

Trade-offs, Transitivity, and Temkin[†]

Dale Dorsey

Larry S. Temkin, *Rethinking the Good: Moral Ideals and the Nature of Practical Reasoning* (Oxford: Oxford University Press, 2012), 616 pages. ISBN: 0199759448 (hbk.). Hardback \$74.00.

The last few years have seen a number of new and important publications in moral theory. Arguably none more important than Larry S. Temkin's *Rethinking the Good: Moral Ideals and the Nature of Practical Reasoning*. This fascinating work highlights and explores in depth a seemingly fundamental inconsistency in our own axiological beliefs. Put roughly, the inconsistency arises given our general willingness to make small trade-offs among individual goods and their distribution among people (such as our willingness to trade-off a slightly larger headache for one individual for the sake of preventing slightly smaller headaches for two individuals), our general *unwillingness* to subject small numbers of people to very large harms to prevent very minor harms to many people (such as causing a single person fifty years of torturous pain for the sake of preventing very minor headaches for a very large number of people), and the transitivity of "better than", i.e., the claim that if A is better than B and B is better than C, then A must be better than C.

This tension is no revelation.¹ What *is* a revelation is the doggedness with which Temkin pursues his inquiry into it, and the depths he is able to plunge. And while I do not believe the book is flawless, it is nevertheless extremely, well, *important*: it contains a number of groundbreaking discussions of the nature of the good, the nature of evaluative ideals, and the extent to which we should hedge our bets concerning the transitivity of the predicate "better than". Furthermore, I admire this book because it has that ever-so-rare property of philosophical texts: it is actually *convincing*, at least convincing enough to convince me that the status of the "better than" predicate is a substantive matter, the intransitivity of which may well

[†]I'd like to thank Theron Pummer for a useful discussion and a set of helpful comments.

¹See Judith Thomson, *The Realm of Rights* (Harvard University Press, 1990), 169; Alastair Norcross, "Comparing Harms: Headaches and Human Lives" in *Philosophy and Public Affairs* 26 (1997); Dale Dorsey, "Headaches, Lives, and Value" in *Utilitas* 21 (2009); and Temkin's own "A Continuum Argument for Intransitivity" in *Philosophy and Public Affairs* 25 (1996), inspired by work by Stuart Rachels (see "Counterexamples to the Transitivity of Better Than" in *Australasian Journal of Philosophy* 76 (1998)).

be an open result. And though I'm not inclined to accept it, I am no longer prepared to refer to intransitivity, as I once did, as the axiological equivalent of a nuclear holocaust.²

One more note of praise. I find that Temkin's book is an instance of exactly the right *way* of doing philosophy. Let me say a little more about what I mean. Some texts in moral theory attempt to argue for a particular conclusion (i.e., that moral theory X is true; or that ϕ -ing is morally wrong, or that S is worse than T, or whatever) on the basis of intuitions or considered judgments that the author believes are widely shared or are at least shared widely enough to warrant pursuit of an argument taking those intuitions for granted. Of course, Temkin asserts no less than his fair share of considered judgments. The book is chock full of them. But Temkin makes clear that he's not *arguing* that one must accept his way through the inconsistency with which he is mainly concerned. According to Temkin, "A large part of this book's aim is to consider various arguments that can be given for and against the different conflicting positions, to explore what positions stand or fall together, and to illuminate some of the implications of accepting or rejecting the different positions," (p. 9). Temkin likens his argumentative strategy to that of a juggler who just cannot keep the various balls in the air; his philosophical task is to explore the implications of letting each drop. But I prefer the analogy of a cartographer: Temkin's book is a highly sophisticated and detailed map of a certain corner of logical space. He doggedly explores the highs and lows of taking (most) any particular route through that space. Of course, he spends a disproportionate amount of time exploring one particular route (i.e., the route of the denial of the transitivity of "better than"). But the general point is this: I admire Temkin's methodology, and I admire the book's ambitions. My hope is that moral philosophers will learn not just about the substantive features of a number of positions by consulting Temkin's book, but also more about how to do philosophy in a constructive and collaborative spirit.

In what follows I offer two brief notes of critique. I cannot possibly hope to engage all of the significant discussions that occur in the midst of *Rethinking the Good*. I'm going to focus only on those points that interest me, insofar as they interest me for a related reason which, I hope, will become clear.

1. *The Inconsistency*

²Dorsey, "Headaches, Lives, and Value", 38.

Chapters Two through Five set up the book's central dialectic. As it turns out, there is a very serious inconsistency between a number of positions that we find immediately plausible. Temkin brings out this inconsistency in various ways, I'll focus for the moment on the first. According to Temkin, what gets us into trouble is the combination of

The First Standard View (FSV)—Trade-offs between Quality and Number Are Sometimes Desirable: In general, an outcome where a larger number of people have a lower quality benefit is better than an outcome where a smaller number of people have a higher quality benefit, *if* the number receiving the lower quality benefit is "sufficiently" greater than the number receiving the higher quality benefit, and *if* the differences in the initial situations of the people benefited and the degrees to which they are benefited are not "too" great. (p. 30)

Take, for instance, the case I explored in the first paragraph. It seems better to alleviate two slightly more minor headaches than one slightly less minor headache. But this trade-off doesn't seem isolated. It seems right to say that there are a very large number of cases in which we would prefer minor benefits for more people to major benefits for fewer people, assuming, of course, that the fewer people aren't substantially worse-off. So FSV seems to express a plausible thought about the good.

Now take:

The Second Standard View (SSV)—Trade-offs between Quality and Number are Sometimes Undesirable Even When Vast Numbers Are at Stake: If the quality of one kind of benefit is "sufficiently" low, and the quality of another kind of benefit is "sufficiently" high, then an outcome in which a relatively smaller number of people received the higher quality benefit would be better than one in which virtually any number of (otherwise) similarly situated people received the lower quality benefit. (p. 32)

Again, SSV sounds plausible. Take the example I offer in the first paragraph. It seems right to say that *if* we could save a single person from fifty years of torture, we should prefer that outcome to saving any number of people from an extremely mild headache lasting two seconds. That seems right. Indeed, Temkin makes SSV sound more controversial than it really is. Temkin spends time arguing against the position he describes as "total utilitarianism", which he takes to be incompatible with SSV (p. 35). But this is

inaccurate. No total utilitarian need be committed to rejecting SSV. Total, aggregative, summing-up utilitarians can believe in lexical dominance relations among goods and hence can believe that total utility is not increased by such a trade-off.³

We have not yet arrived at an inconsistency. Because both FSV and SSV are conditionals, they can both be satisfied trivially. For FSV, for instance, one can offer an interpretation of a “sufficiently greater” number of people, or an interpretation of a “too great” difference between the initial situation of individuals that renders it the case that any such trade-off fails. Hence FSV is neither compatible, nor even in tension, with SSV. (A similar thing can be said about SSV.) However, in generating his inconsistency, Temkin adds a further premise. He calls this the *spectrum* assumption:

The assumption is just that there is, or at least could be, a spectrum of benefits, ranging from the very great to the very minor, such that one could make a series of pairwise comparisons of benefits across the spectrum, where the First Standard View would apply to all pairwise comparisons involving benefits that were “near” each other on the spectrum, while the Second Standard View would apply to all pairwise comparisons of benefits at the opposite ends of the spectrum. (p. 45)

In other words, the “spectrum assumption” holds that one can move, in a finite series of steps, from individual trade-offs that satisfy FSV, and in which a direct comparison between the first state of affairs and the last state of affairs in this spectrum are not fungible, or satisfy SSV.

But wait: do we have an inconsistency yet? No. That’s because both FSV (and SSV) could be satisfied trivially, as noted before; each individual “nearby” trade-off could satisfy FSV, for instance, even if the proposed trade-off is not allowed (so long as the antecedent of FSV, for this trade-off, is false). So the spectrum assumption requires a touch more precision: what Temkin *really* means is that the pairwise comparisons between states of affairs “near” each other on the spectrum satisfy FSV *non-trivially*; the pairwise comparison between states of affairs at the endpoints satisfy SSV *non-trivially*.

Of course, *even if* we accept the spectrum assumption (duly modified), we have not yet generated an inconsistency. The inconsistency develops only

³Of course, this is complicated by the fact that Temkin holds that total utilitarians are committed to cardinal value. (See pgs. 35, 37.) But cardinal value is *not* incompatible with lexical dominance. See, for instance, Gustaf Arrhenius, “Superiority in Value” in *Philosophical Studies* 123 (2005), 106-110.

if we assume the transitivity of better than: only if we assume that because there is a series of “better than” steps from A to Z, that this implies that Z is better than A. And if we make this assumption, there’s a whopper of an inconsistency in our beliefs: because, in each individual step in the spectrum, the latter state is better than the former state (given the operation of FSV without the qualifier; i.e., C is better than B; B is better than A, etc.), one would expect that Z would be better than A, given the transitivity of better than. But, given the non-trivial operation of SSV, A is better than Z. And hence we have our inconsistency. In fact, the inconsistency is entirely captured by:

1. The spectrum assumption (duly modified).
2. The transitivity of better than.

After having explained the inconsistency at issue, much of the remainder of the book is either pitched as a defense of (1), and an exploration of the denial of (2). The remaining sections of this review touch on these topics in order.

2. Intrinsic Value and the Spectrum Assumption

(1) and (2) are inconsistent. But so are the following three principles:

- 1*. Nazism is morally justified.
- 2*. Genocide is not morally justified.
- 3*. Nazism is characterized by genocide.

This inconsistency isn’t particularly worrisome. It’s obvious which one is false. But Temkin claims that there is something special about his inconsistency. In particular, his inconsistency is interesting enough to devote a 616-paged, small-typed, book to because *his* inconsistency is characterized by principles that are all very difficult to give up. But I’m less sure than Temkin that this is correct. I say I’m less sure not because I’m absolutely certain that there isn’t something very compelling about both principles involved in Temkin’s inconsistency. Rather, I do not believe that Temkin has fulfilled his argumentative burden in showing that these principles *are* each compelling.

And herein lies what I take to be the most significant problem with the book. Temkin discusses at length a number of various principles *about* the good, including the capacity of various states of affairs to be traded-off against each other. Most notably among these is the spectrum assumption.

But at no point does he actually engage in any substantial discussion of what *is* good. At various points he talks about the badness of diseases, headaches, etc.; the goodness of lollipops or the avoidance of torture. But at no point does he offer us anything like an account of what *makes* these things good or bad, or what is *intrinsically* good or bad about having a headache, being tortured, having a dreaded disease, or a lollipop. But on the face of it, this is extremely strange. How could it be possible to conduct an inquiry into the evaluative principles between goods—in particular, their tendency to trade-off against each other in sufficient amounts—without first understanding what *is* good?

Thus I think for quite general reasons, we should refrain on commenting about what worlds are better than others until we are at least offered the *beginnings* of such an account. Without knowing more about the goods or bads that constitute the opposite ends of the spectrum, or indeed the “nearby” trade-offs that constitute the spectrum, and *what else of intrinsic value might be lost as a result of such a trade-off*, we cannot responsibly take an epistemic stand on the spectrum assumption. Indeed, the need for such a discussion seems to me *obvious*. How is it possible to vindicate the suggestion that intrinsic value will obey the spectrum assumption (duly modified, of course), which is itself a substantive principle about the plausibility of trade-offs between intrinsic values, without first knowing what is intrinsically valuable? The spectrum assumption may be plausible on its face, but any commitment to it must be vindicated by showing that that principle is plausibly satisfied by the true (or, at least, serious competitors for the true) view of intrinsic value, “utility”, or whatever term you wish to use. Without such a discussion, any account of the various allowable and disallowable trade-offs between intrinsic benefits for persons seems to me frustratingly top-down.

Temkin might complain. Though he does not explicitly defend hedonism or any other theory of the good, he certainly does not conduct his argument in total isolation from claims about the nature of intrinsic values. In particular, he defends the spectrum assumption in large measure by focusing strictly on trade-offs between pleasure and pain. Fair enough. But hedonism as a *complete* theory of welfare is extraordinarily controversial.⁴ And if hedonism is false, there may be very good reasons to believe, e.g., that

⁴In fact, one that Temkin himself denies. The closest Temkin comes to an explicit discussion of the nature of welfare is his suggestion that “our overall assessment of lives is going to be mainly a function of the quality of the relationship within those lives, of the nature of their projects and commitments, of their mental and physical health, of the degree of respect, self-realization, and accomplishments attained, and so on,” (p. 114n).

50 years of torture does damage not simply to one's overall hedonic levels, but to the extent to which one maintains virtually *any* of the other welfare goods available, including, say, long-term relationships, achievements, etc. (For the sake of brevity, I will refer to all non-hedonic goods by the term "achievements".)

Temkin might say: "so what?" Here's why I think the potential affects on other goods is a problem for Temkin's defense. I think it right to say that in making "extreme" comparisons (the ones for which SSV is relevant), people have different intrinsic values in mind than when making "nearby" comparisons (the ones for which FSV is relevant). For instance, when considering opposite ends of the spectrum, it seems to me our considered judgments are colored by the very real possibility that, say, 50 years of torture is not *simply* going to be painful, but is going to cause lots of other bads (including failures of achievement) that are not at issue in the garden-variety "nearby" trade-off. To put the issue more bluntly, in *not* being explicit about the nature of intrinsic value, Temkin has helped himself to an ambiguity in our own intuitions. In allowing the possibility of additional goods (or, at least, not explicitly arguing that they don't exist or that they lack axiological significance), the plausibility of the spectrum assumption seems to lend itself to an explanation that makes explicit reference to a broader range of values than simply pleasure or pain. When I'm considering Temkin's cases, if he remains strictly neutral (beyond the value of pleasure and the disvalue of pain) concerning the nature of intrinsic value, my own reactions to his cases will be colored by this ambiguity: trade-offs at the extreme will seem less plausible because, perhaps even implicitly, my own reaction is not that the trade-off is implausible on grounds of pleasure or pain's intrinsic value, but rather on grounds of pleasure and pain's *instrumental* role in allowing or disallowing other intrinsic goods such as achievement.

Here's an argument that bears this out. Let's just make explicit the claim that there are to be no instrumental benefits or burdens associated with the distribution of pleasure and pain. How might we do this? Simple: we imagine that all affected individuals are in experience machines. In this case, it is surely plausible to believe that FSV is satisfied (non-trivially). But is SSV satisfied (non-trivially)? Not plausibly. Imagine that we could improve the experience machine slightly for some very large number, while this would cost us only the experience of 50 years of torture for one person. Insofar as all individuals are simply in the experience machine, would we be reluctant to make this trade-off? I confess to finding a "no" answer to this question not just plausible but *obvious*. And, in particular, it's obvious in a way that I don't find obvious when we take the cases *out* of the experience

machine (and focus instead on dreaded diseases, tortures, headaches, etc.), into circumstances in which the possibility of additional welfare goods, and their axiological significance, colors (I think, anyway), my own intuitions.

Here's another case. Let's say that in this case the trade-off involves not people in an experience machine, but individuals whose achievements will not at all be affected by the experience of pleasure or pain. Under such circumstances, it seems far less plausible to say that the ends of the spectrum satisfy SSV (without the qualifier). After all, even for the person who experiences 50 years of torture; he or she will maintain *achievements*. (Let's say, for instance, that the trade-off is between granting billions minor pleasure-related benefits and relieving the long-term suffering of, say, a Stephen Hawking-like character who despite the pain he feels nevertheless maintains substantial and valued achievements.) In this case, I find it much less obvious to say that one ought to opt for the relief of extreme pain for one person. If this is correct, then it would appear that much of what is driving our intuition that the opposite ends of the spectrum reverse the general trend of acceptable trade-offs *depends* on our assumption that a life of 50 years of torture isn't something that is compatible with achievement.

Temkin will surely respond that even if our intuitions are colored in this way, this doesn't help much. After all, if we're unwilling to trade-off billions of moments of small pleasure for, say, a life of achievement this is perfectly compatible with the spectrum assumption, insofar as the value of achievement is *fungible*—surely there is a “nearby” trade-off between pleasure and achievements in which, at a certain level of the former, we should prefer it. At some point it is better to have more pleasure, less achievement. But is this principle something that Temkin can just *assume*? I should think not! In fact, the spectrum is *itself* motivation for its rejection. After all, in considering the ends of the spectrum, we are *not willing to trade-off a life of achievement* for billions of moments of individual pleasure. This suggestion, though it does not of course prove the case, lends a high degree of plausibility to the claim that these goods display a lexical dominance relation, and hence are *not* fungible. And if they aren't fungible, FSV is not satisfied (non-trivially) by a “nearby” comparison between these goods. Indeed, I've argued at length for this very conclusion.⁵

Note that I don't mean to be arguing in any serious way here that the spectrum assumption fails. I press a simple methodological concern: in failing to discuss the first-order nature of intrinsic value (in particular, the precise nature of welfare goods and their relationship), Temkin's defense of

⁵Dorsey, “Headaches, Lives, and Value”, op. cit.

the spectrum assumption is too quick.⁶

3. *Transitivity and the Essentially Comparative View*

While, as noted above, Temkin claims to explore a number of different ways of settling the inconsistency, he is clearly most interested in the denial of the transitivity of better than.

According to Temkin, there is an independent argument for intransitivity to be found in what he calls the “*Essentially Comparative View*”. (There are other arguments he presents—see especially chapters 12 and 13—unfortunately I must leave these aside here.) An evaluative comparison between any two outcomes is essentially comparative if and only if the value of one outcome cannot be measured strictly in terms of its intrinsic properties, but instead must be measured in the context of a comparison with different outcomes. For the *Essentially Comparative View*, there is at least one outcome the value of which is essentially comparative. The contrast of the *Essentially Comparative View* is the *Internal Aspects View*, which claims, predictably enough, that the value of any one outcome can be known on the basis of its intrinsic or internal properties. Temkin argues that if the *Essentially Comparative View* is correct, transitivity “should fail” (p. 473). He also argues that the *Essentially Comparative View* is plausible. And hence we should reject the transitivity of better than.

I am inclined to doubt both claims. In particular, I think that we should reject the claim that evaluative relations are essentially comparative, even more so if the alternative is to reject transitivity! But for reasons of space, I’ll leave aside this point and instead focus on the implications of the *Essentially Comparative View* for the transitivity of better than. If we accept the *Essentially Comparative View*, should we accept the intransitivity of “better than”?

I think the answer is “maybe”, but Temkin has not yet shown that we should. The methodological problem I explored in the previous section seems to me to arise again. Temkin’s main argument from the *Essentially Comparative View* to the failure of transitivity involves his claim that the value of equality should be essentially comparative: that whether or not some particular outcome is better than another in terms of equality will depend not simply on the internal aspects of the outcomes being compared,

⁶Note that Temkin discusses the possibility of lexical priority relations only briefly, and in contexts that do not apply to lexical orderings among welfare goods. (See appendix F, pgs. 30, 107.)

but about particular relations between those outcomes. His evidence for this position is a consideration of Parfit's *Mere Addition* cases. In cases of mere addition—cases in which the difference between two states of affairs just is the addition, in the second one, of a worse-off, but still tolerably well-off segment of society—equality doesn't matter; there is no question concerning the relative equality of the second as compared with the first. But in comparing scenarios that do not involve mere addition, equality matters. (And hence the value of equality is essentially comparative.) But if equality is essentially comparative, cases of mere addition involve a sequence of three states of affairs, for which the relation “is better than in terms of equality” cannot be said to be transitive. As Temkin explains the issue, we move from a comparison between A and B, which are equivalent in terms of equality, to B and A+, the former of which is better than in terms of equality than A+, to the conclusion that A+ is *not worse* than A in terms of equality because A+ is simply a case of mere addition in comparison to A. (See p. 365.)

I'm prepared to grant this. For the sake of argument, let's say that equality is essentially comparative and let's also say that this implies that “better than in terms of equality” is not transitive. But what does this say about “better than” *tout court*? According to Temkin, if there is one aspect of value that is itself intransitive, this aspect of value will be *inherited* by our all-things-considered value judgments (p. 211-12). Temkin writes:

Suppose, for example, that an important moral ideal is both relevant and of genuine significance for assessing the goodness of outcomes all things considered, and that the ideal in question is an essentially comparative one. Then, the relevance and significance of the factors for comparing alternatives all things considered will, to at least some extent, depend on the alternatives being compared. . . . But then, where *different* factors are relevant or vary in their significance for comparing different alternatives, as can happen with an essentially comparative ideal, it can be true that even if A is equivalent to or better than B, and B is better than C in terms of the relevance and significance of the factors for comparing *those* alternatives, A may *not* be better than C in terms of the relevance and significance of the factors for making *that* comparison. (p. 381-2).

All perfectly true, no doubt. But what does this say about whether intransitivity in any value relation actually holds? Little. We have not yet been given an argument that any of the supposedly essentially comparative values involved are *actually* significant enough in comparison to overrule

the non-essentially comparative values in any set of comparisons. We need to know, for instance, *how much* equality is worth in comparison to other sorts of values like, for instance, overall welfare, the value of which is most certainly *not* essentially comparative. For instance, even if there is a case in which the value of equality is non-transitive between three different states of affairs, this says nothing about the extent to which we should accept that value on the whole is non-transitive between these states of affairs. Perhaps improvements in aggregative welfare, especially to the least advantaged, are enough to overrule the axiological significance of equality.

Once again, Temkin might complain. There will be cases in which other values that might themselves be transitive in application across outcomes will be *equivalent*, and hence the essentially comparative value will “take over” as it were, even if that value is far less significant in comparison (p. 211.) But we have not yet been given an argument that this should be possible. In particular, Temkin’s clearest case of the operation of essentially comparative values—and one to which he consistently refers as an example of how essentially comparative values wreck transitivity—involves cases of *Mere Addition*, cases that are simply *stipulated* not to be equivalent in terms of values that are plausibly regarded as not essentially comparative: including, most importantly, overall utility.⁷ For all Temkin has shown, transitivity fails for “better than in terms of equality” *only* in comparing worlds that involve, somewhere along the line, mere addition. But *mere addition* is, necessarily, not equivalent in terms of an essential value: total utility. (Indeed, Temkin explicitly claims that if equality is essentially comparative, “better than” *tout court* is intransitive in mere addition cases (p. 382). But this inference isn’t licensed *given* the operation of other values in cases of mere addition.) And so in this particular case at least, we simply don’t have enough information to *know* whether transitivity applies or doesn’t because we don’t know the relative weight of equality in comparison to other values, in particular, overall utility. (This says nothing, of course, of the other cases Temkin discusses later in the book, or additional reasons to treat “better than” as intransitive in mere addition cases.)

I conclude this section by reflecting on the significance of this worry. In particular, the news isn’t all bad for Temkin. Even *if* Temkin must en-

⁷Note that Temkin also defends an “essentially comparative view of utility” which, roughly speaking, holds that the value of utility depends on *who the extra utility goes to*, and hence would render the value of utility essentially comparative (p. 411). But even if the value of utility is essentially comparative in some cases, this does not mean that an “internal aspects” treatment of utility is not itself a value. Regardless of who is in the two worlds, other things equal a world with utility 20 is better than a world with utility 10.

gage in a much deeper axiological inquiry into the nature of the good before he can claim that intransitivity follows from a consideration of mere addition cases, a very significant result arises. What Temkin has, at the very least, shown is that the transitivity of “better than” depends on a number of substantive axiological considerations, considerations that the partisan of transitivity *cannot simply take for granted*. And hence even if Temkin does not establish the truth of intransitivity, we cannot simply treat transitivity as an axiom to which any comparison of outcomes may be applied. Transitivity is, on this view, completely up for grabs depending on the nature and comparative value of the various goods in a given set of comparisons. And so if Temkin has merely indicated that any commitment to transitivity must await a sophisticated inquiry into the substantive first-order facts about the nature of impersonal goodness, this is a very significant conclusion, indeed (even if, as I am, you are inclined to doubt the plausibility of essentially comparative goods).

4. Conclusion

Rethinking the Good is a thick package of serious moral philosophy. I offer a slight methodological quibble that crops up in two of the potentially hundreds of substantive arguments made therein. Hence this methodological problem should not mask the significant results the book delivers. Temkin’s work is essential reading for anyone interested in puzzles of formal, and informal, axiology.

Dale Dorsey
University of Kansas
ddorsey@ku.edu