to us a smaller error, of not noticing incommensurabilities in the population sequence and mistaking them for improvements. But he does not explain why we make this mistake. Our solution, to which we now turn, also attributes to us an error, but it is a relatively minor one, and it is easy to see why we make it.

5 | OUR SOLUTION: DEGREES OF COMMENSURABILITY AND THE CONTINUUM ARGUMENT

We are supposed to intuit that each population along the argument's sequence is *better* than its predecessor. But if we attend to it more closely, is that so clear? Perhaps your judgment that the second population is better than the first is slightly hesitant. Part of you that weighs heavily the *loss* in the quality of life may balk a little. Or even if no part of you weighs quality of life so heavily, you might think that someone else might weigh it more heavily than you do without thereby making a mistake. And so it goes as we move down the sequence. At a first pass, we propose that each population is in fact *incommensurable* with its predecessor, but *only slightly*. Each is not better than its predecessor, but it is *almost better*. In fact, it is so close to being better that we mistake the one relation for the other. We do not notice or we ignore the reasonable weighings that do not favor the second population over the first, because they are overwhelmed by those that do. But it is a minor mistake: *almost better* is almost *better*! Our intuitions are wrong, but almost right. This is the error theory that Parfit needed.

It is natural to think that on the standard ungraded conception, we should either intuit incommensurability when it holds between alternatives, or be simply flummoxed by their comparison. These are *not* our reactions to the successive comparisons that lead to the Repugnant Conclusion. Rather, we feel relatively secure in our 'betterness' judgments. The *graded* conception that we offer explains this: what is almost better might well seem very much like 'better'.

We said that this was our "first pass". While Parfit apparently favors incommensurability holding at *every* step, we pointed out that this position is radical, and stronger than a rejection of the Continuum Argument requires. It sufficed for there to be incommensurability in some (at least two) steps in the population sequence, and *betterness* in the rest. Correspondingly, we need not commit to incommensurability in every step in our solution. We could allow *almost betterness* in some steps, and *betterness* in the rest, and still avoid the Repugnant Conclusion.

Either way, while we agree with Parfit that incommensurability is the key to resisting the Repugnant Conclusion, unlike him we can explain why we so readily intuit that each step involves an improvement. It either does, or almost does.

The ungraded conception of incommensurability in Parfit's solution also rendered mysterious why, both individually and collectively, we should uniformly mistake that relation for betterness.

It was unclear why we don't exhibit a random pattern of mistakes, or a uniform pattern of mistakes in the other direction (worseness). Our account gives a simple explanation: our intuitions, while mistaken, are drawn towards *betterness* because it is this familiar relation that *almost better* closely approximates.

Now, it may seem that a sequence of populations getting *almost better* or *better* at every stage will lead us to a final population that at least is *almost better* than the first. That conclusion, while perhaps not as repugnant as the original one, is still repugnant. But fear not: *almost better*, unlike *better*, is not transitive. In fact, as we will see, such a sequence may lead to a final population that is *worse* than the first—not even merely *almost worse*.

Now let us explore our solution in terms of a graded conception of (in)commensurability in more detail.[15]

6 | OUR SOLUTION: MORE FORMAL PRESENTATION

We now offer a formal model of our solution. The domain we consider consists of populations in which everyone has a life at the same quality level. Life quality and size are the only relevant respects of comparison for such populations, and each permissible preference ranking of populations is generated by some admissible way of weighing these respects.

In our model, we can prove that *almost better* is not transitive, unlike *better*.[16] We may easily have a population sequence such that almost all permissible rankings favor population 2 over population 1, almost all favor population 3 over population 2, and so on to the final population; yet *all* these rankings may favor population 1 over the final population. The small differences between *almost better* and *better* may accumulate over a sequence. A series of almost-improvements can therefore result in a worsening, relative to the starting point.

Here is an example. Suppose there are four options in the sequence, 1, 2, 3, and 4, and the class of permissible preference orderings consists of the three orderings O_1, O_2, and O_3 (with higher-ranked options placed higher):

O_1	O_2	O_3
3	2	1
2	1	4
1	4	3
4	3	2

In the sequence 1, 2, 3, 4, any two adjacent options are incommensurable: there is an ordering that ranks one of them higher and another ordering that ranks it lower. But a majority of permissible orderings ranks 2 above 1 ($\{O_1, O_2\}$), a majority ranks 3 above 2 ($\{O_1, O_3\}$), and a majority ranks 4 above 3 ($\{O_2, O_3\}$). However, the intersection of these three majorities is the empty set. Indeed, 1 is preferred to 4 in every permissible ordering. In other words, 1 is better than 4, yet 2 is almost better than 1, 3 is almost better than 2, and 4 is almost better than 3. Here we assume that a majority of two-thirds suffices for 'almost better'. If we raise the bar for what counts as 'almost better', we can exemplify non-transitivity by suitably increasing the number of options—we show how below.[17]

In the Continuum Argument, we can think of options 1 – 4 as populations that rapidly increase in size but slowly decrease in life quality. Since lives in the first population are excellent and in the last barely worth living, this sequence is obviously too short, but it can be made longer. The

recipe we have used to construct orderings $O_1 - O_3$ of four options can be used for any number n of options with $n \geq 4$. We start with the ordering in which option $n - 1$ comes first, followed by $n - 2$, $n - 3$, and so on, with option 1 coming last. (Note that in this ordering option n does not yet appear.) We then generate new orderings using the cyclical permutation of $n - 1$ into $n - 2$, of $n - 2$ into $n - 3$, ..., of 2 into 1, and of 1 into $n - 1$. Each time we move to the next ordering using the same permutation. In this way we construct $n - 1$ orderings of $n - 1$ options. Then to each ordering we add option n immediately below option 1. When one increases the number n of options in the sequence, the majorities favoring each option over its predecessor become more overwhelming. For every option k such that $k > 1$, k is ranked above $k - 1$ in all $n - 1$ orderings but one. Thus, k is ranked above $k - 1$ by $(n - 2)/(n - 1)$ of the orderings. This proportion gets closer to 1 as n increases. And, of course, option 1 is ranked above option n in all $n - 1$ orderings, as required. Thus, we can set up such an example for any interpretation of 'almost better', however demanding.[18]

So we can model Parfit's judgment that the last population in the sequence is *worse* than the first. Of course, we can also model the Continuum Argument's "repugnant" judgment that the last population is better than the first—just let each population be outright better (not merely almost better) than its predecessor. And we can model the judgment that the first and last populations are incommensurable: we simply use our recipe but terminate the sequence of almost-better populations before it reaches a population that is unanimously ranked below the first population. Indeed, we can model any judgments regarding the comparison of the last population with the first. It could be worse than the first (as Parfit would have it), *almost* worse, paradigmatically incommensurable (with a more or less even split between the orderings favoring one of these populations or the other), *almost* better, or better (as the Continuum Argument would have it). It could approximate being better, worse, or equally good to any specified degree.[19] The traditional relations of *better*, *worse*, *equally good*, and *incommensurable* do not capture just how nuanced value relations may be.

Thus, we can diagnose the slight unease one might feel each time one judges a population to be better than its predecessor, and the significant discomfort one ought to feel when one then judges the final population to be worse than the first and realizes that one's judgments are inconsistent. One is trying to identify *almost better* with one of the traditional relations; the best one can do is to identify it with *better*. But now we can find a place for it in the spectrum of value relations: approximating *better* but not identical to it.

We have modelled Parfit's solution that there is incommensurability at *every* step, but we can also model the less radical and arguably more plausible alternative: that there is incommensurability at some but not all steps. We could allow that for *some* comparisons in the sequence, one population is simply *better* than its predecessor: on that ranking, all permissible orderings agree. It's just that at some *other* points in the sequence we have the next population merely *almost better* than its predecessor: most, but not all, preference orderings place it higher. The following model displays this structure for five populations:

O_1	O_2	O_3
4	3	2
3	2	1
2	1	5
1	5	4
5	4	3

2 is better than 1 (not just almost better), 3 is almost better than 2, 4 is almost better than 3, 5 is almost better than 4, and 5 is worse than 1 (not just almost worse).

Our formal explication of *almost better* is based on a purely ordinal approach to preferences. If we admitted cardinal structure to preferences, we would need a more complicated formal account. To determine how close A is to being better than B, we could consider not merely the proportion of permissible orderings in which A is preferred to B, but also the relative strengths of preference for A and for B, respectively, in different permissible orderings. However, since better on the fitting-attitude analysis is a purely ordinal concept of unanimity in permissible preferences, near unanimity will still be enough for almost betterness. Consequently, the latter relation will still be non-transitive. For our purposes here, the simpler, purely ordinal explication should suffice. Its simplicity also makes it easier to see how it formalizes the more informal reasoning in the previous section.

7 | AN INTERPRETATION

As we have seen, each population in a sequence can be almost better than its predecessor, even though the last population is worse than the first. So far, we have provided formal models that display this phenomenon with suitable collections of permissible orderings of populations. This demonstrates that the idea is consistent. However, it would be nice to provide an *interpretation* that makes it vivid and plausible how such orderings of populations could arise.

Return to our first example:

O_1	O_2	O_3
3	2	1
2	1	4
1	4	3
4	3	2

Each of the orderings is based on some way of weighing life quality against population size. Suppose that each preference ordering postulates a *threshold* in quality, below which lives get radically less favored. At the point where we cross that threshold, no increase in quantity of lives compensates for that loss in quality. But if lives in both of the compared populations are above the threshold or if they already are below the threshold, slight losses in quality can be outweighed by sufficiently large increases in quantity. Different rankings manifest different perspectives on where the threshold should be placed. According to O_1, the threshold is crossed between lives in 3 and 4; according to O_2, between lives in 2 and 3; according to O_3, between lives in 1 and 2. In all three orderings, at the steps at which the threshold isn't crossed, the next population outranks its predecessor. Consequently, almost all orderings (all but one) favor 2 above 1, 3 above 2, and 4 above 3. But they all agree that the threshold is crossed *somewhere* in the sequence, so they all rank 4 below 1.

For example, consider populations of music listeners, with different qualities of music listened to by different quantities of people throughout their lives. There is a number of people listening to Bach (population 1), a larger number of people listening to Strauss instead (population 2), an even larger number of people listening to Kenny G[20] (population 3), and finally a huge number of people listening to the local army band (population 4). All preference orderings presuppose that the quality of music and thereby the quality of a life spent listening to this music decrease

as one moves along this sequence. But as the quality of music declines, the different orderings have different standards for what counts as *muzak*. And let's assume that a life spent listening to muzak is drab.[21]

According to O_1, Bach, Strauss and Kenny G are music, but the local army band is muzak. Thus, the threshold below which a life becomes drab is crossed between 3 and 4.

According to O_2, Bach and Strauss are music, but already Kenny G is muzak. The threshold is crossed between 2 and 3.

According to O_3, Bach is music, but already Strauss is muzak. The threshold is crossed between 1 and 2.

All orderings agree that a slight worsening of music can be compensated by the increase in the number of people listening to it, and that a similar compensation is possible once things have declined to muzak: a slight worsening in muzak can be compensated by sufficiently many more people listening to it. It's the drop from music to muzak that's crucial. *That* drop cannot be compensated by any increase in population size.

More generally, a threshold might be a point at which some 'on-off' cherished value is lost, where this value is a step function of some other graded variable. For example, if one's level of connectedness with others is gradually decreased from a high initial level, there might come a point at which it becomes so low as to make one unable to flourish. A small decrease in the level of connectedness, if it involves crossing this point, deprives one's life of the cherished value of flourishing. (For another example, involving pain levels, see Handfield, 2014; Handfield & Rabinowicz, 2018.)[22]

Permissible orderings might all agree that the relevant value function is a step-function, while disagreeing on its step's locations: they could all agree that a threshold (say, between flourishing and not flourishing) is crossed somewhere in the population sequence, but disagree on where it occurs. Thanks to the agreement, they all agree that the final population, however large, is worse than the first; but thanks to the disagreement, they may yield incommensurabilities between some populations and their predecessors. Indeed, if at a certain point in the sequence, a threshold is crossed in at least one ordering but not in some of the others, then the next population, if large enough, will be incommensurable with its predecessor and will persist in being incommensurable however large we might make it. Size increases can make it almost better than its predecessor, but they can never make it better. (For this notion of persistent incommensurability, see Rabinowicz (forthcoming). See also Handfield & Rabinowicz, 2018.)

It is thus possible to provide an interpretation of our solution. It requires quality thresholds in all permissible orderings, but at different points in the sequence. Could we have done without postulating thresholds? No, not as long as we work with a model in which all permissible preference orderings are assumed to be both transitive and complete. (An ordering is complete if it has no gaps: for every two items in its domain, it ranks one of them higher or ranks both equally.) For suppose O is any such ordering that ranges over all populations, of arbitrary size, in each of which everyone has a life worth living, of the same quality. Consider a sequence of such populations that satisfies two conditions:

(i) For every two adjacent populations P_i and P_{i+1}: O ranks P_{i+1} at least as highly as P_i, unless O ranks *no* population at P_{i+1}'s quality level, however large, that highly.

(ii) O ranks the last population below the first.

Since O is transitive, (ii) implies:

(iii) there is some population P_{i+1} that O does not rank at least as highly as its predecessor P_i.

For otherwise, by transitivity, O would rank the last population in the sequence at least as highly as the first, contrary to (ii). But then by the completeness of O, (i) and (iii) imply that we pass a threshold when we move from P_i to P_{i+1}: O ranks P_i above P_{i+1} and above *any* population at P_{i+1}'s quality level, whatever its size.[23]

Things change, to be sure, if we allow for incomplete preference orderings. Then persistent incommensurabilities in the value ordering can very easily, indeed trivially, be accommodated without introducing thresholds. We can simply allow preference orderings that themselves exhibit persistent gaps. But such preferences would cry out for an explanation. We might think of them as representing the preferential state of an agent who is undecided between different complete preference orderings. But then the preferences between which she is undecided would have to exhibit thresholds. Thus, postulating thresholds might become inescapable at some point.[24]

8 | ANALOGIES: SPECTRUM ARGUMENTS, CONDORCET'S PARADOX

As we noted in section 4.1, the Continuum Argument is an example of a *spectrum argument*. We may clearly extend our model to pronounce on other such arguments, and we may offer parallel diagnoses to ours regarding the Continuum Argument. For example, return to the sequence of pains of slightly diminishing intensity but greatly increasing duration. We may diagnose that at some steps in this sequence, a pain is not *worse* than its predecessor, but merely *almost worse*—and again, *almost worse* is not transitive. Hence, our proposal may solve a class of analogous paradoxical cases, each with its own 'repugnant conclusion', while maintaining the transitivity of *better* and *worse*.

Here is another striking analogy to a slight strengthening of Condorcet's voting paradox. Suppose there are three voters and three options, 1 – 3, which in the voters' respective rankings are ordered as in O_1 – O_3:

$$\begin{array}{ccc} 3 & 2 & 1 \\ 2 & 1 & 3 \\ 1 & 3 & 2 \end{array}$$

Note that the second ordering is generated from the first by permuting 3 into 2, 2 into 1 and 1 into 3. The same cyclical permutation generates the third ordering from the second. The paradox shows that majority voting can lead to a cycle: there is a majority for 3 over 2, for 2 over 1, and for 1 over 3. This seems paradoxical since none of the individual voters has cyclical preferences. In the example we have been using, we have simply added the fourth option, 4, and placed it in each ordering just below option 1. Now there also is a majority for 4 over 3 and *unanimity* for 1 over 4 (not merely a majority). We thus have a strengthened version of Condorcet's paradox. (Cf. Katz, 2011: 4.)

We might take the analogy between our proposal and Condorcet's paradox seriously and use it to illuminate our proposal. We might consider each permissible preference ranking as corresponding to the preferences of a jury member; the set of all permissible rankings determines the jury's collective judgments. For example, jurors may draw the line between music and muzak at

different places: one between Bach and Strauss, another between Strauss and Kenny G, another between Kenny G and the local army band.[25]

The solution to Condorcet's paradox on the same lines as ours above is that the collective jury judgments should be nuanced: if the majority (but not all) of the jurors rank, say, 2 above 1, the collective judgment should not be that 2 is better than 1, but only that it is almost better (if the majority is large enough). Going in the other direction, we may regard the Continuum Argument as a version of Condorcet's paradox.

We might find the 'jury' analogy illuminating even in the case of the ambivalent judgments of an individual. We have imagined you feeling various degrees of unease in your comparisons of options. We might regard this as a kind of fragmentation of your mental state. It's as if you have a group of somewhat conflicting 'jurors' in your head, each corresponding to a permissible preference ordering. Or without the metaphor, *you* are somewhat conflic*ted*. Our model could be interpreted as representing overall judgments in the face of such inter-personal or intra-personal conflict.

9 | VAGUENESS VS INCOMMENSURABILITY

Some philosophers question the very existence of incommensurabilities in value. They suggest that purported incommensurabilities are instances of vagueness (indeterminacy) in value comparisons (Broome, 1997, 2004; Sugden, 2009; Qizilbash, 2012; Elson, 2017; Dorr, Nebel & Zuehl, MS). Some consider supervaluationism to be the best way to model vagueness—this especially brings out similarities to our own approach. (Cf., in particular, Broome, 1997, 2004, who has championed the supervaluationist approach to vagueness in value comparisons. See also Andersson, 2017.) Thus, instead of denying that an item *A* is either better than, worse than, or equally good as *B*, one might say that it is *indeterminate* whether *A* is better, worse, or equally good as *B*. Nevertheless, one might still insist that it is determinate that one of the three standard value relations obtains between *A* and *B*. On this view, *A* and *B* are commensurable, but it is vague in what way: whether *A* is better than *B*, worse than *B*, or equally good. If this position is interpreted on supervaluationist lines, different admissible precisifications of the value ordering all agree that exactly one of these three statements holds but disagree on which it is. On this interpretation, we can define a statement's degree of determinacy as the proportion of precisifications on which it holds.[26] One might then define *almost-determinacy* as a degree of determinacy that is close to 1.

This account could deal with the Continuum Argument in a structurally analogous way to our proposal. Supervaluationists can interpret permissible rankings in our model not as preference orderings, as we do, but as admissible precisifications of the vague value ordering. Where we say that *A* is *almost better* than *B*, they would say that it is *almost determinate* that *A* is better than *B*. And they could then use our model to show how there can be a sequence of populations such that, for each consecutive population, it is determinate or almost determinate that it is better than its predecessor, despite it being determinate (and not merely almost determinate) that the last population is worse than the first. This could also provide an error theory parallel to ours. At some points in the sequence, a population is not *determinately better* than its predecessor, but only *almost determinately better*; we are apt to conflate these evaluations.

There is certainly room for such a vagueness account as an alternative to our incommensurability account. We cannot conclusively adjudicate between the two proposals in this paper—that would require a much more extended discussion. As we have said earlier, we are happy to be ecumenical about solutions to the paradox generated by the Continuum Argument.

For his own part, Parfit (2016: 113) takes a stand against the vagueness interpretation of "impre-cise equality": "Such imprecision is not the result of vagueness in our concepts ..." If two things are imprecisely equal, neither is better than the other, nor are they (precisely) equally good. It is not vague which of these three value relations obtains between those things; none of them does. But the reason we also take this stand is not simply that we want to stay close to Parfit's way of blocking the Continuum Argument. We favor the incommensurability account primarily because we accept the fitting-attitude analysis of value relations. On that analysis, as we have seen, incom-mensurability arises between two items if it is permissible to rank them in different ways. And we think that, in many comparisons, the choice of relative weights assigned to different relevant respects of comparison can permissibly vary, within certain limits. If one item does better in some of these respects and the other item does better in others, and if the former respects can just as well be given a greater or a smaller weight than the latter, it will be permissible to prefer one item to the other, all things considered, and likewise permissible to have the opposite preference. There will not be a unique fitting preference ranking of the items. Thus, the existence of incommensu-rabilities and indeed their common occurrence is to be expected on the fitting-attitude account of value relations. This applies, in particular, to comparisons between populations. It is optional (within some limits) how to weigh the quality of lives and their number. This leeway in relative weights grounds and justifies our incommensurability judgments.

The presence of incommensurabilities does not *exclude* the possibility of vagueness in value comparisons. The two phenomena can be jointly present.[27] We import vagueness into our model if we make the class of permissible preference orderings fuzzy—allowing different precisifications. (Cf. Rabinowicz, 2009a.) It may then be that on some precisifications of that class, two items are commensurable, while on other precisifications they are incommensurable.[28] Applied to the pop-ulation sequence of the Continuum Argument, this means that it may be indeterminate at exactly which points in this sequence incommensurability intervenes, even though it is determinate that it does intervene at some points or others.

We will not complicate our model with this dimension of vagueness. However, we will gener-alize the model in the next section.

10 | GENERALIZATIONS

Up to a point, our job is done. We have provided a 'proof of concept': we have shown how we may model degrees of commensurability, and how they may represent a consistent set of value judgments that approximate the intuitions underlying the Continuum Argument.

However, we can do more. So far, we have assumed that all permissible preference orderings are given equal weight, rather like the equal weight given to the voters in a democracy, or to the members of an egalitarian jury. Then we may simply consider the *proportions* of them that support a given verdict. But perhaps some orderings could be given more weight than others—perhaps, to paraphrase Orwell, all voters/jurors are equal, but some are more equal than others. For example, we may imagine a society in which voters or jurors are given weights according to how senior they are. In our case, permissible preference orderings might be given differential weights according to how reasonable they are—how reasonably they balance different criteria that are relevant for comparing the options under consideration. (Among admissible ways of such balancing, some might still be more sensible than others, even though all must be sensible enough to be admissi-ble.) These weights, which track the degrees of reasonableness of different permissible preference

orderings, specify how much importance these orderings should be given in determining value relations among the options.

We have also assumed that there are only finitely many populations to be ranked. But we are imagining *possible* populations, and there are infinitely many of those. We now generalize our modelling, to allow *unequal* weightings of the permissible preference orderings, and to allow *infinitely many* populations in our orderings. The domain on which these orderings are defined is thus infinitely large and the class of permissible orderings of this domain may also be infinitely large.

To obtain this generalization, we need to posit a non-negative, normalized, additive measure over sets of permissible preference orderings. It assigns to each set its weight. (The weight of an ordering is then just the weight of the singleton set that contains this ordering as its only element.) It looks like a probability distribution over them, but its interpretation is different. It is a *weight* function—and (normalized) weight also obeys the probability calculus. It represents how much consideration each of the sets of orderings should be given. It is regarded as a primitive function, much as the orderings themselves are primitives in the model.

One might wonder where the weights we assign to sets of preference orderings come from. Answer: they are meant to account for our judgments of value relations between options. We seek to show how these value relations can be given a consistent representation by a weight measure on the sets of preference orderings. We reverse-engineer from the value judgments that we accept to a measure that fits them as closely as possible while retaining consistency.[29] The weight of a set of orderings can be understood as the importance these orderings jointly accord to their common part—to the preferences that are common to all of them. The more weight the orderings in the set are given, the more support they accord to their common part. When we add a new ordering to the set, the common part might sometimes remain the same (the addition might not diminish it), but if it does remain the same, its weight might thereby increase (and it will never decrease).

Again, the analogy to voting is helpful. In our imagined society, the total support given to a candidate in an election is given by the sum of the weights (equal or not) of those who vote for her. If she gains a new voter, her support thereby increases.

Once we have a measure, we can immediately generalize to the infinite case, much as Kolmogorov (1950[1933]) generalizes his 'elementary theory of probability' for finitely many events to the infinite case. Like Kolmogorov, we may require our weight measure to be countably additive. Again, what matters are not the sharp numbers generated by a given measure. Other sharp numbers generated by other measures might do the job equally well. We might have many admissible measures—admissible in the sense that they model our judgments well. What matters are the structural features that they share. (Compare the usual representation of imprecise probabilities in terms of sets of precise probabilities.) It is a familiar point about representations: not all their details should be taken to stand for something real. We should not read into the fact that a map of the world colors Australia purple on a piece of paper that Australia is made of purple paper.

Can we have, in this infinite model, a finite sequence of populations with each successive population almost better than its predecessor, but with the last population being worse than the first? Yes. We start with our original example with three orderings and four populations, but now supplant each ordering *O* of four populations by a *set* of orderings on the infinite domain in which these four populations are ranked vis-à-vis each other as in *O*. Suppose that these three sets of orderings, which are mutually exclusive, are also jointly exhaustive of the class of permissible orderings. Suppose also that each of these sets has the same weight, 1/3. This gives us the desired result that the set of orderings in which each population is preferred to its predecessor has weight

2/3 (the combined weight of the two sets of orderings in which this population is ranked above its predecessor), but the first population is preferred to the last in all permissible orderings. We can increase the weight of 2/3 by increasing the number of populations in the sequence, as we have done before. So we can have more demanding interpretations of 'almost better'.

But there are limits to this exercise. 'Almost better' becomes transitive if we interpret this notion in a *maximally demanding* way: *A* is almost better than *B* iff the set of permissible orderings that rank *A* above *B* has weight 1 but still is only a proper subset of the set of all permissible orderings. This is possible in infinite models.[30] It is provable that on this interpretation, 'almost better' is transitive, just as 'better' itself is, and that 'almost better or better' is also transitive. Consequently, this interpretation brings back the original paradox: if each population in the sequence is almost better or better than the preceding one, then the last population cannot be worse than the first. The paradox can thus be avoided only if 'almost better' is given a less demanding reading.

11 | CONCLUSION

Paradoxes often reveal the perils of trying to shoehorn a messy concept into clean categories, or a fine-grained concept into coarse-grained categories. We hope to impose some seemingly sacrosanct principles, often of a logical nature, upon the concept and add some compelling premises, but the concept proves to be recalcitrant. A paradox teaches us that something has to give: at least one of the principles or premises must be abandoned, and often the concept must be refined as a result. If all goes well, philosophical progress is made. We rethink the contours of the concept, or make new distinctions within it.

This kind of progress has been made, for example, with the burgeoning of Bayesian epistemology. Traditional epistemology's coarse-grained, tri-partite distinction of *belief / disbelief / agnosticism* has served us well for many purposes. However, it has come under pressure in various ways, as brought out especially by the lottery paradox and the preface paradox. Bayesian epistemology has come to the rescue with its graded approach to belief—its distinguishing many *degrees of belief*. Belief, disbelief, and especially their remainder, the monolithic catch-all category of agnosticism, are given a more nuanced treatment. Progress has been made on the epistemological paradoxes, and fertile avenues of research in confirmation theory and decision theory have opened up.

So it has gone for epistemology, and so it should go for axiology—or so we suggest. Traditional value theory's coarse-grained, tri-partite distinction of *better / worse / equally good* has served us well for many purposes. However, it has come under pressure in various ways. The need has been recognized for a further category, *incommensurability*, but it too is a monolithic catch-all. We submit that the Continuum Argument has especially brought this out. To say that its conclusion is 'repugnant' is just to say that it presents us with a paradox—its negation is seemingly compelling, but so are the premises that lead to it, with the complicity of the transitivity of *better*, a logical principle that has mostly been regarded as sacrosanct. Parfit's solution, which appeals to incommensurability, is on the right track, but it is too coarse-grained to do justice to the intuitions that drive the paradox.

Degrees of commensurability to the rescue! If two items are commensurable, all permissible orderings agree on their relative ranking. If the agreement is not total but high, they are close to being commensurable. We have quantified degrees of commensurability by the proportion of permissible orderings that agree on the relative ranking of two items. (We have then generalized this model to an infinite-item domain and preference orderings of varying reasonableness by positing a measure on the sets of permissible orderings. In the generalized model, the proportion of

orderings having a certain feature is replaced by the measure of the set of such orderings.) We have analyzed *value* relations in terms of such proportions. Being *almost better* is one such relation. Applied to the Continuum Argument, one population is *almost better* than another if it is ranked above the latter in a large proportion of the permissible orderings—in almost all of them. It might then be mistakenly judged as simply *better*. This potential conflation of almost-betterness with outright-betterness, we have suggested, makes the Continuum Argument so seductive, despite its spuriousness. A sequence of outright-improvements interspersed with some almost-improvements might well result in a last population that is worse than the first: *almost better* is not transitive.

We think that our solution is well-motivated, twice over. Firstly, our graded approach to incommensurability is independently motivated by an examination of incommensurability more generally. Incommensurability typically arises when options need to be compared in different respects, and there are multiple admissible ways of weighing these respects. This gives rise to different permissible all-things-considered preference orderings, which may agree to varying degrees in how they rank the options under consideration. Secondly, our approach is motivated by developments elsewhere in philosophy. An analogous structure is exhibited by the strengthened version of Condorcet's Paradox. Indeed, parallel developments in Bayesian epistemology provide a welcome precedent for our approach. We need degrees of commensurability just as we need degrees of belief. We hope that this approach will also lead to fertile avenues of future research.

ENDNOTES

[1] A pivotal reference is Raz (1986); but see also Griffin (1978).

[2] This is a more demanding interpretation of "incommensurability" than the traditional one, on which it suffices for incommensurability that the values of the items under consideration cannot be represented on the same *cardinal* scale—i.e., there cannot be a common unit of measurement.

[3] Cf. also Parfit (2004). He considered a closely related, but less direct, argument for the same conclusion already in *Reasons and Persons* (1984). "The Continuum Argument" is a misnomer, as it is crucial that this argument involves a *finite* number of *discrete* steps. Nevertheless, this terminology has become somewhat entrenched since Temkin's (1996) use of it (in an argument against the transitivity of betterness that also involved a finite sequence), so we will use it too.

[4] Even if you do not agree with us about this particular sequence of comparisons, we hope you agree that there are sequences displaying this kind of structure.

[5] Here, "weighed" is used in a broad sense. The relative influence of different criteria on the all-things-considered comparisons between the items might be complex. In particular, the influence of a given criterion might not always be independent of how the compared items fare on other criteria: the criteria need not be 'separable' from each other (as they are in simple weighing).

[6] There will also typically be a range of permissible assessments of how the alternatives fare on some relevant criteria. This might lead to further all-things-considered orderings being permissible.

[7] Ewing (1947: 168) distinguishes between this "ought of fittingness" and the moral ought.

[8] Here, we disregard an even more extreme form of incommensurability, which obtains between *A* and *B* if there is no permissible way to rank them vis-à-vis each other *at all* (i.e., if all permissible orderings leave a gap between them). Rabinowicz (2008; 2012) calls this kind of relation *incomparability*. Incomparability clearly obtains between items that belong to different ontological categories—say, between persons and states of affairs. It is less clear whether it can obtain in intra-categorial comparisons.

[9] Any vagueness or context-dependence in what counts as 'almost' should govern both occurrences of the word in this sentence in the same way. We interpret 'almost all' as implying 'not all'.

[10] There is also another spectrum of degrees of commensurability on our view: degrees of approximation to the 'equally good' relation. In this sense, Bach and Einstein are less commensurable than, say, Bach and Mozart. The relation between the latter two approximates equal goodness to a larger extent.

[11] We are indebted to Timothy L. Williamson for the reference to Fishburn's (1988, 188f) strikingly simple and powerful challenge to cyclic preferences. For a discussion, see Williamson (MS).

[12] This sub-section draws on Rabinowicz (forthcoming).

[13] At some points at least, this incommensurability between P_i and P_{i+1} would have to be *persistent*: remaining however much we increased the size of P_{i+1}. Otherwise, the argument's population sequence could be repaired: points of incommensurability could be transformed into improvements by appropriate increases in population size. (See Handfield & Rabinowicz, 2018, and Rabinowicz, forthcoming.)

[14] But earlier in the paper he is less categorical: "We should claim that, of the worlds in this imagined continuum, *many* would be imprecisely equal" (ibid., 119, our emphasis).

[15] We are grateful to a referee for pointing us to an unpublished paper by Dorr, Nebel, and Zuehl (MS) that argues that all comparatives in natural language obey *Comparability*: "If x is at least as F as x and y is at least as F as y, then either x is at least as F as y or y is at least as F as x". Their paper poses an important challenge to the existence of incommensurability, and it deserves a longer response than we can provide here, but we can at least mention some relevant considerations.

    Firstly, their thesis is about *natural language* comparatives. They are well aware that various "terms of art" (39) violate Comparability; an example they consider is 'at least as strong as', meaning *entails*. We regard this comparative as a formal *explication* of natural language's 'at least as informative as', in Carnap's (1962) sense: it is similar to the explicandum, exact, simple, and fruitful. We would claim the same virtues for our formal explication of 'better'; in particular, its application to the Continuum Argument is intended to display its fruitfulness. So the fact that, on our explication, 'at least as good as' violates Comparability should not be worrying: even if it is a term of art, we need it.

    Secondly, while Dorr *et al*'s thesis may be correct about actual natural languages as a matter of empirical fact, we may easily imagine a natural language with comparatives that violate Comparability. Imagine a society in which one's social standing is exclusively determined by how high up one is on one's family tree: parents 'stand higher than' their children, and in general ancestors 'stand higher than' their descendants. Two individuals have equal standing if and only if they have the same ancestors. In this society, 'stands at least as high as' violates Comparability, yet it is not merely a term of art. It is also worth noting that it is not *vague*: there are no borderline cases of the 'is an ancestor of' relation, and it is not sorites-susceptible. This suggests that there can be incommensurability without vagueness. We will discuss vagueness in section 9.

[16] Indeed, this is one of our reasons for regarding 'better' as requiring unanimity among all the orderings, rather than merely a high proportion of them. (Thanks to Justin D'Ambrosio for helpful discussion here.) But the principal reason is that, on the fitting-attitude analysis, what is better ought to be preferred. This does not hold if there are permissible orderings in which it is not preferred.

[17] We have assumed, for simplicity, that all permissible preference orderings are linear. This makes it possible to identify the degree to which the relation between two options, A and B, approximates A being better than B with the proportion of orderings in which A is ranked above B. If *ties* between A and B were allowed, then we would need a slightly more complicated measure of approximation to betterness. (We are indebted to Krister Bykvist for raising this issue.)

    A possible solution, in the spirit of Kemeny and Snell's measure of distance between rankings (Kemeny, 1959; Kemeny & Snell, 1962), is to let ties positively contribute to the degree of approximation to betterness, but make their contribution appropriately smaller: half as large as the contribution of orderings in which A is ranked above B.

    To add another complication, what if we also allow *incomplete* preference orderings, with a gap between A and B? Should we treat gaps in the same way as ties? Or should we treat them differently? Plausibly, we might count only ties but not gaps as providing some positive contribution to the degree of approximation to betterness. But these are questions for future research.

[18] This claim will be slightly qualified when we consider infinite models in Section 10.

[19] At this point, the degrees will be *rational* numbers. We will soon generalize our model to allow irrational numbers as degrees. It also follows from our model that degrees of commensurability are themselves always commensurable—linearly ordered. (Thanks to an anonymous referee here.) This could be given up in a model with more structure. We could distinguish different dimensions in which a relation between two items may be closer to or further away from betterness. For example, if we model preferences with a cardinal structure, then we need to weigh different considerations: the proportion of orderings that rank one item higher than another,

and the relative strengths of the preferences for these items in different orderings. This weighing of considerations could be done in different ways, opening up the possibility of incommensurabilities between degrees of commensurability.

[20] Kenny G's music is often branded "elevator music" by critics.

[21] Homage to Parfit's (2004) imagining lives filled with "muzak and potatoes".

[22] We mentioned earlier that we could add cardinal structure to the permissible preference orderings. Then, one possible interpretation of the thresholds is that, according to each preference ordering, lives above the threshold that it envisages are *much* preferred to those below it. One could then reject the assumption in the Continuum Argument that the quality of lives in each successive population is *slightly worse* than in its predecessor. (Thanks here to an anonymous referee.). At each point in the sequence at which an ordering envisages a threshold, the population just after this point is *not* slightly worse than its predecessor; rather, it is *almost slightly* worse: in almost all, but not all, orderings, its predecessor is slightly preferred.

[23] Obviously, if some permissible orderings do not exhibit such thresholds and consequently place the last population at least as highly as the first, they might still be in a small minority. Then the last population will be *almost worse* than the first one, and we will still avoid the Repugnant Conclusion. But the last population can be *worse* than the first only if all permissible orderings involve thresholds. It might be noted, by the way, that the presence of thresholds located at different places in the orderings prevents the class of permissible orderings from being 'single-peaked'. (A set of rankings is single-peaked over a set of items if the items can be ordered along a line in such a way that in each ranking the items are placed lower the further away they are on the line from this ranking's top item.) A well-known result in social choice theory (Black, 1948) is that majority rule yields a transitive social ranking if the underlying preferences are single-peaked. Translated to our model, this means that if the class of permissible orderings had been single-peaked, *almost better* would have been a transitive relation. (We are indebted here to Brian Hedden.)

[24] Various authors have appealed to incommensurability to avoid the Repugnant Conclusion, and some of them appeal to multiple alternative thresholds, as we do. See the "incomplete critical level utilitarianism" (also referred to as "critical-band utilitarianism") of Blackorby, Bossert, & Donaldson (1996), and other views with similar formal structure: Rabinowicz (2009b), Qizilbash (2007; 2018), Gustafsson (2019). See also the "lexical-threshold totalism" of Nebel (forthcoming). Both these kinds of views entail what we suggest about the Continuum Argument: some but not necessarily all of the populations are incommensurable with their immediate predecessors. (Thanks here to an anonymous referee for pressing this point.)

   We take our distinctive contribution to be the modelling of the fine structure of incommensurability as a spectrum of fine-grained relations, and the consequent non-transitivity of 'almost better'. (All the other 'non-unanimous' relations in our modelling can be shown to be non-transitive too, leaving us with only *better*, *worse*, and *equally good* as the transitive relations—but we appeal only to the non-transitivity of 'almost better' here.)

[25] Note, though, that for this jury interpretation of our model, we cannot have several jurors proposing the same ranking (unless some rankings should be given more weight than others; see the next section).

[26] Or, if the number of admissible precisifications is infinitely large, the degree of determinacy might be defined as the measure of the set of precisifications on which the statement in question holds. If vague statements are considered to be neither (fully) true nor (fully) false, as is sometimes done, then degrees of determinacy are interpretable as degrees of truth. See, for example, Edgington (1992; 1996); McGee and McLaughlin (1995: section V). Early proposals along these lines were put forward in Lewis (1970) and Kamp (1975). For a critical discussion, see Smith (2008): 188ff.

[27] Broome (1997; 2004: §12.2; 2009) argues that vagueness in a value ordering, which we need to allow for anyway, would crowd out incommensurability (or at least determinate incommensurability) from the ordering in question. But the key premise of his argument—the so-called "Collapsing Principle"—has been criticized by several authors (e.g., Carlson, 2004; 2013; Elson, 2014; Gustafsson, 2018).

[28] Parfit (2016: 116) concurs. He allows both cases of determinate imprecise equality and cases in which "it would be indeterminate whether one of two things would be better, or these things would be imprecisely equally good".

[29] By analogy, regarding the standard Stalnaker/Lewis semantics for counterfactuals, one might wonder where the similarity relation on worlds comes from. Lewis (1979) reverse-engineers from our judgments of truth values of counterfactuals to a similarity relation that fits them. That semantics has been thought to illuminate the logic of counterfactuals even without specifying the similarity relation. We likewise think that our model illuminates

the logic of incommensurability—notably, the non-transitivity of *almost better*—even without specifying the measure over permissible orderings.

   The supervaluationist approach discussed in the previous section could similarly give unequal weights to the admissible precisifications. One similarly might wonder where the weights come from. The supervaluationist could give a parallel answer to ours.

[30] Compare how in infinite models, probability 1 does not entail necessity. For example, the probability that a continuous random variable fails to take a particular value is 1, though its taking that value is possible.

REFERENCES

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

Black, D. (1948). On the rationale of group decision-making. *Journal of Political Economy 56*, 23–34. https://doi.org/10.1086/256633

Blackorby, C., Bossert, W., & Donaldson, D. (1996). Quasi-orderings and population ethics. *Social Choice and Welfare 13*, 129–150. https://doi.org/10.1007/bf00183348

Brentano, F. (1969). *The Origin of Our Knowledge of Right and Wrong*. (R.M. Chisholm, Trans.). Routledge & Kegan Paul. (Original work published 1889)

Broome, J. (1997). Is incommensurability vagueness? In R. Chang (Ed.), *Incommensurability, Incomparability, and Practical Reason* (pp. 67–89). Cambridge MA: Harvard University Press.

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

Broome, J. (2009). Reply to Rabinowicz. *Philosophical Issues*, *19*, 412–417. https://doi.org/10.1111/j.1533-6077.2009.00175.x

Buchak, L. (2013). *Risk and Rationality*. Oxford: Oxford University Press.

Carlson, E. (2004). Broome's argument against value incomparability. *Utilitas*, *16*, 220–224. https://doi.org/10.1017/s0953820804000548

Carlson, E. (2013). Vagueness, incomparability, and the collapsing principle. *Ethical Theory and Moral Practice*, *16*, 449–463. https://doi.org/10.1007/s10677-012-9352-9

Carnap, R. (1962). *Logical Foundations of Probability* (2nd ed.). Chicago: The University of Chicago Press.

Dorr, C., Nebel, J. M., & Zuehl, J. (MS). The case for comparability.

Dougherty, T. (2014). A deluxe money pump. *Thought: A Journal of Philosophy*, *3*, 21–29. https://doi.org/10.1002/tht3.91

Dreier, J. (2019). World-centered value. In C. Seidel (Ed.), *Consequentialism: New Directions and Problems* (pp. 31–50). Oxford: Oxford University Press.

Edgington, D. (1992). Validity, uncertainty and vagueness. *Analysis*, *52*, 193–204. https://doi.org/10.1093/analys/52.4.193

Edgington, D. (1996). Vagueness by degrees. In R. Keefe and P. Smith (Eds.), *Vagueness: A Reader* (pp. 294–316). Cambridge MA: MIT Press.

Elson, L. (2014). Borderline cases and the collapsing principle. *Utilitas*, *26*, 51–60. https://doi.org/10.1017/s095382081300023x

Ewing, A. C. (1947). *The Definition of Good*. London: MacMillan.

Fishburn, P. (1988). *Nonlinear Preference and Utility Theory*. Baltimore: The Johns Hopkins University Press.

Griffin, J. (1978). *Are there incommensurable values?* Philosophy and Public Affairs, *7*, 39–59.

Gustafsson, J. E. (2010). A money pump for acyclic intransitive preferences. *Dialectica*, *64*, 251–257. https://doi.org/10.1111/j.1746-8361.2010.01230.x

Gustafsson, J. E. (2013). The irrelevance of the diachronic money-pump argument for acyclicity. *The Journal of Philosophy*, *110*, 460–464. https://doi.org/10.5840/jphil2013110819

Gustafsson, J. E. (2016). Money pumps, incompleteness, and indeterminacy. *Philosophy and Phenomenological Research*, *92*, 60–72. https://doi.org/10.1111/phpr.12248

Gustafsson, J. E. (2018). Does the collapsing principle rule out borderline cases? *Utilitas*, *30*, 483–492. https://doi.org/10.1017/s0953820818000079

Gustafsson, J. E. (2019). Population axiology and the possibility of a fourth category of absolute value. *Economics and Philosophy*, *36*, 81–110. https://doi.org/10.1017/s0266267119000087

Gustafsson, J. E. & Rabinowicz, W (2020). A simpler, more compelling money pump. *The Journal of Philosophy*, *117*, 578–89. https://doi.org/10.5840/jphil20201171037

Hammond, P. (1988). Consequentialist foundations for expected utility. *Theory and Decision*, *25*, 25–78. https://doi.org/10.1007/bf00129168

Handfield, T. (2014). Rational choice and the transitivity of betterness. *Philosophy and Phenomenological Research*, *89*, 584–604. https://doi.org/10.2139/ssrn.2331864

Handfield, T. (2016). Essentially comparative value does not threaten transitivity. *Thought: A Journal of Philosophy*, *5*, 3–12. https://doi.org/10.1002/tht3.188

Handfield, T. & Rabinowicz, W. (2018). Incommensurability and vagueness in spectrum arguments: options for saving transitivity of betterness. *Philosophical Studies*, *175*, 2373–2387. https://doi.org/10.1007/s11098-017-0963-9

Huemer, M. (2008). In defence of repugnance. *Mind*, *117*, 899–933. https://doi.org/10.1093/mind/fzn079

Kamp, H. (1975). Two theories about adjectives. In E. L. Keenan (Ed.), *Formal Semantics of Natural Language* (pp. 123–155). Cambridge: Cambridge University Press.

Katz, L. (2011). *Why the Law is So Perverse*. Chicago: The University of Chicago Press.

Kemeny, J. G. (1959). Mathematics without numbers. *Daedalus*, *88*, 577–591.

Kemeny, J. G. & Snell, J. L. (1962). Preference rankings. In J. G. Kemeny and J. L. Snell, *Mathematical Models in the Social Sciences* (pp. 9–23). Blaisdell.

Kolmogorov, A. N. (1950). *Foundations of the Theory of Probability*. (N. Morrison, Trans.) New York: Chelsea Publishing Company. (Original text published in German 1933.)

Lewis, D. (1970). General semantics. *Synthese*, *22*, 18–67. https://doi.org/10.1007/bf00413598

Lewis, D. (1979). Counterfactual dependence and time's arrow. *Noûs*, *13*, 455–476. https://doi.org/10.2307/2215339

McGee, V. & McLaughlin, B. (1995). Distinctions without a difference. *Southern Journal of Philosophy*, *33* (suppl. vol.), 203–251. https://doi.org/10.1111/j.2041-6962.1995.tb00771.x

Nebel, J. M. (2018). The good, the bad, and the transitivity of better than. *Noûs*, *52*, 874–899. https://doi.org/10.1111/nous.12198

Nebel, J. M. (forthcoming). Totalism without repugnance. In J. McMahan, T. Campbell, J. Goodrich & K. Ramakrishnan (Eds.), *Ethics and Existence: The Legacy of Derek Parfit*. Oxford: Oxford University Press.

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

Parfit, D. (2004). Overpopulation and the quality of life. In T. Tännsjö & J. Ryberg (Eds.), The Repugnant Conclusion*: Essays on Population Ethics* (pp. 7–22). Dordrecht: Kluwer Academic Publishers.

Parfit, D. (2016). Can we avoid the repugnant conclusion? *Theoria*, *82*, 110–27. https://doi.org/10.1111/theo.12097

Qizilbash, M. (2007). The mere addition paradox, parity, and vagueness. *Philosophy and Phenomenological Research*, *75*, 129–151. https://doi.org/10.1111/j.1933-1592.2007.00063.x

Qizilbash, M. (2012). Incommensurability or vagueness? A comment on Rabinowicz and Sugden. *Proceedings of the Aristotelian Society*, *112*, 333–338. https://doi.org/10.1111/j.1467-9264.2012.00338.x

Qizilbash, M. (2018). On parity and the intuition of neutrality. *Economics and Philosophy*, *34*, 87–108. https://doi.org/10.1017/s0266267117000281

Quinn, W. (1990). The puzzle of the self-torturer. *Philosophical Studies*, *59*, 79–90. https://doi.org/10.1007/bf00368392

Rabinowicz, W. (2000). Money pump with foresight. In M. J. Almeida (Ed.), *Imperceptible Harms and Benefits* (pp. 123–154). Dordrecht: Kluwer Academic Publishers.

Rabinowicz, W. (2008). Value relations. *Theoria*, *74*, 18–49. https://doi.org/10.1111/j.1755-2567.2008.00008.x

Rabinowicz, W. (2009a). Incommensurability and vagueness. *Proceedings of the Aristotelian Society*, Supplementary Volume, *83*, 71–94. https://doi.org/10.1111/j.1467-8349.2009.00173.x

Rabinowicz, W. (2009b). Broome and the intuition of neutrality. *Philosophical Issues*, *19*, 389–411. https://doi.org/10.1111/j.1533-6077.2009.00174.x

Rabinowicz, W. (2012). Value relations revisited. *Economics and Philosophy*, *28*, 133–164. https://doi.org/10.1017/S0266267112000144

Rabinowicz, W. (forthcoming). Can Parfit's appeal to incommensurabilities block the continuum argument for the repugnant conclusion? In J. McMahan, T. Campbell, J. Goodrich and K. Ramakrishnan (Eds.), *Ethics and Existence: The Legacy of Derek Parfit*. Oxford: Oxford University Press.

Rachels, S. (1998). Counterexamples to the transitivity of better than. *Australasian Journal of Philosophy*, *76*, 71–83. https://doi.org/10.1080/00048409812348201

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

Scanlon, T. M. (1998). *What We Owe to Each Other*. Cambridge MA: Harvard University Press.

Smith, N. J. J. (2008). *Vagueness and Degrees of Truth*. Oxford: Oxford University Press.

Sugden, R. (2009). On modelling vagueness—and on not modelling incommensurability. *Aristotelian Society Supplementary Volume*, *83*: 95–113. https://doi.org/10.1111/j.1467-8349.2009.00174.x

Temkin, L. (1986). Intransitivity and the mere addition paradox. *Philosophy & Public Affairs*, *16*, 138–210.

Temkin, L. (1996). A continuum argument for intransitivity. *Philosophy & Public Affairs*, *25*, 175–87. https://doi.org/10.1111/j.1088-4963.1996.tb00039.x

Temkin, L. (1997). Rethinking the good, moral ideals and the nature of practical reasoning. In J. Dancy (Ed.), *Reading Parfit* (pp. 290–345). Oxford: Blackwell.

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

Tullock, G. (1964). The irrationality of intransitivity. *Oxford Economic Papers*, *16*, 401–406.

Voorhoeve, A. & Binmore, K (2006). Transitivity, the sorites paradox, and similarity-based decision-making. *Erkenntnis*, *64*, 101–114. https://doi.org/10.1007/s10670-005-2373-1

Williamson, T. L. (MS). A risky challenge for intransitive preference.