Chapter 7: Spectrum Arguments

Luke Elson

June 9, 2025

If it's better that I go out on Thursday than on Friday, and it's better that I go out on Friday than on Saturday, then it's better that I go out on Thursday than on Saturday.

This inference seems absolutely undeniable when thinking about 'all things considered' impartial goodness (which, admittedly, isn't always foremost in the mind when choosing when to go out with friends). I think it obviously correct, following from an obviously correct general principle:

Transitivity of Betterness. For any outcomes X, Y, and Z, if Y is better than X, and Z is better than Y, then Z is better than X.

But *Spectrum Arguments* have been used to argue that such betterness is not transitive, most famously by Stuart Rachels and by Larry Temkin.[1]

This chapter marks a change of subject from preferences ('prefers') to value ('is better'). But the two subjects are clearly linked. I've defended the transitivity of preferences as a rational norm against several challenges, and intransitive betterness would be another threat to it. I think it obvious that we can seek to maximise the impartial good (to strictly prefer X to Y *iff* X is impartially better than Y), and that it could be *rational* to seek such maximisation. But if betterness is intransitive, then seeking to improve the world could give us intransitive preferences, a surprising and undesirable result.

As we've seen, Chrisoula Andreou defends cyclic preferences. In particular, she accepts that there can be cycles in 'is morally preferable to' and tries to reconcile this with a transitive 'is morally better than' (Andreou 2023, 130–31). The two must diverge, which makes for an interesting epicycle but like most such, it would be better if we could avoid it.

That's not even to mention the *buck-passers*, who think that value is reducible to fitting preferences: for going out on Thursday to be better than going out on Friday is for there to be (overall) more reason to prefer going out on Thursday over Friday, or for such a preference to be fitting. For them, the distinction between betterness and (fitting) preference is merely verbal. I discuss a fitting-attitude

[1] The most developed and widely-discussed spectra are in Rachels (1998) and Temkin (2012); see references therein to their earlier work.

account of the Spectrum Argument below.[2]

If we hope to rule out intransitivity in rational preferences, we can't rest easy whilst there appears to be intransitivity in betterness. Much of my discussion below is an evaluative parallel to that in Chapter 5; I apologise for any repetition, but a relatively self-contained discussion of Spectrum Arguments is worth providing. Spectrum Arguments have been a niggling moral problem since at least Parfit's Repugnant Conclusion, which says that given any world with a happy population, that world would be better with some (larger, possibly vastly larger) population whose lives are just barely worth living.[3]

I am at least tempted by Global Normative Nihilism, so I am at least tempted to think that none of the objective evaluative claims below is true, but I'll aim to show that vagueness can provide a resolution to spectrum arguments even within the broken framework that is moral realism. Other error theorists and allies have taken a different tack: the arguments highlight a real inconsistency in the morality system, one inconsistent with moral realism.[4] At the risk of overreach, I aim to dissolve that problem in this chapter.

1 Depression Spectrum

The spectra in Spectrum Arguments are complex cases, often involving simultaneous variation in the number of people experiencing some harm or benefit, the duration of that harm or benefit, and its intensity. The complexity of the cases engenders a sense that there *must* be some kind of trickery in them. And as we'll see, that trickery looks like vagueness. But they are powerful arguments.

I will focus on a spectrum due to Larry Temkin, because (like Stuart Rachels) he explicitly rejects transitivity. Or rather, he argues that transitivity is inconsistent with with two other near-undeniable principles, so one of them has to go, and he is 'prepared, at least in principle, to reject the transitivity of "better than"' (Temkin 2012, 65).

Temkin comprehensively rejects attempts to construe—or debunk—his spectra as sorites. Focusing on just one example or author is risky, because it might not be representative of all spectra. But even if so, we can draw conclusions about all spectra *of that kind*. And focusing on the details of that kind allow us to show precisely how vagueness models the spectrum in question.

Depression is harmful. Table 1 shows our main example—taken verbatim from (Temkin 2012, 47ff) with an adjustment for Anglo/American differences in the meaning of 'quite'—which includes a range of depressive harms, from feeling pretty down one day a month to being extremely seriously depressed six days a week.

[2] See Nebel (2018) pp. 883-884 for a similar point about the connection between goodness and fitting preference, or reasons to prefer.

[3] Adapted from Parfit (1984), section 131. Unsurprisingly given the name, Parfit rejects this conclusion.

[4] See especially Cowie (2022) and Cowie (2023).

Table 1: Depression Spectrum.

Outcome	Harm (how each victim feels)	Number of victims
O1	pretty down once a month	100,000
O2	awfully down in the dumps once every two weeks	20,000
O3	slightly depressed 1 day a week	6,000
O4	pretty depressed 2 days a week	2,000
O5	fairly seriously depressed 3 days a week	800
O6	seriously depressed 4 days a week	200
O7	very seriously depressed 5 days a week	50
O8	extremely seriously depressed 6 days a week	10

Each outcome O1–O8 specifies a harm level (an intensity and duration of depression) and a number of victims. The outcomes are akin to stages in Chapter 5's Self-Torturer, being a bundle of suffering (electrical pain, depression) and reward (money, fewer victims). We can assume that all else is equal for the rest of the human population. I'll say that the victims suffer at 'level n' in outcome On.

The spectrum argument has just three premises. The first is a substantive evaluative claim about outcomes O1–O8. It's an instance of what Temkin calls the 'First Standard View', that sometimes it is better for a larger number of people to have a lower-quality benefit than for a smaller number of people to have a higher-quality benefit.[5] This will be so if the difference in the number of people is sufficiently big, and the difference in the quality of the benefits is not too big. It's better for a million people to receive £45 than for a thousand people to receive £50.

Temkin claims that the First Standard View has the following upshot in the spectrum:

Pairwise-Later-Better. In each adjacent comparison in Depression Spectrum, the latter is better—O2 is better than O1, and so on.

Pairwise-Later-Better is hard to deny. For example, though the depression in O3 is (speaking simply) half as bad as that in O4, it affects *three times* as many victims. Here it seems clear that O4 is an improvement, that the benefit (relief from depression) to 4,000 outweighs the harm (a worse level of depression) to 20,000.

The second premise is another substantive evaluative claim. Temkin's 'Second Standard View' says that it can be better for a small number of people to receive

[5] Adapted from Temkin (2012), p. 30.

a higher-quality benefit than for a huge number of people to receive a lower-quality benefit.[6] This can happen if the difference in the quality of the benefits is sufficiently big: it's better for one person to receive £1m than for ten million people to receive 10p, for example.

Here is its application to the current spectrum:

Distant-Earlier-Better. In Depression Spectrum, O1 is better than O8.

This claim is also very hard to deny. As Temkin points out, serious depression can be debilitating.[7] The ten victims in O8 have lost much that makes life worth living. Meanwhile, feeling pretty down once a month as in O1 looks like a baseline for normal human existence, to the point where it's not even clear that avoiding this would *improve* our lives. It's not totally undeniable that O8 is worse than O1, but it's as close to undeniable as we are likely to get.

The third premise is an instance of transitivity:

Depression Transitivity. In Depression Spectrum, if O2 is better than O1, and O3 is better than O2, and ..., and O8 is better than O7, then O8 is better than O1.

The three premises are jointly inconsistent. Repeated applications of Pairwise-Later-Better together with Depression Transitivity imply that O1 is worse than O8, which is of course the denial of Distant-Earlier-Better. One of them has to go. He thinks that conflict between the short-range verdicts of Pairwise-Later-Better and the long-range verdicts of Distant-Earlier-Better engenders (or reflects) genuine intransitivity, leading to the three principles being inconsistent.

2 Defending Transitivity

Let's agree that one outcome can't be both better and worse than another outcome, which would be even more revisionary than intransitivity. Then the three main premises *are* inconsistent, so to save transitivity we must reject one of the two substantive claims.

It need not be the same one in every spectrum argument. Sometimes the analogue of Distant-Earlier-Better is false. For example, (Temkin 2012, 51) claims that 400,000 people living without the top third of one finger is better than 10 people living without two arms and a leg; I disagree.

But often the analogue of Distant-Earlier-Better is extremely plausible. Most of us reject simple impartial aggregation of welfare—most of us sometimes judge it worse for a few people to suffer terribly than for many people to suffer slightly, even if the total suffering is greater in the latter case—and so it's hard to deny Distant-Earlier-Better in all cases.

That leaves us Pairwise-Later-Better. Depression Spectrum seems as good a case as any to attack it, but Pairwise-Later-Better is compelling here. Perhaps predictably, I'll argue that it's a sorites tolerance principle, and that this explains

[6] Adapted from Temkin (2012), p. 32.

[7] Temkin (2012), p. 49.

why it's both false and compelling. Many have suggested that the spectrum argument is an instance of the sorites paradox.[8] Others have argued that the spectra rest on incommensurability (incomparability), and since I think *that* phenomenon is vagueness, if they are right then I think the spectrum is a sorites (using that term broadly to mean 'a fallacious argument that trades on vagueness').[9]

But for a sorites you need vagueness; moreover, you need a convincing explanation of *where* the vagueness is. We faced a similar challenge when arguing for vague preferences in Chapter 2.

Before that hard work begins, here's some circumstantial evidence. In a Spectrum Argument, many individually-plausible steps lead to an implausible conclusion: many instances of Pairwise-Later-Better, together with Depression Transitivity, imply that O8 is better than O1. This structure is characteristic of the sorites paradox, so its presence here is suggestive of vagueness. But merely suggestive, because differing verdicts depending on whether we reach a destination (such as a comparison between O1 and O8) by few or many steps is also—by definition—characteristic of intransitivity.

The phenomenology smells of vagueness. If you'll forgive the self-quoting, here's what I wrote about Pairwise-Later-Better above:

> Pairwise-Later-Better is hard to deny. For example, though the depression in O3 is (speaking simply) half as bad as that in O4, it affects *three times* as many victims.

But as I wrote that, I hesitated. Is it really obvious that that O3 is worse than O4? Granted O4 has has far fewer sufferers, but it also has much greater intensity of suffering. The pairwise comparison is awfully lacking in detail—especially phenomenological detail about what the levels of suffering are like—and perhaps its plausibility could be explained away. Certainly the pairwise comparisons look like the soft underbelly of Depression Spectrum.

Temkin has a response to such qualms:

> [you could be] be uncertain about the desirability of the different trade-offs given the particular numbers I've chosen ... [one might increase the numbers suffering the worse depression] ... alternatively, one could always insert an intermediate illness between the two described and then rerun the argument with an extra step[10]

If, for example, you are unsure about the comparison of O3 and O4, adding an intermediate outcome will help, as in Table 2.

[8] See, for example, Qizilbash (2005), Nebel (2018) and Thomas (2021).

[9] Chang (2016), pp. 205–211; Parfit (2016). See also Elson (2017).

[10] Temkin (2012), p. 49.

Table 2: Adding an intermediate stage to Depression Spectrum.

Outcome	Harm (how each victim feels)	Number of victims
O3	slightly depressed 1 day a week	6,000
O3.5	a bit more than slightly depressed 1.5 days a week	3,000
O4	pretty depressed 2 days a week	2,000

Temkin claims, and it seems correct, that adding the extra step between O3 and O4 makes the pairwise betterness-verdicts more compelling. But in light of Chapter 5, this really should ring alarm bells. The procedure described by Temkin is an analogue of the *filtered series* we saw in the Puzzle of the Self-Torturer. More precisely, the procedure is a refinement or anti-filtering of Depression Spectrum, adding outcomes instead of taking them away.

Like Quinn, Temkin sees that we can create the appearance of intransitivity by introducing intermediate steps, or remove that appearance by removing them. I showed in Chapter 5 how vagueness neatly explains the effectiveness of filtered series in the Self-Torturer—if you increase the number of steps in a sorites, each step becomes more plausible—so we should be confident that the same maneouvre will work here too. But genuine intransitivity could also offer an explanation for this phenomenon, since that too manifests with many 'small steps'. The evidence is again circumstantial.

3 A Soritical Model

As I mentioned, others have argued that the spectrum could be a sorites. The central idea behind sorites models is that some instance of Pairwise-Later-Better is false, but it's vague which one. The existing model in the literature unfortunately fails for many spectra, I'll argue—including Depression Spectrum.[11] I'll present that model, explain the problem with it, and then fix it.

Consider a principle such as

Depression-Threshold. There is some threshold level of depression so debilitating that there is some number n (e.g., 10) such that n people suffering at the threshold level is worse than *any* number of people suffering any milder depression.

It's better that any number of people suffer less than the threshold level than that ten people suffer at or worse than the threshold. Depression-Threshold explains why O8 is worse than O1: there are far more victims in O1, but the suffering in

[11] The clearest version of the model is given by Thomas (2021).

O8 exceeds the threshold whereas that in O1 does not. So Distant-Earlier-Better is true.

Depression-Threshold would follow from what I'll call the

Numbers-Resistant Evaluative Principle. Fewer people suffering from depression is pro-tanto good; people suffering less badly from depression is pro-tanto good; some levels of depression are so harmful that *any* outcome in which n people suffer at those levels is worse than *any* outcome in which fewer than n people suffer at those levels.

In other words, some levels of depression are lexically inferior to lesser levels of depression. In what follows, I'll assume that n is 1 for readability: the threshold level of depression is so bad that even a *single* person suffering it is worse than any number of people suffering less intensely.

Imagine we can choose an outcome O1–O8 to instantiate. We can decrease the intensity of the suffering (moving up Table 1) only if we also ratchet up the number of sufferers, and we can decrease the number of sufferers (moving down Table 1) only by making things worse for the remaining sufferers. The soritical model says that as we make adjacent pairwise comparisons (O1 to O2, then O2 to O3, ...) we reach the best outcome by getting as close as possible to the threshold *without* crossing it. That is the point where as few people as possible suffer from depression without any of them suffering unacceptably (lexically) badly. Their suffering may still be terrible, but not enough to automatically outweigh far larger numbers.

At some point, the remaining sufferers are taken beyond this lexical threshold level of harm—perhaps their lives are wasted, not worth living—and that outcome is worse than *all* previous outcomes. (This may remind you of Chapter 5's shepherd, where instead of a collapse there is a surge in utility when he's moved enough stones for a cairn.)

On this model, Pairwise-Later-Better has a false instance where the threshold is crossed, and so its universally-quantified form is false. It may be that some instances of Pairwise-Later-Better are clearly, determinately true, but the model needs just one false instance—false by a clear enough margin to make O8 worse than O1. Lexical inferiority certainly achieves that, and is how the current strategy defends transitivity against Pairwise-Later-Better.

The job of vagueness in this model is to is explain away the plausibility of Pairwise-Later-Better. That principle is already the weak point of Depression Spectrum. If Numbers-Resistant Evaluative Principle is true, then there's a threshold level of depression and it corresponds to what I'll stipulatively call a *wasted life*. Any outcome in which someone's life is wasted is worse than an outcome in which nobody's life is wasted. But precisely how much depressive suffering is required for a life to be wasted? Even without vagueness, this will be an enormously difficult first-order evaluative question, with a lot of interpersonal variation in the answer.

In the spectrum *most* instances of Pairwise-Later-Better are true, because in most adjacent pairs of outcomes, the later outcome has a lower number of sufferers and the substantive threshold isn't crossed; to identify the false instance we need to find the pair of adjacent outcomes that straddles the threshold, which could be very difficult. We can see why each instance of Pairwise-Later-Better would look very plausible—especially given the vague descriptions of the outcomes in the spectrum—but still hesitate about it as a fully-general claim.

As the series is refined, there are more adjacent pairwise comparisons, but still only one of them is false—only one straddles the threshold—so our credence in most instances of Pairwise-Later-Better should increase. ('Most' because perhaps some comparisons near the start or end of the spectrum were always obviously true.) That is how the spectrum could be explained as uncertainty or ignorance, and hence perhaps as vagueness with epistemicism.

But what about indeterminacy? Rather than it being merely unknown which of the transitions (say) O4–O5, O5–O6 or O6–O7 wastes a life, it instead could be indeterminate precisely how much depression is needed to cross that substantive evaluative threshold.

There will then be three (classes of) sharpenings:

- Sharpening s_5 says that level 5 is the first life-wastingly bad level, so **O5 is worse than O4**, but O6 is better than O5 and O7 is better than O6.
- Sharpening s_6 says that level 6 is the first life-wastingly bad level, so O5 is better than O4, **O6 is worse than O5**, and O7 is better than O6.
- Sharpening s_7 says that level 7 is the first life-wastingly bad level, so O5 is better than O4 and O6 is better than O5, and **O7 is worse than O6**.

The underlying tolerance principle is 'if the suffering in On doesn't exceed the threshold level (of wasting a life), then the suffering in O(n+1) doesn't exceed the threshold level', and has fundamentally the same structure as 'if n grains of sand are a heap, then (n-1) grains of sand are a heap'. Given the Numbers-Resistant Evaluative Principle, the tolerance principle implies that each outcome is better than the previous one, which is Pairwise-Later-Better. But whether we construe it as uncertainty, ignorance, or indeterminacy, assuming a tolerance-denying account of vagueness as I have throughout this book, it's determinate that *one* of the transitions crosses the threshold. Tolerance principles about heaps and wasted lives are hard to deny, but must be false.

Since the tolerance principle has a false instance, if we are trying to make the world as good as possible by moving down the table, then we are caught in the kind of (single-threshold) practical sorites discussed in Chapter 5. The precise project is reducing the number of sufferers (reducing the number of sufferers *all else equal* always makes the world better); the vague project is keeping the suffering levels above the threshold level. The outcome is much worse if even one sufferer goes over that threshold and her life is wasted—and its location is indeterminate. The different projects in the practical sorites correspond to different clauses in Numbers-Resistant Evaluative Principle.

We can even reconcile the falsity of Pairwise-Later-Better with the First Standard View. That View says that an outcome is better '[...] and *if* the differences in the initial situations of the people benefited and the degrees to which they are benefited are not "too" great'.[12] If a phenomenally small difference can be *too* great because it crosses some substantive threshold (for example, wasting a life), then Pairwise-Later-Better can be false without taking the First Standard View down with it.

Teruji Thomas argues convincingly that many spectrum arguments are susceptible to this kind of sorites story, giving an example in terms of 'ecstatic' pleasures—the ecstatic/non-ecstatic boundary is the threshold for his version of Pairwise-Later-Better. On Thomas's model, just as the correct theory of vagueness explains why heap-tolerance principles are so plausible, it explains why Pairwise-Later-Better is so plausible, even though they are both false.[13]

On this picture, a spectrum argument is not fundamentally different to the challenge that faces Chapter 5's (solvent) shepherd or Chay. It's also what my previous, failed model of the self-torturer looked like. Insofar as you can vary the number of depression sufferers and the intensity of their suffering in the way here described, and you want to maximise the good, Numbers-Resistant Evaluative Principle puts you in a single-threshold practical sorites.

4 Problems with Numbers-Resistance

That model has an advantage. It is strong: the appearance of intransitivity is dissolved even when the number of people in O1 is increased *arbitrarily* highly. Temkin and his allies can't simply rescue the Spectrum Argument by increasing the number of people in O1 into the millions or beyond. One person suffering more than the threshold amount is worse than *any* number of people suffering at level O1. This is numbers-resistance.

But it also has a disadvantage. It is strong: considered purely as an evaluative claim, numbers-resistance is far from obvious. Could there really be a level of depression so bad that one person (or ten thousand people, or...) suffering it outweighs the badness of *any* number of people suffering mild depression, even say ten trillion? Severe depression is one of the worst things that can happen to a person. But when we are talking about impartial goodness, 'from the point of view of the universe' as they say, I find such numbers-resistance hard to believe. Depression Spectrum wouldn't be half as compelling if there were ten trillion sufferers in O1.

Your verdict may differ (if you seek to reject the Repugnant Conclusion, then it likely does) and need not be the same in every spectrum argument.[14] But if the price for using vagueness to save transitivity is such numbers-resistance, then in many cases—including Depression Spectrum—that price may be too high for some. We'd rather not be *forced* into it.

Numbers-resistance also has surprising consequences near the threshold. As Handfield and Rabinowicz show, this kind of model implies that 'there are adjacent, qualitatively very similar harm types in the spectrum such that one is

[12] Temkin (2012), p. 30. Emphasis in original.

[13] Thomas (2021), p. 5.

[14] See Huemer (2008) for a thorough defence of the Repugnant Conclusion.

radically inferior to the other'.[15] The threshold is a cliff-edge. For the solution just described to work, it must be better for ten trillion people to suffer a level of depression just below the threshold than for one person (or whatever n is) to suffer a level just above it. Better that ten trillion people almost waste their lives than that one person wastes her life. Handfield and Rabinowicz show that any numbers-resistant solution to the spectrum must imply such a cliff around the threshold level

The idea behind this 'cliff-edge problem' is clear: if one harm far above some lexical threshold is thereby radically superior to another harm far below it, and this is explained by their standing on opposing sides of the threshold, then (at least if we are assuming no other oddities) there must also be such radical superiority where both harms are arbitrarily close to the threshold. The point looks at first to be devastating for vagueness models of the spectrum argument, because as (Handfield and Rabinowicz 2018, 2385) put it, such a cliff-edge 'is still counterintuitive, whether or not it is indeterminate where this point occurs'.

If we are asked to choose which is more implausible—intransitivity or radical inferiority between similar levels of harm—the answer is far from clear. I'll show below that we can model the spectrum as vagueness *without* radical inferiority but with the cost, if it is a cost, of accepting the analogue of the Repugnant Conclusion.

But first, a word in defence of the cliff-edge. Lexical inferiority *is* sometimes plausible. Returning to preferences, it's plausible that there's some amount of pain in the Self-Torturer's ordeal that isn't compensated by *any* amount of money. Despite the very large amounts of money involved, the Self-Torturer will reach a point where the pain isn't worth enduring for any riches.

As we've just seen, there therefore must be two adjacent settings, one which can be compensated by some (extremely large) amount of money, and one which can't. But *pace* Handfield and Rabinowicz, much of the implausibility here *is* mitigated by the location of the adjacent settings being vague. If you are at a pain level near this cliff, you likely have intense lifelong pain (how much will vary enormously between individuals, of course) that limits your enjoyment of life. But you have enough money to experience fantastic adventures, help those you love, pursue the projects that give your life meaning, and so on.

The persistent pain means that you don't get that much joy from visiting Angel Falls, but you prefer this to no pain but also no trip to Venezuela. The pain is also psychological: you know that you put yourself in this situation—it's not as if you have an involuntary chronic condition—for a mercenary reason, and you fear that you are wasting your life, and that your friends and family are just hangers-on, using you for your money. Is it worth it? The calculation is on a knife-edge, so it wouldn't take much more pain for you to cross the threshold—*no amount of money is worth this!*—to a point where the persistent physical and psychological pain means that you'd would prefer instead a poor but painless life with your family and friends.

The point is that this is a very fine-grained complex choice, and it's not hard to

[15] Handfield and Rabinowicz (2018), p. 2385. Chang (2016), p. 2017 also puts this point clearly.

see how the three kinds of indeterminacy I described in Chapter 2 could arise. Indeterminacy makes the cliff-edge more palatable: settings that can *determinately* be compensated by some amount of money are likely quite distant from settings that are determinately not worth any money. There is no sudden switch in what you consistently counterfactually choose. So in the Self-Torturer radical inferiority is not excessively implausible. (The repeating model I defended in Chapter 5 doesn't require it, but I think it plausibly occurs nevertheless.)

However, Depression Spectrum involves not preferences but impartial goodness, not one person but many. So here my appeal to counterfactual choices (as in the Self-Torturer) looks less convincing. Now perhaps evaluative radical inferiority is less implausible with truly large numbers: perhaps there is a *hellish* level of suffering where a countable infinity of people suffering at the hellish level is worse than any bigger infinity of people suffering at a very slightly less-than-hellish level. But there's no denying that the evaluative cliff-edge problem is—at least—surprising.

Again, vagueness in the cliff's location does at least mitigate this surprise. It *would* be implausible if we had two populations on either side of the cliff-edge, where a 'Slightly Better-Off Population' is barely distinguishable in welfare from a 'Slightly Worse-Off Population', but still knowably lexically superior, so that even if the Slightly Worse-Off Population were a trillion times bigger, instantiating the Slightly Better-Off Population would make for a determinately, knowably better outcome. But this is not the situation.

If the cliff-edge is indeterminate then there aren't two adjacent populations such that the Slightly Better-Off Population is *determinately* lexically superior to its neighbour. One might have to go quite far down the welfare spectrum to find a population that's determinately lexically inferior to the Slightly Better-Off Population. "There are two adjacent suffering levels, one lexically inferior to the other" is a supertruth, and has the same shadowy semi-plausibility as other supertruths. Vagueness mitigates the implausibility in one hair making a man not-bald, why not here?

That argument is perhaps less effective under epistemicism, but I think it does some work: part of the implausibility of a cliff-edge is that there could be two arbitrarily similar populations, as close as you like in welfare, except that one is knowably radically superior to the other. Epistemicism spares us that.

And how similar would the populations on either side of the cliff be? Handfield and Rabinowicz are correct that the harm types on either side of the cliff-edge might be 'qualitatively very similar', but this doesn't imply that they are *evaluatively* very similar. Though the differences in felt pain between adjacent outcomes are very small, there could be big leaps in evaluatively-significant properties depending on those pain levels: 'life not worth living' is a classic candidate, or my own 'wasted life', or Thomas's 'ecstatic'. There will also be broadly descriptive but evaluatively-significant properties lurking along the spectrum, such as the ability to sustain a life with deep personal relationships, sufficient leisure time, or enough security to allow democratic co-operation.

It's not excessively implausible that a small difference in lifelong pain levels could tip the balance in the application of these properties and that this could have very serious evaluative upshots. So even if the cliff-edge problem remains is unwelcome, it's not as bad as we might have thought. Given a forced choice, I would accept an evaluative cliff-edge instead of intransitivity in 'better than'.

But you might not agree: you might think that the cliff-edge problem is a decisive problem for vagueness models of at least some spectra. Or even if you accept numbers-resistance in some cases—for example, to avoid Parfit's Repugnant Conclusion—you might deny it elsewhere, such as in Depression Spectrum. If nothing else, having an *option* to model the spectrum as a sorites without lexical inferiority would be preferable, so we want to show that vagueness can explain away the plausibility of Pairwise-Later-Better whilst also accepting the analogue of the Repugnant Conclusion.

5 Spectra without Radical Inferiority

So we'd like a *numbers-sensitive* solution, which brings

Conservative Inferiority. If n people are suffering at some level, no matter how bad, then there's always some larger number of people nk such that it's *worse* for nk people to suffer at a slightly less severe level.

Conservative inferiority says that there's always a finite *exchange rate* between numbers and suffering. A certain change in the number of sufferers can always be compensated by a certain change in the level of their suffering, and vice versa. This exchange rate is likely to be highly variable, of course—there are (probably) no linear scales for numbers and for harms, and no constant exchange rate between them.

Assuming precision along the dimensions (harm and numbers, for example) the vagueness model of spectrum arguments with conservative inferiority is that the exchange rate is vague. But exchange rates are rather abstract, so I'll first put things in terms of a concrete evaluative property. Instead of lexically-inferior wasted lives, the numbers-sensitive model uses a slightly weaker evaluative property: a *damaged* life is conservatively inferior to a non-damaged life. A damaged life *can* be compensated by numbers.

To give an example without vagueness, suppose that O8 is conservatively inferior to O7, using the numbers from Depression Spectrum.

In that spectrum, there are five times as many sufferers in O7 as in O8. But suppose the exchange rate is much more demanding: to compensate for the increased harm and damaged lives in O8, there would need to be *six* times as many sufferers in O7. We have a false instance of Pairwise-Later-Better. Not because we cross some one-off lexical threshold, but simply because (at least) one of the trade-offs isn't worth it: the move from O7 to O8 imposes a lot of harm on the sufferers, and a five-fold decrease in numbers isn't enough to compensate. This is merely conservative inferiority, so Depression Spectrum can be revived by increasing the number of sufferers in O7, and (presumably) also increasing the number of sufferers in O1–O6 to keep constant the ratio of sufferers in adjacent outcomes.

But then—and this will be key to my purported solution to the spectra—often, with the higher numbers Distant-Earlier-Better is false, and transitivity is saved. So, for example, suppose we increased the number of sufferers in O7 and its predecessors ten-fold. Then we have 500 sufferers in O7 and a million in O1. As I said, I have my doubts about the badness of O1, so let's focus on O2: now there are 200,000 people awfully down in the dumps every two weeks. Is this clearly better than O8, which has 10 people extremely seriously depressed six days a week? It is not clear to me, and I think this instance of Distant-Earlier-Better is false.

The crucial question facing this kind of model is this: if O8 is worse than O7, why is the opposite verdict so plausible? Because we've not injected any vagueness yet. It can be determinate that one of the trade-offs is not worth it at current exchange rates, but vague which one.

To put this clearly, it'll be necessary to attach some numbers to the harms in Depression Spectrum. Assigning precise values is an absurd exercise in some ways, but that's part of the point of the sorites model: any attempt to attach precise values to the harms endured in the spectrum based on their short, cryptic descriptions is foolhardy. I don't think the vagueness of Depression Spectrum need involve any evaluative ontic vagueness; the descriptions of the harm levels simply involve vague language.

Suppose that outcome O1 has a value of 100 units (1 unit for every thousand people suffering the very mild depression in O1). The simplest numbers-sensitive model then says that the existence of a damaged life 'costs' 150 units of value. All else equal, if one outcome contains a person whose life has been damaged and another outcome doesn't, then the latter outcome is better by 150 units. The mere existence of a damaged life in an outcome causes the loss of value—no additional disvalue accrues to additional damaged lives.

Now, suppose it's indeterminate between sharpenings s_5, s_6, and s_7 which outcome (O5, O6, or O7) is the first to contain a damaged life. Then it's indeterminate at which stage the value plummets by 150 units, as in Table 3. I've supposed for simple modelling that at all other stages, value increases by 20 units.

Table 3: Depression Spectrum with a damaged life.

Outcome	Value (s_5)	Value (s_6)	Value (s_7)
O1	100	100	100
O2	120	120	120
O3	140	140	140
O4	160	160	160
O5	30	180	180
O6	50	50	200
O7	70	70	70
O8	90	90	90

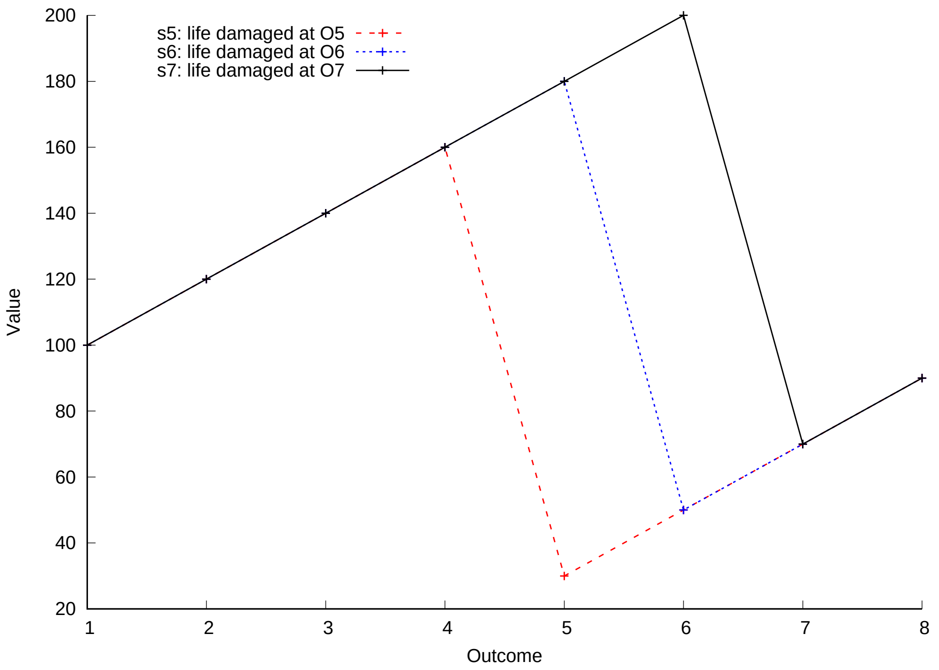

Figure 1: Value loss from a damaged life

This spectrum is graphed in Figure 1. The value at stage x is

$$80 + 20x - 150\lfloor x/s \rfloor$$

where s is the outcome at which lives are damaged on the sharpening in question. '$80 + 20x$' is an additional 20 units for each outcome, starting at 100 in O1, to reflect the shrinking number of sufferers; '$150\lfloor x/s \rfloor$' is a one-off cost of 150 units at outcome s and beyond to reflect a damaged life. The similarity to the utility functions for the vague projects in Chapter 5 is clearly no coincidence.

Outcome values are determinate (the same on all sharpenings—the lines on the graph overlap) in outcomes O1–O4 and O7–O8. But it's indeterminate where in O5–O7 the value plummets; Pairwise-Later-Better is false, but it's indeterminate whether the false instance is the comparison between O4 and O5, O5 and O6, or O6 and O7. 'Better than' is transitive on every sharpening, and therefore determinately transitive.

Crucially, this is merely conservative inferiority: the 150 unit value loss from a damaged life is enough to make O8 (barely) worse than O1, but were the population in O1 sufficiently larger, then the value at O1 would be higher, O1 would be at least as good as O8, and Distant-Earlier Better would be false. This is what we were looking for: there is no evaluative cliff-edge, but something like a hillside.

I have talked in indeterminist terms, but once again this picture is compatible with epistemicism about vagueness, or even ordinary uncertainty. This shouldn't be shocking, as I indicated above: the comparisons involved here are fairly abstract and the verdicts are not obvious. I would not be shocked to find out that O5 is far worse than O4, for example. I don't have strong evaluative intuitions. The model simply draws out *one* way in which ignorance or indeterminacy could raise its head: there is a severe (but not lexical) loss of value at one outcome as one life is damaged by severe depression, and it's not clear precisely how intense that depression must be, or many days a month it must strike, for a life to be damaged, which will again be subject to enormous interpersonal variation.

6 Exchange Rate Spectra

The model I've just sketched shows how we can have a sorites model of Spectrum Arguments without radical inferiority, without the cliff-edge.

As always in a spectrum argument, we must keep in mind several variables when comparing two outcomes: the number of people in the outcomes, the harm they endure, which of the outcomes is worse, and *by how much* it is worse. In order for O8 to be determinately worse than O1, there must be a severe (though as we've seen, not necessarily lexical) drop in value at the point where a life is damaged: in the numbers I'm using, -150. The first outcome where a life is damaged must be sufficiently worse than its predecessor to outweigh the improvements in the other pairwise comparisons. But given the large numbers of people involved, how *could* one life have that effect?

Any implausibility here can be avoided. The large drop is a consequence of loading all the vagueness into a single property (whether there is a single damaged life). In the language of chapters 5 and 6, I've modelled the spectrum as a *single-threshold* practical sorites, where we are trying to avoid a single drop (no longer able to build a cairn or write a book, damaging a life). That drop must be deep enough to outweigh all of the gains at other outcomes.

As with the Puzzle of the Self-Torturer, the model can be made more sophisticated by making the sorites *repeating*, so that each—or at least many—of the pairwise comparisons is indeterminate, but the long-term trend is determinately downwards. Instead of there being a single important evaluative threshold (being life-damaging), there are 'mini-thresholds' along the way (affecting one's work, affecting one's family, affecting one's hobbies, harming one's children...) and the structure described above with life-damaging plays out at each of them. Because there are many drops, no single one need be severe enough to make O8 worse than O1.

Another advantage of a repeating model is that because many of the adjacent pairwise comparisons will be vague (because many outcomes will be near the mini-thresholds), we can explain the hesitant phenomenology of Pairwise-Later-Better: many instances of it are borderline.

Fundamentally, the mini-thresholds reflect and perhaps determine the underlying exchange rate. What change in numbers compensates for a particular increase or decrease in depressive suffering? The most faithful model will show how this structure operates at the level of the exchange rate.

Here's a mini-spectrum with four levels of suffering, to show the structure of a repeating vagueness spectrum argument. Let there be four levels of suffering, and we start in outcome P1 with 1000 people at level 1. As in the original spectrum, we move through outcomes reducing the number of sufferers but increasing their level of suffering. Thus whether the move is an improvement or a worsening will depend on the exchange rate between numbers and intensity of suffering. Suppose that there are three sharpenings of the exchange rates, where 'three times worse than' means that a tripling in numbers is required to compensate, and so on:

- Sharpening E1: level 2 is three times worse than level 1; level 3 is four times worse than level 2; level 4 is five times worse than level 3.
- Sharpening E2: level 2 is three times worse than level 1; level 3 is five times worse than level 2; level 4 is four times worse than level 3.
- Sharpening E3: level 2 is five times worse than level 1; level 3 is four times worse than level 2; level 4 is three times worse than level 3.

These numbers are fairly low—level 1 is three to five times less harmful than level 2, for example—and comparable to the population ratios of adjacent outcomes in Depression Spectrum, in which O1 has five times as many sufferers as O2. These numbers are also not random. By multiplying the numbers in each sharpening, you can see that level 4 is determinately sixty times worse than level 1. The exchange rate between level 1 and level 4 is determinate, but the different sharpenings distribute it differently between the intermediate levels. The case could also have been presented in non-exchange rate terms with sufferers having determinate welfare levels in P1 and P4, but indeterminate at P2 and P3.

We can now present a spectrum using the specified exchange rates, with the populations in Table 4. Here's an illustration.

Table 4: A mini exchange-rate spectrum.

Outcome	Sufferers	E1	E2	E3
P1	1000	———	———	———
P2	300	better than P1	better than P1	worse than P1
P3	70	better than P2	worse than P2	better than P2
P4	20	worse than P3	better than P3	better than P3

So, for example, P2 has 30% the sufferering population of P1 does. Is this enough of a reduction to compensate for their greater suffering? Yes, according to sharpenings E1 and E2. P2 is better than P1, because the suffering in P2 is three times worse than in P1 but P2 has fewer than 1/3 the sufferers of P1. So the population has been reduced by more than enough to compensate for the increased suffering.

No, according to sharpening E3. The suffering in P2 is *five times more harmful* than that in P1. To be as good as P1, P2 would need to have 1/5 the sufferers (200) or fewer. It has more than that, so the decrease in sufferers is not enough to compensate for the increase in harm, and E3 says that P2 is worse than P1. Since the sharpenings disagree, it's indeterminate whether P1 is better or worse than P2, and similarly for the other adjacent pairwise comparisons.

The last outcome is determinately worse than the first, because the suffering in P4 is 60 times worse than that in P1, and the population in P4 is more than one-sixtieth of that in P1 (16 or 17, depending on rounding). Each pairwise comparison between neighbours is indeterminate, but the distal comparison between P1 and P4 is determinately for the worse. We have the same structure as a spectrum argument, but without lexical inferiority or even large but non-lexical drops in value at substantive thresholds. We simply have indeterminate exchange rates between harms and numbers.

The model could be made more 'realistic' in several further ways. We could increase the number of outcomes (and so the number of adjacent comparisons), and like the Self-Torturer, make it repeating—we could, for example, have 40 outcomes repeating the structure of Table 4 ten times. Then every comparison between adjacent outcomes would be indeterminate, but comparisons at longer ranges (separated by more than 4 outcomes) would always be for the worse. We could increase the number of sharpenings, so that each adjacent comparison is positive on 'the vast majority of' sharpenings, making Pairwise-Later-Better psychologically compelling but still determinately false.

Finally, we could replace 'suffering at level 2' and the like with brusque descriptions ('somewhat serious nagging pain'). All of these changes might make it closer to the Spectrum Arguments we've seen. But at the cost of clarity, and the existing model in Table 4 has all the core features of the spectra *without* sacrificing transitivity.

Why accept this model? I don't simply get to assume that exchange rates between numbers and suffering are unknowable or indeterminate—that would be question-begging. As I mentioned above, part of the story is that the brief phenomenal descriptions in spectrum arguments are too vague to pick out determinate, knowable levels of suffering and harm, even ignoring the broader evaluative issues. I've just illustrated how vague exchange rates can engender something very like the Spectrum Argument, and the descriptions in the original case are exactly what we would expect to lead to such vagueness. 'Feeling pretty down' is less precise than 'is a heap', so we should independently expect vague exchange rates in its application.

In any case, the dialectical pressure goes the other way: uncertainty, ignorance, and indeterminacy of value are not inherently implausible, and they allow us to model the Spectrum Argument without intransitivity or lexical inferiority (which *are* implausible, or at least more contested). That itself is good reason to think that one of the former phenomena must be present in the spectra: we can avoid an implausible claim by making a less implausible one. I don't say that this is decisive reason to think that the spectra are sorites, but it puts a lot of pressure on those who would deny it.

7 Temkin contra Sorites

Temkin one of them. He is of course not blind the possibility that the Spectrum Argument involves a sorites. He argues that the 'standard sorites' is structurally inadequate to model the Spectrum Argument, and that the 'revised sorites paradoxes' which might indeed look like Spectrum Arguments are so different from standard ones that the sorites response fails as a defence of transitivity.[16]

The real question about the sorites model is not "how similar is it to the canonical sorites?" but "is vagueness doing the work?". Though the sorites model of the Spectrum Argument is far more complex than a simple sorites involving removing one grain from a heap, the additional complexity is due to multi-dimensionality. Vagueness, not intransitivity, remains the underlying phenomenon.[17]

I'll illustrate this with hair. Consider an ordinary sorites, involving a tolerance principle such as "if a man with n hairs is hairy, then a man with (n-1) hairs is hairy". Temkin draws a disanalogy with the Spectrum:

> the Standard Sorites paradox does *not* claim that if someone is hairy, and you remove a single hair from his head, then he will be clearly *hairier* than before [...] for the argument to parallel mine, it would have to be arguing that the person is clearly getting hairier with *each* hair removal, though at the end he has become much *less* hairy than he was initially.[18]

Pace Temkin, and as Thomas notes about a similar case, in the sorites reconstruction of Depression Spectrum, it is *not true* on most accounts of vagueness (including those I'm working with) that 'the person is clearly getting hairier with *each* hair removal'. It looks that way, but this claim simply a sorites tolerance principle.[19] So a sorites model of the spectrum doesn't need to make a parallel claim true—depending on the theory of vagueness, quite the opposite.

I'll respond to Temkin's argument by presenting a numbers-sensitive single-threshold model—which is easier to visualise—of a hairiness sorites series akin to the spectrum. The correct hairiness analogy with the Spectrum Argument is multidimensional, in that we are weighting two components against each other. Similarly, in Depression Spectrum two dimensions—the number of depressed and the level of their suffering—interact. But in both cases we are still dealing with a sorites, because ultimately vagueness along one dimension (whether hair-numbers or suffering exceeds some threshold) does the work.

Your goal is to infiltrate the Hirsute Men's Club, so it's better to have more and non-balding men. It's better to have more non-balding men than fewer, because that increases the likelihood of a successful infiltration. But balding men are still useful, so it's all-else-equal better to have more balding men than fewer.

You initially start with five very hairy men. But you can make trade-offs in a very specific way, increasing the number of men but redistributing the available hairs amongst them, as in Table 5.

[16] Temkin (2012), pp. 277ff.

[17] Jacob Nebel also suggests that multi-dimensionality will allow for a sorites model of the spectra. See Nebel (2018), pp. 892–894. Compare also Pummer (2018) on the '2D sorites argument' (p. 1737).

[18] Temkin (2012), p. 278.

[19] Thomas (2021), p. 12.

Table 5: Hairiness Spectrum.

Stage	Head status	Number of men
H1	Extremely hairy, looks like a shaggy dog	5
H2	Very hairy, long unkempt hair	10
H3	Somewhat long (shoulder-length) dense hair	15
H4	A business haircut	20
H5	A slight receding hairline	25
H6	A noticeably receding hairline and bald spots	30
H7	Widespread bald spots	35
H8	Completely hairless head except for a goatee	40

As we go down the table, for your purposes things improve at each step, but then there is a sharp discontinuity and things get dramatically worse when we cross from the non-balding to the balding men.[20] The situation can be visualised if we attach some (again, arbitrary) numbers.

Let's suppose that each non-balding man is worth 10 units, and each balding man is worth 1 unit. (So this is conservative inferiority, and the exchange rate between the two types of men is 10:1). It's vague at what stage the men are balding. As usual, let's say it's indeterminately either at stage H5, H6, or H7. The sharpenings of the value function are graphed in Figure 2. You can see that H1 has some positive value, H2 is better, and so on, but the value collapses as the men become balding. It then rises again, much more shallowly, as balding men are added. There is a sharp discontinuity, but because of the the vagueness of 'balding' it's indeterminate at what stage that discontinuity is.

This model is clearly a sorites, trading on the vagueness of 'balding'. But it is multidimensional and mirrors the structure of the spectrum argument. Transitivity of both 'balder than' and 'better than' are true here, and the equivalent of Pairwise-Later-Better is false. Note also that if you don't find any of the adjacent comparisons compelling, adding intermediate stages between (say) H4 and H5 would make them much harder to deny, which is another feature of the spectrum argument. There are no structural reasons why spectra can't be sorites.

8 Degrees of Incommensurability?

I've argued that vagueness can provide a good model of the spectrum argument, especially when generalised to include indeterminate exchange rates rather than lexical inferiority. My argument has been that since evaluative vagueness is more plausible than evaluative intransitivity, we should accept the vagueness model.

But perhaps there's another model, an alternative to both? Alan Hájek and

[20] I use the phrase 'sharp discontinuity' because that's how Temkin phrases the putative contrast between spectrum and sorites.

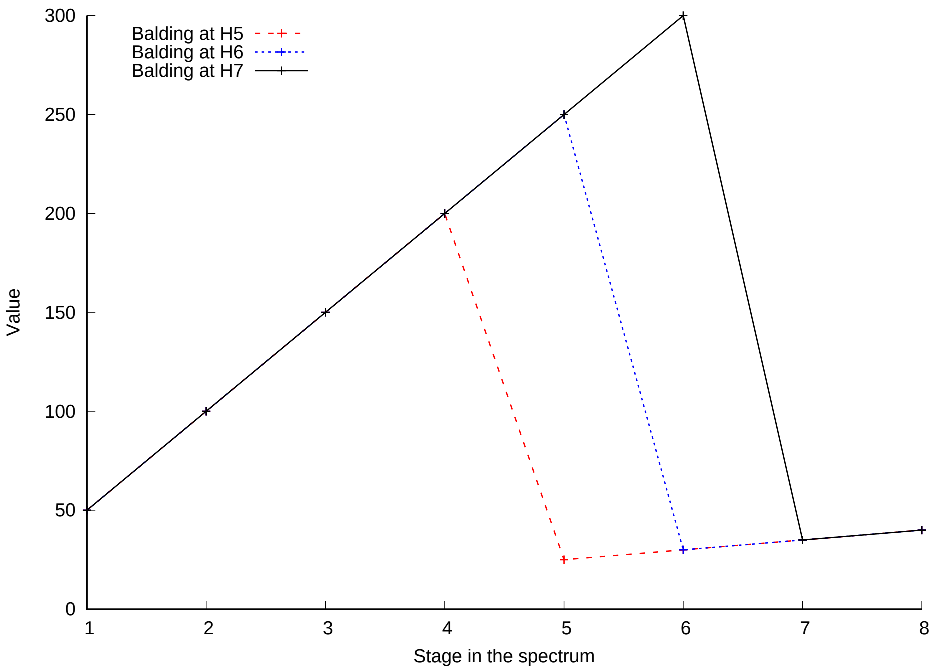

Figure 2: Baldness sorites

Wlodek Rabinowicz argue that the Fitting-Attitude account of value can explain crucial features of the spectrum argument.[21] They claim that incommensurability is not an all-or-nothing matter, and that Pairwise-Later-Better is false because at least some of its instances are outright false, because some outcome is *slightly incommensurate* with the next one: O4 may be merely 'nearly better' than O3, for example. This can happen when there are multiple permissible preference orderings, most but not all of which rank O4 above O3.

Their model is impressively technically adept, going beyond slogans and into, for example, infinite populations and unequal weightings of preference orders. The formal structure of their account is—as Hájek and Rabinowicz note—awfully close to vagueness, especially once degrees of truth are admitted: instead of saying that *O4 is nearly better than O3*, one might say that that *it's nearly determinate that O4 is better than O3*, which would take us very close to John Broome's 'graded' vagueness account of incommensurability.[22] They ultimately reject the vagueness view as the primary diagnosis of the spectrum argument, because they are attached to the Fitting-Attitude account of value.[23]

But I think the vagueness model could be paired with such an account of value with minimal distortion, in a broadly supervaluational manner: it is indeterminate which things are better, because indeterminate what the permissible preference orderings are, and what Hájek and Rabinowicz identify as the set of permissible preference orderings is really the set of *indeterminately*-permissible preference orderings. (Indeed, this could be a plausible account of how vagueness arises in impartial goodness, a question I've not considered in this chapter be-

[21] Hájek and Rabinowicz (2022).

[22] See Broome (2021), especially pp. 45ff.

[23] Hájek and Rabinowicz (2022), p. 15.

yond my remarks about short cryptic evaluative descriptions.)

Setting aside such speculative views, why prefer my vagueness approach to their degrees of incommensurability approach? My main argument is phenomenological: vagueness better explains hesitancy in the face of pairwise comparison. Their diagnosis of why each instance of Pairwise-Later-Better seems compelling but often attracts some hesitancy is almost-betterness: O4 is ranked higher than O3 on *nearly all* permissible orderings so the disparity in measure explains the overall tendency to prefer O4 and but the exceptions explain the hesitancy, hesitancy that 'reflects acknowledging that one could reasonably have other preferences' by preferring O3.[24]

If both sets of preferences—O3 $\prec$ O4 and O4 $\prec$ O3—are genuinely, determinately permissible, then it's not obvious that we would or should hesitate. Whichever way one prefers one is preferring in a permissible way, albeit not the only permissible way. So why hesitate? Vagueness has a better explanation here: whichever way one prefers, one prefers in only an *indeterminately*-permissible way, and so we should expect some hesitancy and unease, as we see under incomplete preferences and the sorites. And under epistemicism we should expect the same, since we do not (perhaps cannot) know if we are choosing the best option.

Moreover, if O3 is almost worse than O4, then on the face of it the Hájek-Rabinowicz degrees approach should deem a 'minority' preference for O3 over O4 wholly comprehensible and not criticisable. It is a minority, but a permissible minority—not merely borderline-permissible. But in practice it's only schooling in a theory of value (or in the sorites) that would prevent us from seeing the minority preference as a serious mistake.

Debatable phenomenological considerations aside, I am inlined to reject Hájek and Rabinowicz's construal because the vagueness approach is more ecumenical and not beholden to the Fitting Attitudes account. But if you are already committed to that account, then they may have offered a more serious contender for you.

9 Conclusion

Numbers-resistance is an upshot of existing sorites models of the Spectrum Argument. I've argued that this is not as implausible as we might fear. But having it forced upon us is a cost. So I've presented a numbers-sensitive vagueness model of the argument, and we can avoid both intransitivity and radical inferiority.

Let me restate (in indeterminacy terms) the difference between the numbers-resistant and numbers-sensitive models of Depression Spectrum:

- In numbers-resistant models, there is some absolute level of suffering so bad that no reduction in the number of sufferers can outweigh suffering at that level—it is lexically or radically inferior. This threshold level is indeterminate. No number of hours of push-pin is as good as one hour of poetry, but it's indeterminate where poetry stops and push-pin begins.

[24] Hájek and Rabinowicz (2022), p. 4.

- In numbers-sensitive models, there's no lexical or radical inferiority—any increase in suffering can be compensated by a reduction in the number of sufferers. But it's indeterminate precisely what the 'exchange rate' is between suffering and numbers. Some number of number of hours of pushpin is as good as one hour of poetry, but it's indeterminate what that number is.

We can think of numbers-resistance or lexical inferiority as being numbers-sensitivity with an occasionally infinite exchange rate, but I think that masks some of the underlying evaluative structure.

I'll restate my (numbers-sensitive) diagnosis of Depression Spectrum. The trick is in the numbers of sufferers in O1, and relatedly the ratios of sufferer numbers in adjacent outcomes. At low numbers (only 100 people in O1, say), Distant-Earlier-Better is clearly true but at least one instance of Pairwise-Later-Better is obviously false: if O7 didn't even halve the number of sufferers compared to O6, then we'd say without a doubt that O7 is worse.

At very high numbers of sufferers, the ratios of sufferers in adjacent outcomes can be high enough to avoid indeterminacy and make Pairwise-Later-Better true. But there is no intransitivity because Distant-Earlier-Better is false: it *is* better for ten—10—people to be extremely seriously depressed 6 days a week than for ten trillion—10,000,000,000,000—people to be awfully down in the dumps once every two weeks. (I put things in terms of the harm in O2, here, because as I say, the harm in O1 is so mild that it's not obviously at all bad.)

Depression Spectrum is so puzzling is because it straddles the line between the very high numbers where Distant-Earlier-Better is obviously false and the very low numbers where Pairwise-Later-Better is obviously false. It inherits (conflicting) plausibility from both extremes, leading to apparently contradictory verdicts. By seeing this and the vagueness involved at intermediate numbers, we can preserve transitivity without embracing radical or lexical inferiority.

Numbers-sensitivity *does* allow for the numbers to be increased to the point that the initial outcome is better than the final outcome—it embraces the depressive cousin of Parfit's Repugnant Conclusion—but as I've argued, that is intuitively plausible for Depression Spectrum. A galaxy-full of people feeling down every two weeks is indeed worse than ten people suffering extreme depression. If you don't share that judgement, the numbers-resistant model avoids repugnance at the cost of a cliff-edge. Even those of us who find numbers-resistance implausible should object even more strongly to a denial of the transitivity of 'better than'.

In both models, *if* the final outcome in the spectrum is indeed worse than the first outcome, there is one or more false instance of Pairwise-Later-Better—false by a big enough margin to carry the evaluative judgement. The false instance is very hard to find, perhaps because it's indeterminate which instance it is. Nevertheless, transitivity is saved.

References

Andreou, Chrisoula. 2023. *Choosing Well: The Good, the Bad, and the Trivial*. 1st ed. New York: Oxford University Press. https://doi.org/10.1093/oso/9780197584132.001.0001.

Broome, John. 2021. "Incommensurateness Is Vagueness." In *Value Incommensurability*, by Henrik Andersson and Anders Herlitz, 1st ed., 29–49. New York: Routledge. https://doi.org/10.4324/9781003148012-3.

Chang, Ruth. 2016. "Parity, Imprecise Comparability and the Repugnant Conclusion." *Theoria* 82 (2): 182–214. https://doi.org/10.1111/theo.12096.

Cowie, Christopher. 2022. "A New Argument for Moral Error Theory." *Noûs* 56 (2): 276–94. https://doi.org/10.1111/nous.12357.

———. 2023. "Why Moral Paradoxes Support Error Theory." *The Journal of Philosophy* 120 (9): 457–83. https://doi.org/10.5840/jphil2023120927.

Elson, Luke. 2017. "Incommensurability as Vagueness: A Burden-Shifting Argument." *Theoria* 83 (4): 341–63. https://doi.org/10.1111/theo.12129.

Hájek, Alan, and Wlodek Rabinowicz. 2022. "Degrees of Commensurability and the Repugnant Conclusion." *Noûs* 56 (4): 897–919. https://doi.org/10.1111/nous.12388.

Handfield, Toby, and Wlodek Rabinowicz. 2018. "Incommensurability and Vagueness in Spectrum Arguments: Options for Saving Transitivity of Betterness." *Philosophical Studies* 175 (9): 2373–87. https://doi.org/10.1007/s11098-017-0963-9.

Huemer, Michael. 2008. "In Defence of Repugnance." *Mind* 117 (468): 899–933. https://doi.org/10.1093/mind/fzn079.

Nebel, Jacob M. 2018. "The Good, the Bad, and the Transitivity of Better Than." *Noûs* 52 (4): 874–99. https://doi.org/10.1111/nous.12198.

Parfit, Derek. 1984. *Reasons and Persons*. Oxford [Oxfordshire]: Clarendon Press.

———. 2016. "Can We Avoid the Repugnant Conclusion?" *Theoria* 82 (2): 110–27. https://doi.org/10.1111/theo.12097.

Pummer, Theron. 2018. "Spectrum Arguments and Hypersensitivity." *Philosophical Studies* 175 (7): 1729–44. https://doi.org/10.1007/s11098-017-0932-3.

Qizilbash, Mozaffar. 2005. "Transitivity and Vagueness." *Economics and Philosophy* 21 (1): 109–31. https://doi.org/10.1017/S0266267104000410.

Rachels, Stuart. 1998. "Counterexamples to the Transitivity of *Better Than*." *Australasian Journal of Philosophy* 76 (1): 71–83. https://doi.org/10.1080/00048409812348201.

Temkin, Larry S. 2012. *Rethinking the Good: Moral Ideals and the Nature of Practical Reasoning*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199759446.001.0001.

Thomas, Teruji. 2021. "Are Spectrum Arguments Defused by Vagueness?" *Australasian Journal of Philosophy*, June, 1–15. https://doi.org/10.1080/00048402.2021.1920622.