

*The Good, the Bad, and the Transitivity of Better Than**

JACOB M. NEBEL
New York University

Abstract

The Rachels–Temkin spectrum arguments against the transitivity of *better than* involve good or bad experiences, lives, or outcomes that vary along multiple dimensions—e.g., duration and intensity of pleasure or pain. This paper presents variations on these arguments involving combinations of good and bad experiences, which have even more radical implications than the violation of transitivity. These variations force opponents of transitivity to conclude that something good is worse than something that isn't good, on pain of rejecting the good altogether. That is impossible, so we must reject the spectrum arguments.

According to

The Transitivity of *Better Than*: For any bearers of value *A*, *B*, and *C*, if *A* is better than *B*, all things considered, and *B* is better than *C*, all things considered, then *A* is better than *C*, all things considered.¹

This paper is about some purported counterexamples to transitivity developed by Rachels (1993; 1998; 2001) and Temkin (1987; 1996; 2012). Following Temkin, I call them *spectrum arguments*.

The spectrum arguments against transitivity present a sequence of outcomes, lives, or experiences, which vary along multiple dimensions—e.g., duration and intensity of pleasure or pain. Here is an example due to Rachels, which I call

The Bad Spectrum:

A: 1 year of excruciating agony.

B: 2 years of pain slightly less intense than the pain in *A*.

C: 4 years of pain slightly less intense than the pain in *B*.

D: 8 years of pain slightly less intense than the pain in *C*.

...

Y: 2²⁴ years of pain slightly less intense than the pain in *X*.

Z: 2²⁵ years of pain slightly less intense than the mild pain in *Y*.

The excruciating agony in *A*, for example, might be the worst kind of torture you can imagine. The mild pain in *Z* might be the discomfort of a hangnail or a waning

*Thanks to Kara Dreher, Ben Holguin, Simon Knutsson, Michael Otsuka, Derek Parfit, Jim Pryor, Trevor Teitel, Jake Zuehl, and two anonymous referees for comments. Thanks, most of all, to Larry Temkin for inspiration, advice, and many hours of discussion.

mosquito bite. Following Temkin (2012, 284), I assume that the experiences in each spectrum are endured in lives of equal length, and that each life is similar to its successor, except with respect to the experiences described. We might imagine that, aside from these painful experiences, the life is otherwise neutral—e.g., because the rest of it is spent unconscious, or in ways that are neither better nor worse than unconsciousness (as imagined by Rachels 1998, 78). To many people, *A* seems better than *B*, which seems better than *C*, which seems better than *D*, . . . , which seems better than *Y*, which seems better than *Z*, but *Z* seems better than *A*. If each comparison is as it seems, then *better than* is not transitive.

Consider next

The Good Spectrum:

A: 1 year of extraordinary ecstasy.

B: 2 years of pleasure slightly less intense than the pleasure in *A*.

C: 4 years of pleasure slightly less intense than the pleasure in *B*.

D: 8 years of pleasure slightly less intense than the pleasure in *C*.

. . .

Y: 2^{24} years of pleasure slightly less intense than the pleasure in *X*.

Z: 2^{25} years of pleasure slightly less intense than the mild pleasure in *Y*.

In this case, *Z* seems better than *Y*, . . . , which seems better than *B*, which seems better than *A*, but (to many people) *A* seems better than *Z*. If each comparison is as it seems, then *better than* is not transitive.

I call anyone who accepts these comparisons a *cyclist*. Cyclists believe that there are, or could be, betterness cycles—i.e., sequences of items x_1, \dots, x_n in which each x_i is better than x_{i+1} , and yet x_n is better than x_1 . One could, in principle, reject transitivity without being a cyclist.² But the most powerful arguments against transitivity involve apparent betterness cycles, as in the spectrum arguments above.

These two examples—the good spectrum and the bad spectrum—will serve as adequate samples of the Rachels–Temkin spectrum arguments. Temkin (2012) presents a wide variety of other spectrum arguments, but the ones above will allow us to bring out the relevant issues in the simplest form. What all of these arguments have in common is (among other things) that, within each spectrum, the items are either all good or all bad, all things considered. The experiences in the bad spectrum, for example, are bad; those in the good spectrum are good. What do I mean by “good” and “bad”? For now, I’m taking some ordinary sense of these words for granted. Plausibly, a year of extraordinary ecstasy is good if anything is, and a year of horrible agony is bad if anything is.³

We might hope that this common feature of the Rachels–Temkin spectrum arguments would mitigate the deliberative crisis posed by a rejection of transitivity. If transitivity fails, then it may seem impossible for the *better than* relation to guide our choices or attitudes in many cases. This is because, for any outcome we choose or prefer, there may be some better alternative that we ought to have chosen or preferred instead. The rejection of transitivity might, therefore, lead to skepticism about practical reasoning, or at least about the role of the good in practical reasoning.⁴ But this deliberative crisis might not be too severe if there is some

partition of outcomes into those that are good, those that are bad, and those that are neither good nor bad. For we would at least know not to choose or prefer outcomes that are bad rather than alternatives that are good. For example, it would be wrong to bring about one of the outcomes in the bad spectrum when we could instead bring about one of the outcomes in the good spectrum. That would be nice.

But we can combine good and bad experiences to form a new, and no less compelling, kind of spectrum argument. This new kind of spectrum argument will show that the reasoning behind the Rachels–Temkin arguments has consequences that are even more radical than the violation of transitivity.

I present my modified spectrum arguments in section 1. I then argue, in sections 2 and 3, that even those who might be willing to reject the transitivity of *better than* should not embrace the more radical implications of my variations. We should, therefore, reject the spectrum arguments against transitivity. I conclude, in section 4, by considering how these arguments might be rejected. I tentatively suggest that the arguments are instances of the sorites paradox.

1. Combined Spectrum Arguments

We begin with an outcome *A* that contains a single year of excruciating agony and also contains some amount of ecstasy. How much ecstasy, you ask? Enough so that *A* is neutral—neither good nor bad overall. For some, that might just be a single year. For others, it takes much more pleasure to counterbalance pain. I will suppose that it takes three years of extraordinary ecstasy to counterbalance the year of agony.⁵ (Readers who think that no amount of ecstasy could counterbalance a year of agony—i.e., that any life with one year of agony is bad, no matter how much ecstasy it contains—might instead imagine some shorter duration of agony, which could be counterbalanced by three years of ecstasy. Some of my arguments would, I think, succeed even for such shorter durations. And some of my arguments might succeed even if no duration of ecstasy could counterbalance any duration of agony.)

Think of it this way. Suppose that we could choose whether or not to bring someone into existence with a certain kind of life. If this life would be miserable overall, it would be wrong to bring her into existence, other things being equal. That would be true, for example, if her life contained a year of excruciating agony and many years of neutrality. But surely (I think) there is some number of years of extraordinary ecstasy such that, if her life contained such years in addition to the year of agony (and many years of neutrality), it would not be wrong to bring her into existence.⁶ I am supposing that three is one such number. Readers who deny that three is such a number, or that there is any such number, are invited to imagine longer durations of ecstasy, or shorter durations of agony, instead.

Now consider

The Bad Spectrum with Ecstasy:

A: 1 year of excruciating agony and 3 years of ecstasy.

B: 2 years of pain slightly less intense than the pain in *A* and 3 years of ecstasy.

C: 4 years of pain slightly less intense than the pain in *B* and 3 years of ecstasy.

D: 8 years of pain slightly less intense than the pain in *C* and 3 years of ecstasy.

...

Y: 2^{24} years of pain slightly less intense than the pain in *X* and 3 years of ecstasy.

Z: 2^{25} years of pain slightly less intense than the mild pain in *Y* and 3 years of ecstasy.

As we descend down the alphabet, the outcomes seem to get worse and worse. And yet *A* seems worse than *Z*. These judgments together violate transitivity.

But consider the following. *A* is neither good nor bad, all things considered. (If you disagree, then adjust the duration of pleasure or pain as needed.) However, some of *A*'s successors along the spectrum are surely bad overall—*D*, for instance. (If you disagree, then move further down the alphabet.) Now consider

The Badness Principle:

For any value bearers *x* and *y*, if *x* is bad and *y* is worse than *x*, then *y* is also bad.

If *D* is bad, and if *F* is worse than *D*, then *F* must also be bad. *D* is indeed bad. And *F* certainly seems worse than *D*. So *F* must also be bad. Since each outcome seems worse than its predecessor—i.e., than the outcome labeled by its immediate predecessor in the alphabet—iterations of this reasoning tell us that *Z* is bad, too. But *A*, which is not bad, seems worse than *Z*. This violates the badness principle. (Note that this argument does not assume transitivity.)

We have just seen that, given the badness principle, the following claims are inconsistent:

- (1) In the bad spectrum with ecstasy, each outcome is worse than its predecessor.
- (2) In the bad spectrum with ecstasy, *A* is worse than *Z*.
- (3) Some of the outcomes in the bad spectrum with ecstasy are bad.
- (4) Not all of the outcomes in the bad spectrum with ecstasy are bad.

Claims (1) and (2) generate the argument against transitivity. If we add claims (3) and (4), then we get a violation of the badness principle. More generally, if we have a betterness cycle, then the badness principle requires that either *all* or *none* of the outcomes in the cycle be bad.⁷ One bad apple really can ruin the barrel!

Which claim should we reject? I am inclined to see this as a *reductio* of the conjunction of (1) and (2)—i.e., the crucial judgments in the Rachels–Temkin arguments against transitivity.⁸ I am not certain which of these two claims I would reject (although I make some speculative remarks in section 4). But these two claims seem to me collectively less plausible than (3), (4), and the badness principle.

Consider (3). Surely, if anything is bad, then at least one of the outcomes in the bad spectrum with ecstasy is bad. To show this, consider a month of moderate pain. If anything is bad, that's bad. It would be wrong to bring someone into existence

if her life would contain that experience alone. But, I submit, at least one of the outcomes in the bad spectrum with ecstasy is *worse* than that, all things considered. Sure, they each involve three years of extraordinary ecstasy. But they also involve vastly greater durations and intensities of pain than our month of moderate pain. So, if anything is bad, then at least one of the outcomes in the bad spectrum with ecstasy is bad.

That conditional is not enough to secure (3). For maybe nothing whatsoever is bad. I consider that possibility in section 3. For now, it suffices to note that rejecting (3) would be a radical move.

Next consider (4). Some might claim that *all* of the outcomes in the bad spectrum with ecstasy are bad. But we designed *A* to be neutral. Even if three years of ecstasy is not enough, then surely there is some greater duration of pleasure that would be enough to prevent *A* from being bad overall.⁹ Some might follow Gurney (1887) in holding that no amount of pleasure can compensate for sufficiently intense agony. But then consider *Z*, which contains only mild discomfort. Surely *some* quantity of pleasure would be enough to make *Z* good, all things considered.¹⁰ And yet, if we followed Gurney, that quantity would not make all of the outcomes good, so we would continue to accept (3).

That leaves the badness principle. I discuss it in section 2, after expanding my argument.

We can now consider another spectrum of this kind:

The Good Spectrum with Agony:

A: 3 years of ecstasy and 1 year of agony.

B: 6 years of pleasure slightly less intense than the pleasure in *A*, and 1 year of agony.

C: 12 years of pleasure slightly less intense than the pleasure in *B*, and 1 year of agony.

D: 24 years of pleasure slightly less intense than the pleasure in *C*, and 1 year of agony.

...

Y: 3×2^{24} years of pleasure slightly less intense than the pleasure in *X*, and 1 year of agony.

Z: 3×2^{25} years of pleasure slightly less intense than the mild pleasure in *Y*, and 1 year of agony.

Each outcome seems better than its predecessor, and yet *A* seems better than *Z*. This violates transitivity. But it is also plausible that at least one outcome in this spectrum is good, all things considered. *D* seems good overall, for example. (Readers who disagree might imagine that each life in this spectrum contains some shorter period of agony.) But if *D* is good overall, and if each outcome is better than its predecessor, then *E* must also be good overall, by

The Goodness Principle:

For any value bearers *x* and *y*, if *x* is good and *y* is better than *x*, then *y* is also good.

So Z must be good all things considered. But then A , which seems better than Z , must be good. However, we have designed A so that it is neutral. And Z , in any case, does not seem good. This illustrates a similar inconsistency, given the goodness principle:

- (5) In the good spectrum with agony, each outcome is better than its predecessor.
- (6) In the good spectrum with agony, A is better than Z .
- (7) Some of the outcomes in the good spectrum with agony are good.
- (8) Not all of the outcomes in the good spectrum with agony are good.

As before, claims (5) and (6) ground the challenge to transitivity. If we add (7) and (8), then we violate the goodness principle. If we reject (5) or (6), then we can reject the Rachels–Temkin spectrum arguments against transitivity.

We can now turn to a third set of outcomes, which I consider to be the most compelling variation on the spectrum arguments:

The Combined Spectrum:

A : 3×2^{25} years of mild pleasure, and 1 year of agony.

B : 3×2^{24} years of pleasure slightly more intense than the pleasure in A , and 2 years of pain slightly less intense than the pain in A .

...

Y : 6 years of pleasure slightly more intense than the pleasure in X , and 2^{24} years of pain slightly less intense than the pain in X .

Z : 3 years of ecstasy even more intense than the pleasure in Y , and 2^{25} years of pain even less intense than in the pain in Y .

In the combined spectrum, each outcome seems much worse than its predecessor, and yet A seems worse than Z . But some of the outcomes (e.g., A) seem bad, all things considered, and others (e.g., Z) seem good. Given the goodness and badness principles, these claims are inconsistent:

- (9) In the combined spectrum, each outcome is worse than its predecessor.
- (10) In the combined spectrum, Z is better than A .
- (11) Some of the outcomes in the combined spectrum are either good or bad.
- (12) Not all of the outcomes in the combined spectrum are good, and not all of them are bad.

Note that (11) is weaker than the corresponding claims about the good spectrum with agony and the bad spectrum with ecstasy. (11) is satisfied just in case at least one of the outcomes in the combined spectrum is either good or bad. Suppose that we rejected this premise. Then, I think, we should conclude that *nothing* is good or bad.¹¹ For if some outcomes were good or bad, then we would expect them to be better or worse than some outcomes that could be situated in some purported betterness cycle. For example, five hundred years of excruciating agony would be bad if anything is. But it would also be worse than than fifty years of slightly more excruciating agony, along with some massive duration of mild pleasure. And

the latter outcome could be situated in some purported betterness cycle like the combined spectrum above.

Claim (12) is stronger than its analogues regarding the good spectrum with agony and the bad spectrum with ecstasy. But it is still eminently plausible. If we think, with the cyclists, that a sufficient duration of extraordinary ecstasy would be better than any duration of mild pleasure, and that a sufficient duration of horrible agony would be worse than any duration of mild pain, then we should judge *A* to be bad and *Z* to be good—at least, given suitable durations for our pleasures and pains. After all, on this way of thinking, *A* would be worse than a few minutes of mild pain, which would be bad, and *Z* would be better than a few minutes of mild pleasure, which would be good. (12) would then follow by the goodness and badness principles. If, on the other hand, we think that for any duration of ecstasy, there is some duration of mild pleasure that would be better, and that for any duration of agony, there is some duration of mild pain that would be worse, then we should still accept (12): we would judge *A* to be good and *Z* to be bad, given suitable durations of mild pleasure and pain. Therefore, we should accept (12) regardless of whether we share the cyclists' judgments about long durations of mild pleasure and pain.

That leaves two options for the cyclists. The first option is to reject the goodness and badness principles. I consider this option in section 2. The second option, which I consider in section 3, is to deny that anything is good or bad. I argue that neither option is plausible.

2. The Goodness and Badness Principles

The goodness and badness principles seem to me obvious. I do not have a decisive argument for them. But, in this section, I mention several considerations that support these principles.

2.1. *Monotonicity*

Many people would accept the goodness and badness principles as instances of a more general principle:

Monotonicity:

For any gradable adjective *F*, if *x* is *F* and *y* is *F*er than *x*, then *y* is *F*.

Note that the positive form *F* must be understood relative to the same context or comparison class in both the antecedent and the consequent. For example, if Sally is tall for a kindergartener and Muggsy Bogues is taller than Sally, it does not follow that Muggsy Bogues is tall for an NBA player. But if someone counts as tall in some context, relative to some fixed comparison class, then anyone taller than that person would also count as tall in that context, relative to that same comparison class. I return to this kind of issue in section 3. But, for now, let us bracket it.

Monotonicity is widely accepted by linguists and philosophers.¹² One reason for this is that monotonicity follows from the most natural views about the semantics

of the positive and comparative forms of gradable adjectives. There are two kinds of such views: the first kind takes the comparative form as basic; the second kind takes the positive form as basic.

Proponents of the first kind of view tend to think of gradable adjectives as relations between individuals and degrees (Kennedy 2004). For example, the adjective *tall* expresses a relation between an object x and a degree of height d such that the object's height is at least as great as d . For the adjective's positive form, the relevant degree argument is some contextually relevant standard. For example, one counts as tall in some context if and only if there is a degree d to which one is tall such that d is at least as great as the contextually relevant standard (Barker 2002, 7).

Monotonicity can then be obtained in either of two ways. The first is by assuming the transitivity of the *at least as great as* relation over the set of degrees. By transitivity, if y is taller than x and x is at least as tall as the relevant standard of comparison, then y is at least as tall as that standard; therefore, y too counts as tall. It would, of course, beg the question to appeal directly to transitivity in the present context. But monotonicity can also be obtained in a second way. It is generally assumed that if x is F to degree d , then x is also F to any degree d' where $d \geq d'$.¹³ This entails monotonocity. For if I am taller than you, then I am tall to any degree that you are. So if there is a degree to which you are tall which exceeds the contextually relevant standard, then I am tall to that degree too.

I have just explained why monotonicity follows from the first kind of view about gradable adjectives, on which the positive form is derived from the comparative form. Monotonicity also follows from the second kind of view, on which the comparative form is derived from the positive form. On this kind of view, Muggsy is taller than Sally just in case there is some comparison class relative to which Muggsy counts as tall and Sally does not (Klein 1980). Suppose, for *reductio*, that Sally is tall (in some context), that Muggsy is taller than Sally, but that Muggsy is not tall (in that context). Then there is some comparison class (namely, the one relevant to our present context) relative to which Sally counts as tall but Muggsy does not. But, by our definition of *taller than*, Sally would then count as taller than Muggsy. We have both that Sally is taller than Muggsy and that Muggsy is taller than Sally. This violates the asymmetry of *taller than*. Therefore, monotonicity must be true.

It might seem unimportant that monotonicity follows from these influential semantic views. For what do the workings of our language have to do with which outcomes are better or worse? The semantic facts seem to me important because the spectrum arguments purport to show that the relation expressed by the phrase *better than* is not transitive. Whether or not this is true depends, at least in part, on the meaning of that expression, which is just one of many relevantly similar expressions in our language. Of course, cyclists might reject the semantic views to which I have appealed (as Temkin seems to do—I discuss his argument in section 2.2). But then they owe us an alternative picture of the semantics of *better than*. For their arguments use this expression, and we should not be convinced by these arguments if we do not know what their premises mean.

Cyclists might claim not to care which expression we use in their arguments. In response to the charge that transitivity is a logical truth about comparatives (Broome 2004), Temkin claims that although there is a sense of *better than* which is transitive as a matter of logic or of language, there is also a sense which refers to an important normative relation that need not be transitive (Temkin 2012, 13). And it may not matter to the cyclists which words we use to express that relation. In section 2.3, I argue that Temkin's own account of this important normative relation entails the goodness and badness principles. But, before that, I consider an argument that Temkin presents against the more general principle of monotonicity.

2.2. Temkin on the Monotonicity of *Hairy*

Temkin (2012, 288–290) grants that some adjectives (e.g., *tall*) are monotonic—i.e., that if x is tall and y is taller than x , then y is also tall. But he seems to deny the monotonicity of *hairy*. I shall summarize Temkin's argument and then explain why I find it unconvincing. I shall then, in section 2.3, argue that even if monotonicity fails for some gradable adjectives, we have independent reason to accept the goodness and badness principles.

Temkin seems to reject the monotonicity of *hairy* based on a spectrum argument from hairiness to baldness (originally due to Wasserman 2005). Beginning with a head covered evenly by a large number of hairs, and which therefore seems hairy, he imagines making the distribution slightly less even while adding significantly more hairs. This change would seem to make the head hairier. But a sequence of many such changes would seem to lead to a head with a pronounced bald spot of (say) six inches in diameter, which Temkin imagines would make the head bald, however many hairs it contains. So we seem to have a sequence of heads h_1, \dots, h_n such that each head is hairier than its predecessor, and in which h_1 is hairy and h_n seems not only less hairy than h_1 (in violation of the transitivity of *hairier than*), but also bald. Iterations of the monotonicity of *hairy*, however, yield that h_n must be hairy, which seems impossible if h_n is bald. Temkin seems to suggest that we should, therefore, reject the monotonicity of *hairy*.

Temkin motivates this move as follows:

The key to recognizing this possibility is to recognize that while our criteria for being “hairy” and our criteria for being “hairier than” are intimately related to each other, we actually have *different* criteria for the two notions. So, while we might be willing to countenance the possibility that even a tiny difference in the pattern of hair distribution can make *all* the difference in terms of whether someone counts as hairy or not, we *don't* believe that such a tiny difference *alone* can settle the question of whether one head is hairier than another. (2012, 289, emphasis his)

The violation of monotonicity is possible, Temkin suggests, only because hairiness depends on multiple dimensions of variation—i.e., both distribution and number. The multidimensionality of *hairy* distinguishes it from (say) *tall*, which Temkin agrees is monotonic. Because *good* and *bad* are plausibly multidimensional—the goodness of pleasure and badness of pain, for example, depend on both intensity

and duration—Temkin’s view might predict that *good* and *bad* will fail to be monotonic as well.

I find Temkin’s argument unconvincing. For I see no reason why the criteria for *hairy* should differ so wildly from the criteria for *hairier than*. Temkin’s reasoning seems to be that certain properties—e.g., having a bald spot of some size—make it the case that one’s head is not hairy, no matter what other hairiness-relevant properties one has, but that one’s head could have such properties while being hairier than a head that lacks them, because other factors—e.g., quantity of hair overall—are relevant to the comparative *hairier than*. I find this very hard to believe. Those other factors should *also* be relevant to whether or not one’s head is hairy.

It is incredible that there should be a sharp threshold between heads that are hairy and heads that are not hairy. There are, however, powerful arguments for that view (Williamson 1994). But it seems to me *absurd* to think that (i) making one’s pattern of hair distribution less even to some (perhaps barely noticeable) degree makes it the case that one is no longer hairy, no matter how much more hair one has as a result, even though (ii) despite that difference, one would nonetheless be *hairier* if the number of increased hairs were sufficiently large. And that is how Temkin seems to be thinking of the hairy head that is less hairy than the not-hairy head. We could, of course, have stipulative uses of *hairy* that allowed for such cases: e.g., if we count as hairy, for some practical purpose, anyone who has at least n hairs and has no bald spot with a radius exceeding r . But then we would not be using the sense of *hairy* whose comparative form is our ordinary, multidimensional *hairier than*—much like the mathematical sense of *odd* is not the positive form of our ordinary, multidimensional *odder than*.

I am, therefore, unconvinced by Temkin’s argument for why monotonicity might fail for *hairy*. But even if some gradable adjectives are not monotonic, as Temkin claims of *hairy*, there is nonetheless strong reason to accept the goodness and badness principles.

2.3. *Reasons, Desires, and Preferences*

Following Parfit (2011), when we call something

good, in ... the *reason-implying* sense, we mean roughly that there are certain kinds of fact about this thing’s nature, or properties, that would in certain situations give us or others strong reasons to respond to this thing in some positive way, such as wanting, choosing, using, producing, or preserving this thing. (38)

Parfit understands *better than* similarly. On Parfit’s view, an outcome is (all things considered) better than some alternative in the reason-implying sense just in case we have stronger reasons to want this outcome to obtain, or to hope that it will obtain, than we have regarding its alternative (41).

I do not insist that we *analyze* or *reduce* goodness and badness to reasons to want or to prefer. But it seems to me that there is a true biconditional in the vicinity: that an outcome x is good if and only if there is strong reason, or one ought

(Ewing 1939), or it is fitting (Broad 1933) to desire x , and that something similar holds for betterness.

The reason-implying nature of goodness, badness, and betterness suggests an argument for the goodness and badness principles. Suppose that some outcome x , which we can bring about, is good. We have strong reasons to want x to obtain. Suppose too that some alternative y is better than x . Then our reasons to want y to obtain are even stronger than our reasons to want x to obtain. But if the goodness principle is false, then y might not be good. So we might not have strong reasons to want y to obtain. But that seems impossible, if we have strong reasons to want x to obtain and even stronger reasons to want y to obtain.¹⁴

Cyclists might reject this argument for the goodness and badness principles. They might argue that I have simply assumed monotonicity for the strength of reasons. Why else should we expect that if we have strong reasons to want x to obtain and even stronger reasons to want y to obtain, then we have strong reasons to want y to obtain? They might also reject Parfit's specification of the reason-implying sense of *better than*. So let us instead consider what Temkin claims to be the relevant sense of *better than*. Temkin (2012, 9) suggests that y is better than x just in case one would have most reason to prefer y to x . And, we can assume, if one has most reason to ϕ , then one ought to ϕ (Parfit 2011, 33).

Temkin's reason-implying sense of *better than* still supports the goodness principle. If Temkin's suggestion is correct, and if we reject the goodness principle, then it may be true that we ought to want x to obtain and that we ought to prefer y to x , and yet false that we ought to want y to obtain. But that combination of attitudes seems to me not just irrational, but impossible. I can see how someone might have nontransitive preferences; there is decent empirical evidence that we often do have such preferences. But I cannot see how any agent—let alone a rational one—could desire x , prefer y to x , and yet not desire y . That would be like desiring x , not desiring $\neg x$, and yet preferring $\neg x$ to x .¹⁵ That isn't just a combination that one ought not to have, but probably a combination that one *cannot* have.¹⁶

Why might this combination of attitudes be impossible? Intuitively, it's because our desires and preferences are not entirely independent. One set of these attitudes seems to supervene on the other. Many decision theorists, for example, assume that degrees of desire are derived from a set of preferences satisfying certain constraints. Alternatively, some philosophers (e.g., Pollock 2006) take a monadic, desire-like state as primitive, and understand preferences in terms of cardinally measurable desires. On views of either kind, our preferences and desires don't float freely of each other. Our conative lives would be beyond understanding if they did.

2.4. Summary

I have not given a decisive argument for the goodness and badness principles. But I have pointed to several considerations that support these principles. First, I observed that they are instances of a more general principle entailed by the two most influential views about the semantics of the positive and comparative forms of gradable adjectives. Second, I briefly considered Temkin's argument for why multidimensional predicates can violate monotonicity, and have found it wanting.

Temkin's view would require an implausibly sharp borderline for the positive form while maintaining tradeoffs between dimensions for the comparative form. That seems to me an unstable combination of views. And third, I claimed that violations of the goodness and badness principles would require agents to have seemingly impossible—and certainly irrational—combinations of attitudes, given reason-implicating conceptions of the relevant notions. I acknowledge that none of these considerations is, on its own, decisive. But together, and given the independent plausibility of the goodness and badness principles, these considerations seem to me to militate strongly against rejecting the principles.

Let me mention one final point: rejecting the goodness and badness principles might not be enough to avoid the implausible implications raised in section 1. Although I have framed the discussion in terms of the goodness and badness principles, as well as the more general principle of monotonicity, we can also ask of each outcome along each spectrum whether it is good (or bad) if its predecessor is. Given our more general principles, this conditional claim is entailed by the cyclists' comparative judgment that each outcome is better (or worse) than its predecessor. But the conditional claim would generate a problem even without our general principles: in the combined spectrum at the end of section 1, for example, it implies that if *A* is good then so is *Z*. (I discuss this further in section 4.) And, even without these general principles, the conditional claim seems no less independently plausible than the cyclists' comparative judgment. The kinds of considerations that support their comparative judgment would also seem to support the conditional claim that each outcome is good (or bad) if its predecessor is. So if the cyclists wish to escape the problems of section 1 by rejecting the goodness and badness principles, they would still have to explain why, although they deny that each outcome is good (or bad) if its predecessor is, they are convinced that each outcome is better (or worse) than its predecessor. And although that conditional claim has unwanted implications, I can see no principled rationale for rejecting it while accepting the cyclists' comparative judgment. So, ultimately, I suspect that rejecting the goodness and badness principles is a non-starter.

3. Rejecting the Good

The remaining possibility is that, on the cyclists' view, nothing is good or bad. Now, this seems to me to be a *reductio* of their view. The cyclists' judgments that lead to this conclusion are far less compelling than that the outcomes in the good spectrum are good, and that the outcomes in the bad spectrum are bad. But there may be various ways in which this conclusion could be softened or clarified to be less implausible.

3.1. *Intrinsic Value*

One way is to qualify the kind of goodness and badness to be rejected as *intrinsic* goodness and badness—i.e., the goodness or badness that something has in virtue of its intrinsic properties. This notion might seem to be challenged by the original Rachels–Temkin spectrum arguments, even before considering my variations on

them. Temkin, for example, claims that transitivity is plausible on a certain kind of picture about value—what he calls the *internal aspects view*—on which the value of an outcome is entirely a function of “internal” (i.e., intrinsic) features of that outcome, and which of two outcomes is better is entirely a function of the value of each outcome (so conceived). Temkin suggests that one can (more) plausibly reject transitivity by rejecting this kind of picture. If that is correct, then wouldn’t the cyclists already reject that outcomes can be intrinsically good or bad? In that case, we would learn little about their commitments from my combined spectrum arguments.

But this reply would miss the mark, for three reasons.

First, rejecting an internal aspects view would not commit the cyclists to the view that nothing is intrinsically good. To see this, consider the view that Temkin proposes in opposition to the internal aspects view. According to what he calls

The Essentially Comparative View:

There is some outcome such that either (i) there is no answer to the question of how good this outcome is, considered just by itself, or (ii) this answer has no special significance or relevance to how good or bad this outcome is in comparison with other outcomes.¹⁷

And this view falls far short of the claim that nothing is intrinsically good, in multiple ways: (a) the essentially comparative view is a merely existential generalization; (b) it could be true that every outcome is either good, bad, or neutral even if there is no answer to the question of *how* good or bad it is; and (c) it is compatible with the view that every outcome has some degree of intrinsic value or disvalue, but that betterness is not entirely a function of intrinsic value (e.g., because the goodness and badness principles fail). It would, therefore, be hasty to conclude that the cyclists are already committed to the view that nothing is intrinsically good or bad.

Second, in stating my modified spectrum arguments, I did not appeal to the quasi-technical notion of *intrinsic* value. The goodness and badness principles, for example, state no such restriction. To see one reason why this is important, consider the view that intrinsic value is extensionally distinct from noninstrumental value, or the value that something has as an end, or for its own sake (Smith 1948; Korsgaard 1983; Rabinowicz and Rønnow-Rasmussen 2003). If we could devise spectrum arguments involving bearers of extrinsic noninstrumental value—e.g., works of art that derive some of their value from historical properties—then the kind of value that the cyclists would reject need not be intrinsic.

Third, if rejecting the quasi-technical notion of intrinsic value, or any other circumscribed notion of value, is supposed to make the cyclists’ view any more plausible than the hardline view that nothing is good or bad, they need to explain the sense in which, on their view, it *can* be kosher to say that things are good or bad, and why things that are intuitively good or bad are indeed good or bad in that kosher sense. Absent some such account, the view remains absurd, for some things (e.g., long durations of pain, or of pleasure) are obviously good or bad in an important sense.

For an example of how one might provide such an account, consider what I said in section 2.1 about the importance of context for the positive form of gradable adjectives, like *tall*. Even if heights—i.e., how tall one is—are intrinsic, one cannot be tall *simpliciter*, because whether one counts as tall in some context depends on whether one is tall relative to some contextually salient comparison class—e.g., the average height in one’s kindergarten class or in the NBA. (It is no threat to the monotonicity of *tall* that a short NBA player can be taller than a tall kindergartner, because the positive form of the adjective must be understood relative to the same context when applying monotonicity.) But obviously there are important senses in which some people are tall. In denying that anything is tall *simpliciter*, we can point to a clear sense in which things can nonetheless be tall—namely, by being taller than some contextually relevant standard. The question is whether the cyclists are in a position to do anything like that. I shall argue that they are not.

3.2. McTaggart on Good

Might *good* be relevantly like *tall*? On such a picture, whether some outcome counts as good or bad depends on some comparison class which varies with context, and that is why nothing is intrinsically good or bad. This may seem implausible, because it seems clear that undeserved suffering, for example, is bad in some context-independent sense. But McTaggart (1927) suggests a way around this implication. Let me explain McTaggart’s suggestion and then explain why he ultimately rejects the picture under consideration.

McTaggart (p. 410) suggests that “qualities” (e.g., pleasure and pain) can be good or bad, even if the things that possess or instantiate those qualities (e.g., lives, outcomes, and token experiences) could not be. These claims are compatible because “all that is meant by a quality being good or evil is that it makes any substance which possesses it better or worse than it would have been otherwise.” And McTaggart suggests that the goodness or badness of qualities is more important, for practical purposes, than the goodness or badness of the things which possess those qualities. For what matters, for practical purposes, is that we make our lives and the world better, rather than worse. We can know how to do that via knowledge of the qualities that make things better or worse. We do not need to know whether our lives, or the world, is good or bad, all things considered. It may, therefore, seem promising to think of *good* and *bad* as relevantly like *tall* and *short*.

McTaggart, however, rejects this kind of view.¹⁸ His reason is that it seems unable to make sense of the fact that the goodness and badness of outcomes and lives are related, in certain ways, to the desirability of their existing or obtaining. For example, if a life were filled only with undeserved suffering, then it would be undesirable, and would make things worse, if someone lived it. The best explanation for this claim would seem to appeal to the fact that the life is just plain *bad*.¹⁹ Just as bad qualities (e.g., pain) make their possessors worse than they would otherwise be, and are undesirable to possess, the existence of bad lives makes the world worse than it would otherwise be and is therefore undesirable. But if the life of suffering only counts as bad in some contexts and counts as good in others—e.g., a context in which our lives are even worse than it—then it would seem difficult to explain

why, in a context where it counts as very good, its existence would be undesirable and would make the world worse.

One way of avoiding McTaggart's problem would be to claim that some lives may count as good, and others as bad, in all contexts. But if something which counts as good (bad) in all contexts is worse (better) than some element of a combined spectrum, then this strategy is not available to the cyclists. For then the elements of the combined spectrum will all count as good (bad) in all contexts, given the badness principle and the cyclists' cyclical judgments. That would be an unwelcome result.

Even if we are not persuaded by McTaggart's objection, I doubt whether the appeal to different contextual standards for *good* and *bad* can accommodate our intuitions about goodness and badness in the spectrum arguments. Consider the combined spectrum, which takes us from (A) 3×2^{25} years of mild pleasure and 1 year of agony to (Z) 3 years of ecstasy and 2^{25} years of mild pain. Each outcome is supposed to be worse than its predecessor, but A seems worse than Z. We can ask, of each outcome, whether it is good relative to some *fixed* contextual standard. For example, let the standard be an outcome in which there are no sentient beings. In this context, an outcome is good just in case it is better than that outcome, and bad just in case it is worse. Intuitively, the spectrum contains some outcomes that are good *in that context* and bad *in that context*. And, intuitively, if x counts as good (bad) within some context and y is better (worse) than x , then y must also count as good (bad) within that context. What this shows is that, given the cyclists' judgment that each outcome is worse than its predecessor, we are left with inconsistent intuitions even when our judgments about good and bad are restricted to a single context. And so the rejection of context-independent goodness and badness, on the model of *tall*, does not save our intuitive judgments about the cases under consideration.

3.3. Reference-Dependent Betterness

The cyclists, therefore, need a more radical picture than the one on which *good* and *bad* are like *tall*. Consider a different analogy. According to the *A-theory* of time, there is a meaningful and objective partition of times into past, present, and future. *B-theorists* deny this. But there are two ways of being a B-theorist. One kind of B-theorist thinks that although there is no privileged present, past, or future, the relations *earlier than* and *later than* are not similarly suspect. Just like there is a context-independent *taller than* ordering but no context-independent sense of *tall*, this kind of B-theorist thinks that there is an objective *earlier than* relation but no objective property of being in the future or the past. But other philosophers are B-theorists for reasons having to do with special relativity. According to the special theory of relativity, there is no such thing as *absolute simultaneity*; there is only simultaneity relative to a frame of reference. This makes it hard to be an A-theorist, because an objective present time would seem to require absolute simultaneity, or at least some privileged frame of reference relative to which simultaneity amounts to being present. If we think that there is no such privileged frame of reference, our view is more radical than what we said about *tall*. For although whether or not

one is tall depends on the contextually salient standard for tallness, the ordering imposed by the relation *taller than* does not depend on context. By contrast, which of two events precedes the other, and not just whether or not some event is in the future, depends on our frame of reference.

We can perhaps understand value along the lines of relative simultaneity. In other work (Nebel 2015), I propose a *reference-dependent* notion of value, inspired by Tversky and Kahneman (1991)'s notion of reference-dependent preferences.²⁰ Instead of a single betterness relation over all outcomes, there are many betterness relations, each indexed to some privileged reference point. These reference points may be the status quo, some anticipated outcome, or some other standard (e.g., par in golf). Economists have used the idea of reference-dependent preferences to explain a wide array of empirical phenomena; we might similarly hope that the idea of reference-dependent betterness can explain some important normative phenomena.

To get a feel for reference-dependent betterness, consider what Adams (1979) says about our lives' "retrospective preferability." Adams supposes that if Helen Keller had not been deafblind, then she would have led a happier life. That may have given us reason to hope for her sake (before she became deafblind) that her sight and hearing would be spared. But that life would have been extremely different from her actual life. As Adams puts it,

That other, happier life would have contained few of the particular joys and sorrows, trials and triumphs—in short very little of the concrete content—that she cared about in her actual life. Her never having been blind or deaf would have been very like her never having existed. (60)

So, given what Keller's life was actually like, it is reasonable not to wish, for her sake, that her sight and hearing had been spared.

We might understand this in terms of reference-dependent betterness.²¹ If Keller hadn't been deafblind, that would have been better for her only from certain frames of reference—e.g., before she acquired the projects and values that guided her actual life, or in a state of affairs where Keller isn't deafblind.²² But, from the reference point of the actual state of affairs, or of the world since she lost her sight and hearing, it would have been worse (or, at least, not better) for her if she hadn't been deafblind.

The idea of reference-dependent betterness may seem to promise a way of reconciling our intuitive judgments about the goodness and badness of the outcomes in the spectrum arguments with the monotonicity of goodness and badness. The key idea is that our reference point shifts as we move along the spectrum. Consider the combined spectrum from (*A*) 3×2^{25} years of mild pleasure and 1 year of agony to (*Z*) 3 years of ecstasy and 2^{25} years of mild pain. Suppose we claim only that *Z* is better than *A* from reference points *Z* and *A*, that each outcome is worse than its predecessor relative to its (and its predecessor's) reference point, and that some (but not all) of the outcomes are good (or bad) from some reference point or other. We can then maintain that if *x* is good (relative to some reference point *r*) and *y* is better than *x* (relative to *r*), then *y* must also be good (relative to *r*). If our intuitive

judgments about which outcomes are good or bad, and which are better and worse, are made relative to different reference points as we move along the spectrum, then these judgments need not conflict with the goodness and badness principles. That is because these principles hold only given a fixed reference point: if x is good relative to reference point r_1 , and y is better than x relative to some other reference point r_2 , it need not be true that y is good relative to r_1 or r_2 ; for x might be bad relative to r_2 , or better than y relative to r_1 .

We might object that, intuitively, our judgments about the outcomes along the spectrum are reference-independent, in that they do not depend on which outcome obtains or is the standard of reference. But, as with simultaneity and futurity, proponents of reference-dependent betterness might hold that there is no “absolute”—i.e., reference-independent—betterness relation or goodness property to be instantiated.²³

The notion of reference-dependent betterness may also be able to avoid McTaggart’s problem. McTaggart’s problem was that unless certain lives or outcomes are just plain bad, we cannot plausibly explain why their existence makes the world worse, and is therefore undesirable. I mentioned that this claim could be avoided if some lives counted as bad in every context. The cyclists cannot make this move if betterness is reference-independent, because such lives seem better than some elements of some combined spectrum. Their badness would then infect all the elements of the betterness cycle, including ones that are intuitively not bad. But if we can appeal to reference-dependent betterness, then we might claim that although the life is bad from every reference point, no element of a combined spectrum is such that, from every reference point, it is *worse* than this life. So no element of a combined spectrum needs to count as bad from every reference point.

But there was another problem with the view that *good* is like *tall*, and reference-dependent betterness seems not to avoid this problem. The problem was that, even within a fixed context, some elements of the combined spectrum seem good and others seem bad. Suppose, for example, that our reference point is not one of the outcomes along the spectrum, but instead some neutral framing. We might be deciding whether to bring someone into existence, and the outcomes along the combined spectrum represent different possible lives that she might lead. Intuitively, some of the lives (e.g., Z) are good relative to this neutral reference point, and others (e.g., A) are bad relative to this same reference point. Which of the lives, if any, we have most reason to choose depends on the betterness relation tied to our neutral reference point. And each outcome still *seems* better than its successor, not just relative to the reference point of each outcome and its successor, but from this neutral point of view as well.²⁴ Since Z also seems better than A from this reference point, our neutral reference-dependent betterness relation seems to contain a cycle, which (given monotonicity) would rule out A being bad and Z being good relative to this neutral reference point.

My claim is not that there *is* a cycle here, but rather that the idea of reference-dependent betterness does not explain away the *appearance* of a cycle from this neutral reference point. So it does not capture our judgments about goodness and badness in the combined spectrum, unless we are willing to reject the intuition that

each outcome is better than its successor from our neutral reference point. But if we are willing to reject this intuition about our neutral reference point's betterness relation, why shouldn't we reject it when framed in reference-independent terms, as in the initial spectrum arguments? The intuition seems to me no less compelling when couched in terms of a neutral reference-dependent betterness relation than when couched in terms of plain old betterness. But the purpose of reference-dependent betterness in the present context was to make the radical conclusion of the combined spectrum arguments—that nothing is good or bad—seem less implausible. If, however, we are willing to reject a premise of these arguments, then we need not try to make their radical conclusion seem less implausible. We can simply reject the conclusion.

I, therefore, believe that the notion of reference-dependent betterness does not preserve our intuitive judgments about the combined spectrum. Perhaps there is some other way in which cyclists might claim that some of the outcomes along the spectrum are good or bad in some sense that (a) is compatible with the goodness and badness principles and (b) maintains some semblance of our intuitive judgments about the outcomes. But I do not know what that way would be.

4. How (Not) to Reject the Spectrum Arguments

This leaves us with the conclusion that, on the cyclists' view, nothing is good or bad.²⁵ That is absurd. We should, therefore, reject the spectrum arguments against transitivity.

But *how* should we reject these arguments? Which of their assumptions should we reject, and how can we dispel their apparent compellingness?²⁶

I wish to conclude by considering two extant responses to the spectrum arguments. In light of the modifications I introduced in section 1, one of these responses now seems less promising, and the other more, than they might have otherwise seemed.

4.1. Incommensurability

One response to the spectrum arguments appeals to *incommensurability*. Some philosophers claim that some of the elements in each spectrum are incommensurate in value with their predecessors, in the sense that they are neither better than, nor worse than, nor exactly as good as their predecessors (Parfit 2012; Handfield 2014; Chang 2016). These experiences, therefore, do not give rise to a betterness cycle.

The relevant kind of incommensurability would seem not to be *incomparability*. For each element along the spectrum is quite similar to its predecessor, whereas the most plausible examples of incomparability involve radically different kinds of things—e.g., Shakespeare and $\sqrt{2}$ —which do not bear the same kind of value. The incommensurate elements might instead be held to be *on a par* (Chang 2002).²⁷

This response seems to me unpromising in the context of my combined spectrum. For we can plausibly strengthen the goodness and badness principles as follows:

The Strengthened Goodness Principle:

For any comparable value bearers x and y , if x is good and y is not, then x is better than y .

The Strengthened Badness Principle:

For any comparable value bearers x and y , if x is bad and y is not, then x is worse than y .

These principles entail the goodness and badness principles.²⁸

The strengthened principles can be motivated by considering a thing's value on the assumption that it isn't good. Of course, some things aren't good because they don't bear any value whatsoever—e.g., $\sqrt{2}$. But our concern is with things that bear some relevant kind of value, and are comparable to things that are good or bad. If some such thing isn't good, then it is presumably either neutral or bad. Maybe it can be indeterminate which of these it is. But, on either assumption, it is not only *not better* than anything good (as the goodness principle claims), but also *worse*. It is hard to see how something good could fail to be better than something neutral or bad. This is true not only for the goodness of lives, outcomes, and experiences, but also for "attributive" goodness—e.g., the goodness of artists. If one artist is good and another is not, how could the one fail to be a better artist than the other?²⁹

The strengthened principles, if true, make it futile to appeal to parity in the combined spectrum. For if A is bad, and each element is either worse than or on a par with its predecessor, then B must also be bad, according to the strengthened badness principle. For if B were not bad, then it would be better than A . So B is bad. Iterations of this reasoning yield that Z is bad. But Z seems good, not bad. The appeal to parity, therefore, does not strike me as a promising response to the spectrum arguments, in light of my combined variations.

4.2. Vagueness

A different response appeals to vagueness. Like most expressions in our language, *better than* is vague. And many people claim that the spectrum arguments are relevantly like sorites arguments, which exploit intuitions about vague predicates in order to generate absurd conclusions.³⁰ But the cyclists claim that the arguments are disanalogous.³¹

It is not obvious how exactly the spectrum arguments are supposed to be soritical. Consider, for example, the following sorites argument involving the predicate *rich*:

Someone with one dollar isn't rich. For any n , if someone with n dollars isn't rich, then someone with $n + 1$ dollars isn't rich. So a billionaire isn't rich.

But obviously a billionaire is rich. We know this argument to be unsound. The spectrum arguments, though, do not seem to have this simple structure, in which some monadic predicate is claimed to be tolerant to a certain change in respects relevant to its application.³²

However, the cyclists' judgments about the *combined* spectrum do give rise to an argument with this structure. The cyclists claim that each element is worse than

its predecessor, and so (by the badness principle) that each element is bad if its predecessor is. This conditional gives rise to a sorites:

Z (3 years of ecstasy and 2^{25} years of mild discomfort) is not bad. And, for any n and m , if n years of pleasure and m years of pain are not bad, then neither are $2n$ years of slightly less intense pleasure and $\frac{m}{2}$ years of slightly more intense pain. So A (3×2^{25} years of mild pleasure and 1 year of agony) is not bad.

But, intuitively, A is bad. In this argument, the monadic predicate *bad* is claimed to be tolerant to a certain change in respects relevant to its application. This leads us to a counterintuitive conclusion.

Of course, there are differences between the two arguments. One difference that might seem relevant lies in the motivation for the conditional premise. It seems that if someone with n dollars isn't rich, then someone with $n + 1$ dollars isn't rich either, *despite* the fact that the latter is *richer* than the former. But, in the combined spectrum, it seems that if some element isn't bad, then neither is its predecessor, *because* each element seems *worse* than its predecessor. The conditional seems true because, not in spite, of the comparative claim, which goes in the direction opposite that of its analogue for *rich*.

But this difference is not relevant. There are many kinds of vague predicates and many kinds of sorites arguments. Unlike *rich* and *richer*, which (we can suppose) depend only on a person's net worth, *bad* and *worse* are multidimensional. For example, the badness of a painful experience depends on its duration as well as its intensity. Whereas the sorites argument for *rich* appeals to the intuition that a marginal difference in net worth cannot make a difference to whether or not one is rich, the sorites argument for *bad* appeals to a different (but related) intuition about multidimensionally vague predicates. The intuition is that a marginal difference along some dimensions (e.g., intensity of pleasure or pain) cannot make a difference to whether or not an outcome is bad (or worse than some alternative), when there are large differences along other relevant dimensions (e.g., duration of pleasure and pain). It seems that such marginal differences should not be decisive, and that the large differences should outweigh the small differences. That is why it seems that every element should be worse than its predecessor, and therefore bad if its predecessor is. This leads to a counterintuitive conclusion because the small differences can add up to differences so great that we are inclined to think they are decisive.

This seems to be a very general feature of multidimensionally vague predicates, as we saw with *hair* in section 2.2.³³ To take another example, health is a function of several dimensions. I might be just barely healthy in each of these respects—e.g., because my blood pressure, pulse, blood sugar levels, and other relevant factors are just slightly within the normal range—so that I am nonetheless healthy overall. A small difference along one of these dimensions (say, blood sugar) seems like it should not be decisive, with respect to whether or not I am healthy, when there is a large difference along some other dimension. A slightly higher level of blood sugar should not make me no longer healthy if it trades off with a significant improvement with respect to some other factor—e.g., becoming perfectly healthy in one of the other respects. But such tradeoffs add up. Even if I am perfectly healthy in every

other respect I should no longer count as healthy if my blood sugar levels are far outside the normal range.³⁴ Like the combined spectrum argument, the problem stems from a kind of tolerance which is characteristic of multidimensionally vague expressions.

So how should we reject the seemingly soritical argument from the combined spectrum? Perhaps by rejecting the conditional premise, which different solutions to the sorites paradox do in different ways. On one view, there might be some particular element that isn't bad even though its predecessor is, although it is unknowable which element that is (Williamson 1994). On other views, even if the conditional premise is false, it need not be true of any particular element that it isn't bad even though its predecessor is (Fine 1975). On one view of this kind, each of the conditional's instances might even be *true* (Kamp 1981). These strategies would also apply to the cyclists' initial claim that each element is worse than its predecessor. And, however we reject that claim in the context of the combined spectrum, we could reject it in the same way in the context of the original spectrum arguments. So, even though the original spectrum arguments seem to lack the form of a sorites argument, they rest on a claim that seems to give rise to sorites arguments, which we ought (somehow) to reject.

I do not insist that any solution to the sorites paradox will straightforwardly explain what is wrong with the spectrum arguments. Nor do I insist that there are no relevant differences between the spectrum arguments and the sorites paradox. I remain open to the possibility that, for any duration of agony (or ecstasy), there is some duration of mild discomfort (or mild pleasure) that would be worse (or better). But I find that possibility hard to accept. So I hope that the reasoning which leads to it can be rejected in some familiar way; my variations on the spectrum arguments make me somewhat more confident that it can.

Here is what we have learned from my variations on the spectrum arguments. The cyclists must reject the monotonicity of *good* and *bad* or else deny that anything is good or bad. Neither option is attractive. So we must reject the spectrum arguments against transitivity. Perhaps we can reject them by solving the sorites paradox.

Notes

¹ In what follows, I generally omit "all things considered" and speak only of "transitivity," but this should always be understood as the transitivity of the *all-things-considered better than* relation.

² This would require some *A*, *B*, and *C* such that *A* is better than *B*, which is better than *C*, and yet *C* is neither better nor worse than *A*. This could be because *A* and *C* are equally good. An anonymous referee suggests that a cycle could then inevitably be constructed by making *C* very slightly better (call it *C+*). But an ensuing cycle is guaranteed only on the assumption that if *C+* is better than *C*, which is just as good as *A*, then *C+* must be better than *A*. This assumption is entailed by the transitivity of *at least as good as* (Sen 1970, 10–11). But opponents of the transitivity of *better than* might well also reject the transitivity of *at least as good as*. So they could, in principle, maintain that *better than* is acyclic.

A more interesting way of rejecting transitivity while maintaining acyclicity would be to hold that if *A* is better than *B*, which is better than *C*, then *C* may not be worse than *A*, although it cannot be at least as good as *A*—i.e., *A* and *C* may be *incommensurate* in value. This combination of claims may be of interest for two reasons. First, it is compatible with the condition of *Suzumura consistency* (Suzumura 1976), which prohibits weak betterness cycles containing any instances of strict betterness: that is, for

any set of x_1, \dots, x_j such that each x_i is at least as good as x_{i+1} and x_n is at least as good as x_1 , there is no k and j such that x_k is better than x_j . This condition is stronger than acyclicity, neither implies nor is implied by the transitivity of *better than*, and is both necessary and sufficient for the absence of money pumps (Bossert and Suzumura 2010). Second, some might be inclined to make such judgments about Parfit's (1984) mere addition paradox. Many people are inclined to judge that some population (A) of excellent lives is better than any population (Z) of mediocre lives; that any perfectly equal population (such as Z) is better than any same-sized population ($A+$) of lives with lower average and (therefore) total wellbeing; and that if $A+$ differs from A only via the addition of lives within some range, then A and $A+$ are incommensurate.

³ One might take the lesson of the Rachels–Temkin spectrum arguments to be that nothing is bad *simpliciter*. I discuss this possibility in section 3. If you are tempted by it, you can read my claims about badness here in the conditional spirit just stated, as what would be true if anything is bad. Note also that even philosophers who find it unintelligible to talk of things as being just plain good or bad (e.g., Thomson 2001) might accept that the experiences under discussion are good or bad in some intelligible and normatively relevant sense. Thomson, for example, accepts that things can be good or bad *for you* (p. 13), which is surely true of the experiences in the spectrum arguments.

⁴ Temkin (2012, 519) attributes this worry to Derek Parfit. Temkin (2014, 471) notes the connection between Parfit's worry and the argument of this paper (cited there as Nebel 2012).

⁵ I do not suppose that this kind of balancing is precise, as if for every quantity of bad, there is exactly one quantity of good such that combining those quantities is neutral. Perhaps there is a neutral range with vague upper and lower bounds, although see Broome (2004) for objections to this idea; see Rabinowicz (2009a; 2009b) for responses to Broome's objections. If there is just a single neutral level of value—i.e., a “zero” level—then my arguments below could take an even simpler form.

⁶ Of course, this heuristic raises difficult questions in the ethics of population. But it's just a way to get you to think about the balance between pleasure and pain in a third-personal way.

⁷ This is also observed by Carlson (2013). Danielsson (1996) makes a similar point in terms of choiceworthiness.

⁸ Strictly speaking, one could reject (1) and (2) while maintaining their analogues in the original spectrum arguments. But that position, I think, has no plausibility.

⁹ It may be hard to say that A is *precisely* neutral, because neutrality may be vague, as Broome (2004) argues. But surely there is some version of A that is determinately *not bad*, although it may be vague whether it is merely neutral or good.

¹⁰ There are issues, of course, having to do with the length of time in Z . It may therefore help to imagine that these outcomes are experienced by creatures that don't get bored.

¹¹ My quantifier here is restricted to things like outcomes, lives, experiences, states of affairs, or whatever other kinds of things give rise to spectrum arguments.

¹² See, e.g., Barker (2002), Burnett (2014), Fine (1975), Fara (2000), and Williamson (1994). Interestingly, monotonicity has received little attention in the theory of value. See, however, Carlson (2011; 2013; 2016) and Danielsson (1996). Another exception may be Aristotle, whom Sahlin (1993, 185) quotes as seeing “nothing to prevent even things which are not good from being preferable to things which are good” (*Topics*, III.2 117a20–22).

¹³ See, e.g., Heim (2000) and especially Nouwen (2011). Note that the claim that I am 6 feet tall, on this view, entails that I am also 5 feet tall. This may sound strange, because the sentence, “I am 5 feet tall,” seems to mean that I am exactly 5 feet tall. But this strangeness can be explained as a merely pragmatic effect, in terms of scalar implicatures. In other contexts—e.g., “You must be 5 feet tall to ride the Vominator”—the *at least* reading is clearly correct. (Thanks to Ben Holguin for the example.)

¹⁴ An anonymous referee suggests that, on the cyclists view, reasons might not have an absolute, context-independent strength. On this kind of view, whether or not we have most reason to desire some outcome might depend on the context, including the available alternatives. I would consider this intriguing possibility to be a variant of a view I discuss in section 3, on which outcomes are only good or bad, desirable or undesirable, relative to a context or frame of reference. I suspect that my objections there would apply here, as well.

¹⁵ What about doxastic analogues of these attitudes? An anonymous referee notes that if belief is just credence above a certain threshold, and if the threshold for a given proposition depends on the stakes

of acting on that proposition, then one can believe that p , be more certain of q , and yet not believe q . This is an interesting possibility. I suspect that beliefs and credences are not relevantly analogous to desires and preferences, respectively. Desiring is not preferring above a certain threshold, and whether one desires a proposition does not depend on stakes in a way that is independent of what one prefers. But I shall not try to defend these claims here.

¹⁶ This argument might seem to assume that if it's necessary that (if you ψ , then you ϕ), then we can infer that you ought to ϕ from the fact that you ought to ψ . And this inference sometimes seems bad (Broome 2013, 121). It seems bad when one doesn't ψ and yet ϕ ing has some disadvantage in worlds where one fails to ψ . But such cases for desiring and preferring would seem to involve the "wrong kind of reason" problem (see Rabinowicz and Rønnow-Rasmussen 2004), and the proponent of a fitting-attitudes analysis already needs a solution to that problem. So I don't think we should expect the inference to be bad in this instance.

¹⁷ This is a paraphrase of Temkin (2012, 371).

¹⁸ My interpretation of McTaggart's objection is indebted to Broad (1938).

¹⁹ Some may also claim that living such a life would be worse for the person who lives it than nonexistence (Roberts 2011; Arrhenius and Rabinowicz 2015; Fleurbaey and Voorhoeve 2015). Others would find this claim incoherent (Broome 2004). I need not take a stance on it here.

²⁰ In Nebel (2015), I propose reference-dependent value as a way of making sense of rational status quo bias as well as deliberation under an essentially comparative view of the good. Ideas similar to reference-dependent betterness have been suggested by Brennan and Hamlin (2016, "state-dependent value"), Frick (2014, "context-dependent betterness"), and Dreier (2016, "world-centered value").

²¹ Adams doesn't understand it this way. He supposes that Keller's life, had she not been deafblind, would have been better, but that her actual life is nonetheless preferable in retrospect, because the better life bears no self-interest relation to her actual life. But if we wish to maintain a strong connection between betterness and what we have reason to prefer, then the idea of reference-dependent betterness seems to me essential.

²² When we state an evaluation and tie it to some reference point, this may make it sound as if the evaluation is subjective—i.e., as if one outcome is better than another from some reference point just in case one who occupies that reference point would judge the one better than the other. But I intend the reference-dependent evaluation to be thoroughly objective and independent of our attitudes.

²³ Some might object that if betterness is reference-dependent, then it isn't *all-things-considered*. But all-things-considered betterness is betterness overall, as opposed to in some respect or along a certain dimension. Reference-dependent betterness relations may take into account betterness relations over multiple dimensions of value.

²⁴ Some proponents of reference-dependent betterness might claim that we are psychologically incapable of holding the reference point fixed—i.e., that we cannot help but shift the reference point when attending to different pairs along the spectrum. But this would seem to me a strange and unmotivated fact about our psychology, and without further evidence that it is true, I am inclined to believe that we are perfectly capable of considering the various outcomes from the same neutral frame of reference.

²⁵ Again, my quantifier here is restricted to the kinds of things that can be better or worse than the things that show up in the spectrum arguments.

²⁶ I am indebted to an anonymous referee for encouraging me to explore this question here.

²⁷ For an account of parity and other varieties of incommensurability in terms of fitting attitudes, see Rabinowicz (2008; 2012).

²⁸ Suppose, for *reductio*, that the strengthened goodness principle is true but that the non-strengthened version is false. Then there is some x and y such that (i) x is good, (ii) y is better than x , but (iii) y isn't good. x and y are comparable, given (ii). So (i) and (iii) entail that x is better than y , by the strengthened goodness principle. But that contradicts (ii), by the asymmetry of *better than*.

²⁹ Note that if one person wasn't an artist at all, then it would be easy to see how she wouldn't be a worse artist. But that would be a case of incomparability, due to noninstantiation of the relevant property.

³⁰ See Qizilbash (2005) and Wasserman (2005). Similar claims about similar arguments have been made by Ackerman (1994), Danielsson (1996), Arrhenius and Rabinowicz (2005), and Voorhoeve and Binmore (2006).

³¹ For the most comprehensive list of disanalogies, see Temkin (2012, sec. 9.2).

³² For the notion of tolerance, see Wright (1975).

³³ Grinsell (2013) argues that this feature is relevantly like continuity assumptions in the theory of social choice. He argues that the sorites paradox is just an instance of Chichilnisky (1982)'s impossibility theorem.

³⁴ See Sassoon (2013) for a theory of multidimensional gradability which can explain this judgment.

References

- Ackerman, Felicia. 1994. "Roots and Consequences of Vagueness." *Philosophical Perspectives* 8: 129. doi:10.2307/2214167.
- Adams, Robert Merrihew. 1979. "Existence, Self-Interest, and the Problem of Evil." *Noûs* 13 (1): 53. doi:10.2307/2214795.
- Arrhenius, Gustaf, and Wlodek Rabinowicz. 2005. "Value and Unacceptable Risk." *Economics and Philosophy* 21 (2): 177–97. doi:10.1017/S0266267105000556.
- . 2015. "The Value of Existence." In *The Oxford Handbook of Value Theory*, edited by Iwao Hirose and Jonas Olson. Oxford University Press.
- Barker, Chris. 2002. "The Dynamics of Vagueness." *Linguistics & Philosophy* 25 (1): 1–36.
- Bossert, Walter, and Kōtarō Suzumura. 2010. *Consistency, Choice, and Rationality*. Cambridge, Mass.: Harvard University Press.
- Brennan, Geoffrey, and Alan Hamlin. 2016. "Conservative Value." *The Monist* 99 (4).
- Broad, C. D. 1933. "Is 'Goodness' a Name of a Simple Non-Natural Quality?" *Proceedings of the Aristotelian Society, New Series*, 34 (January): 249–68. <http://www.jstor.org/stable/4544242>.
- . 1938. *An Examination of McTaggart's Philosophy*. Vol. II. Cambridge: Cambridge University Press.
- Broome, John. 2004. *Weighing Lives*. Oxford; New York: Oxford University Press.
- . 2013. *Rationality Through Reasoning*. Blackwell.
- Burnett, Heather. 2014. "Penumbra Connections in Comparative Constructions." *Journal of Applied Non-Classical Logics* 24 (1-2): 35–60. doi:10.1080/11663081.2014.911519.
- Carlson, Erik. 2011. "Defining Goodness and Badness in Terms of Betterness Without Negation." In *Descriptive and Normative Approaches to Human Behavior*, edited by Ehtibar N. Dzhafarov and Lacey Perry, 51–66. Advanced Series on Mathematical Psychology. World Scientific.
- . 2013. "Intransitivity." In *International Encyclopedia of Ethics*. Blackwell Publishing Ltd.
- . 2016. "'Good' in Terms of 'Better'." *Noûs* 50 (1): 213–23. doi:10.1111/nous.12061.
- Chang, Ruth. 2002. "The Possibility of Parity." *Ethics* 112 (4): 659–88. doi:10.1086/339673.
- . 2016. "Parity, Imprecise Comparability and the Repugnant Conclusion." *Theoria* 82 (2): 182–214. doi:10.1111/theo.12096.
- Chichilnisky, Graciela. 1982. "Social Aggregation Rules and Continuity." *The Quarterly Journal of Economics* 97 (2): 337–52. doi:10.2307/1880762.
- Danielsson, Sven. 1996. "The Refutation of Cyclic Evaluations." *Theoria* 62 (1-2): 161–68. doi:10.1111/j.1755-2567.1996.tb00535.x.
- Dreier, Jamie. 2016. "World-Centered Value." Unpublished manuscript.
- Ewing, A. C. 1939. "A Suggested Non-Naturalistic Analysis of Good." *Mind, New Series*, 48 (189): 1–22. <http://www.jstor.org/stable/2250596>.
- Fara, Delia Graff. 2000. "Shifting Sands: An Interest Relative Theory of Vagueness." *Philosophical Topics* 28 (1): 45–81. Originally published under the name "Delia Graff."
- Fine, Kit. 1975. "Vagueness, Truth and Logic." *Synthese* 30 (3-4): 265–300.
- Fleurbaey, Marc, and Alex Voorhoeve. 2015. "On the Social and Personal Value of Existence." In *Weighing and Reasoning: Themes from the Philosophy of John Broome*, edited by Iwao Hirose and Andrew Reisner. Oxford University Press.
- Frick, Johann David. 2014. "'Making People Happy, Not Making Happy People': A Defense of the Asymmetry Intuition in Population Ethics." Ph.D. Dissertation, Harvard University. <http://dash.harvard.edu/handle/1/13064981>.

- Grinsell, Timothy Wood. 2013. "Avoiding Predicate Whiplash: Social Choice Theory and Linguistic Vagueness." In *Proceedings of SALT*, 22:424–40. <http://elanguage.net/journals/salt/article/download/22.424/3479>.
- Gurney, Edmund. 1887. *Tertium Quid: Chapters on Various Disputed Questions*. <https://archive.org/details/tertiumquidchap00gurngoog>.
- Handfield, Toby. 2014. "Rational Choice and the Transitivity of Betterness." *Philosophy and Phenomenological Research* 89 (3): 584–604. doi:10.1111/phpr.12050.
- Heim, Irene. 2000. "Degree Operators and Scope." In *Semantics and Linguistic Theory*, 10:40–64. <http://journals.linguisticsociety.org/proceedings/index.php/SALT/article/viewFile/3102/2825>.
- Kamp, Hans. 1981. "The Paradox of the Heap." In *Aspects of Philosophical Logic*, edited by Uwe Mönnich, 225–77. Synthese Library 147. Springer Netherlands. http://link.springer.com/chapter/10.1007/978-94-009-8384-7_8.
- Kennedy, Christopher. 2004. "Comparatives, Semantics of." *Concise Encyclopedia of Philosophy of Language and Linguistics*, 68–71.
- Klein, Ewan. 1980. "A Semantics for Positive and Comparative Adjectives." *Linguistics and Philosophy* 4 (1): 1–45. <http://link.springer.com/article/10.1007/BF00351812>.
- Korsgaard, Christine M. 1983. "Two Distinctions in Goodness." *The Philosophical Review* 92 (2): 169. doi:10.2307/2184924.
- McTaggart, J. Ellis. 1927. "The Nature of Existence, Volume II." <https://ia600400.us.archive.org/29/items/natureofexistenc02mctauoft/natureofexistenc02mctauoft.pdf>.
- Nebel, Jacob M. 2012. "Crossing the Zero Level: Spectrum Arguments and Intrinsic Value." Junior Paper, Princeton University.
- . 2015. "Status Quo Bias, Rationality, and Conservatism About Value*." *Ethics* 125 (2): 449–76. doi:10.1086/678482.
- Nouwen, Rick. 2011. "Degree Modifiers and Monotonicity." In *Vagueness and Language Use*, 146–64. Springer. http://link.springer.com/chapter/10.1057/9780230299313_7.
- Parfit, Derek. 1984. *Reasons and Persons*. Oxford: Clarendon Press.
- . 2011. *On What Matters. V. I*. Oxford: Oxford University Press.
- . 2012. "Towards Theory X: Part One." Unpublished manuscript.
- Pollock, John L. 2006. *Thinking About Acting: Logical Foundations for Rational Decision Making*. Oxford ; New York: Oxford University Press.
- Qizilbash, Mozaffar. 2005. "Transitivity and Vagueness." *Economics and Philosophy* null (01): 109–31. doi:10.1017/S0266267104000410.
- Rabinowicz, Wlodek. 2008. "Value Relations." *Theoria* 74 (1): 18–49. doi:10.1111/j.1755-2567.2008.00008.x.
- . 2009a. "Broome and the Intuition of Neutrality." *Philosophical Issues* 19 (1): 389–411. <http://onlinelibrary.wiley.com/doi/10.1111/j.1533-6077.2009.00174.x/full>.
- . 2009b. "I—Incommensurability and Vagueness." In *Aristotelian Society Supplementary Volume*, 83:71–94. Wiley Online Library. <http://onlinelibrary.wiley.com/doi/10.1111/j.1467-8349.2009.00173.x/full>.
- . 2012. "Value Relations Revisited." *Economics and Philosophy* 28 (02): 133–64. doi:10.1017/S0266267112000144.
- Rabinowicz, Wlodek, and Toni Rønnow-Rasmussen. 2004. "The Strike of the Demon: On Fitting Pro-Attitudes and Value." *Ethics* 114 (3): 391–423. doi:10.1086/381694.
- . 2003. "Tropic of Value." *Philosophy and Phenomenological Research* 66 (2): 389–403. <http://www.jstor.org/stable/20140542>.
- Rachels, Stuart. 1993. "A Theory of Beneficence." Undergraduate Thesis, University of Oxford.
- . 1998. "Counterexamples to the Transitivity of Better Than." *Australasian Journal of Philosophy* 76 (1): 71–83. doi:10.1080/00048409812348201.
- . 2001. "A Set of Solutions to Parfit's Problems." *Nous* 35 (2): 214–38. doi:10.1111/0029-4624.00294.
- Roberts, Melinda A. 2011. "An Asymmetry in the Ethics of Procreation." *Philosophy Compass* 6 (11): 765–76. doi:10.1111/j.1747-9991.2011.00435.x.

- Sahlin, Nils-Eric. 1993. "Worthy of Choice." *Theoria* 59 (1-3): 178–91.
- Sassoon, Galit. 2013. "A Typology of Multidimensional Adjectives." *Journal of Semantics* 30 (3): 335–80. doi:10.1093/jos/ffs012.
- Sen, Amartya. 1970. *Collective Choice and Social Welfare*. San Francisco: Holden-Day San Francisco.
- Smith, James Ward. 1948. "Intrinsic and Extrinsic Good." *Ethics* 58 (3): 195–208.
- Suzumura, Kōtarō. 1976. "Remarks on the Theory of Collective Choice." *Economica* 43 (172): 381–90. doi:10.2307/2553273.
- Temkin, Larry S. 1987. "Intransitivity and the Mere Addition Paradox." *Philosophy & Public Affairs* 16 (2): 138–87.
- . 1996. "A Continuum Argument for Intransitivity." *Philosophy & Public Affairs* 25 (3): 175–210.
- . 2012. *Rethinking the Good: Moral Ideals and the Nature of Practical Reasoning*. Oxford Ethics Series. Oxford ; New York: Oxford University Press.
- . 2014. "Rethinking the Good: A Reply to My Critics." *Analysis* 74 (3): 439–88. doi:10.1093/analys/anu043.
- Thomson, Judith Jarvis. 2001. *Goodness & Advice*. The University Center for Human Values Series. Princeton, N.J: Princeton University Press.
- Tversky, Amos, and Daniel Kahneman. 1991. "Loss Aversion in Riskless Choice: A Reference-Dependent Model." *The Quarterly Journal of Economics* 106 (4): 1039–61. doi:10.2307/2937956.
- Voorhoeve, Alex, and Ken Binmore. 2006. "Transitivity, the Sorites Paradox, and Similarity-Based Decision-Making." *Erkenntnis* 64 (1): 101–14. doi:10.1007/s10670-005-2373-1.
- Wasserman, Ryan. 2005. "Paradoxes of Transitivity." Unpublished manuscript.
- Williamson, Timothy. 1994. *Vagueness*. London; New York: Routledge.
- Wright, Crispin. 1975. "On the Coherence of Vague Predicates." *Synthese* 30 (3/4): 325–65. <http://www.jstor.org/stable/20115035>.