The Costs of Transitivity: Thoughts on Larry Temkin's *Rethinking the Good*

Shelly Kagan
Yale University
shelly.kagan@yale.edu

Abstract

In *Rethinking the Good,* Larry Temkin argues that the common belief in the transitivity of better than (all things considered) is incompatible with various other value judgments to which many of us are deeply committed; accordingly, we should take seriously the possibility that the better than relation is not, in fact, a transitive one. However, although Temkin is right, I think, about the mutual incompatibility of the beliefs in question, for the most part his examples don't leave me inclined to deny transitivity. Nonetheless, there is one example, involving infinity, that does seem to me particularly troubling.

Keywords

good – better than – transitivity – value judgments – infinity

Like you, I have read "author meets critics" symposia before, and I know what I am supposed to do here.[1] I am supposed to praise the book that is the subject of the current symposium and tell you that it is a major contribution to the philosophical literature on the topic, but then humbly explain how its central

1   This is a revised version of a paper originally presented at an author meets critic session on *Rethinking the Good* (Oxford, 2012), held at the Eastern Division meetings of the American Philosophical Association, December 29, 2013. I am grateful to Larry Temkin for his comments at the time, and for countless discussions of these matters over the years.

thesis is obviously false, how several of its major assumptions are misguided, how the arguments are weak and the examples unhelpful. I am also supposed to show you how if only one were to adopt a different way of looking at this issue—*my* way of looking at the issue—things would be so much better, philosophically speaking, that we would avoid all of the difficulties the author unwittingly got himself into.

Well, that's what I am supposed to do, but I am not going to do it. I'm not even going to do the very first part, the part where I tell you that Larry Temkin's *Rethinking the Good* is a major contribution to the literature on the topic. For the truth is, that doesn't come close to doing justice to it. It would be somewhat closer to the truth to say that *Rethinking the Good* introduces us to a *new* topic, one that almost no one else has so much as considered before. Temkin himself, of course, has developed many of the ideas presented here in various articles over the years; and Temkin readily acknowledges a debt of his own to Stuart Rachels. But that's pretty much it. Temkin has noticed something that the rest of us have been more or less blind to. And it is something important.

What's more, I am not going to try to argue that the central thesis is false, let alone obviously false, nor that its assumptions are misguided, or its arguments weak. It seems to me that Temkin is simply right about his central claim, and our only question is what we are going to do in the face of recognizing it.

So what is that claim? As a *first* pass, we might put it this way: the concept of one outcome being better than another, all things considered, is not a transitive one. That is to say, A might be better than B (all things considered—a qualification I won't keep repeating), B better than C, and yet, for all that, it might *not* be true that A is better than C. Indeed, C might be better than A. The better than relation is not a transitive one.

I trust you can see just how troubling—and important—it will be if Temkin is right about this. The assumption that the better than relation is transitive pervades pretty much every aspect of our thinking about the good. It is central to all practical deliberation, both prudential and moral, central to decision theory, expected utility theory, economics, and more. If we have to agree that better than is *not* transitive, it is really going to matter.

Still, this was only a first pass at stating Temkin's thesis. Let me note two qualifications to this rough characterization of his view. The first qualification is relatively minor. Strictly speaking, it might be that it isn't so much that the better than relation isn't really transitive, as that the better than relation doesn't apply nearly as widely as we normally take it to. So we can't use it in anything like the way we want to—in the wide range of cases in which we are accustomed to wielding it. Temkin expresses this idea by saying that better than may be *non*transitive—indicating that it may not apply where we thought

it did—rather than necessarily being *in*transitive. For the most part, this qualification won't be important in what follows; but it will be at the end.

The second qualification is more important. Strictly speaking, Temkin isn't trying to convince us that better than *is* nontransitive. What he argues, rather, is that there are various views to which most of us are deeply committed, and that these views are mutually inconsistent. So one or more of these widely held views will have to go. Now, one of the views in the mutually inconsistent sets is the assumption that better than is transitive. So one possibility that we should take seriously, Temkin argues, is the possibility that better than is *not*, in fact, a transitive relation. But Temkin doesn't *insist* that belief in the transitivity of better than is the one that should go. It might be that we should give up one or more of the *other* common assumptions instead. But whatever it is that we give up, there is going to be a cost, and a significant one. There is no avoiding that.

In a moment I want to turn to some of Temkin's examples, ways in which he tries to convince us that belief in the transitivity of better than is incompatible with other deeply intuitive beliefs. First, though, I want to briefly discuss a reaction that many will have to Temkin's examples, which is to say that even if he is right in showing that we are going to have to give up some belief that we won't want to abandon, Temkin certainly cannot be right when he claims that we should be open to the possibility of giving up the transitivity of better than. Many people, I am sure, will think that the transitivity of better than is simply true by definition, that transitivity is built into the very *idea* of one outcome being better than the other, *all* things considered. So whatever it is that we ultimately decide to do to avoid Temkin's problem cases, we *cannot* coherently think that one genuine option is to declare that better than is nontransitive.

Like Temkin, I think that this position is of no real use. Suppose we agree—if only for the sake of argument—that there is indeed a concept of better than for which it is true, by definition, that it is transitive. As Temkin notes, we still must be open to discovering that this is not the comparative relation that is significant for practical deliberation. It might be, after all, that when we judge one outcome more choiceworthy, whether from a personal or a moral point of view—when we judge, say, that there is more reason to choose a given outcome from an impartial standpoint—it might be that *this* relation, of there being more reason all things considered to choose one outcome rather than another, is *not* transitive. Whether it is, or is not, is precisely the issue that we need to investigate. And so, if we simply stipulate that better than is transitive, then the question simply becomes whether better than is indeed the relation to which we are appealing when we judge one outcome more choiceworthy in this reason-generating sense.

So we gain nothing by insisting that it is a mere analytic truth that better than is transitive; we only complicate the wording of the discussion that we are still going to need to have anyway. It may or may not turn out to be the case that "more choiceworthy, all things considered, in the sense of there being more reason to choose" really is a transitive relation. But we might as well make things easier on ourselves when discussing this question, by allowing ourselves to state it in language that asks whether or not it is really true that *better than* is transitive.

More reasonably, one might wonder *how* better than could turn out to be nontransitive. Temkin offers a helpful answer, which we will understand better if we first look at a toy example that he offers of a clearly intransitive concept. Temkin asks us to consider a notion of "larger than" where X is stipulated to be larger than Y just in case X is either taller than Y or heavier than Y. As Temkin notes, it could certainly turn out that A is taller than B, hence *larger*, while B is heavier than C, hence larger as well, and yet, for all that, A is both shorter *and* lighter than C, so A is *not* larger than C. Thus A is larger than B, who is larger than C, yet A is not larger than C.

What's going on here? Roughly speaking, it's this: the concept of larger than is such that judgments about who is larger can depend on different underlying dimensions—sometimes height, sometimes weight. As a result, when we string together a series of such judgments—A to B, and B to C—the underlying criteria doing the work in any given such judgment might change, from judgment to judgment, leaving it completely unsettled how the final judgment—A to C—will come out. Putting the point loosely, but intuitively, if a concept is multidimensional in terms of what underlying dimensions the comparisons using that concept turn on, and if the precise dimensions at work (or their relative weights) can vary, depending on the particular comparison being made— that is, on what is being compared to what—then the concept in question might well turn out to be intransitive.

And Temkin's thought, of course, is that more or less this is what we find when we make judgments about one outcome being better or worse than another. Such judgments are multidimensional in the sense that different underlying dimensions can play a role in making it true that one outcome is better than another, and the precise dimensions that matter (or the weight appropriately given to any given dimension) can vary, depending on the particular comparison being made. As a result, better than might well fail to be transitive, in something like the way that larger than fails to be transitive.

Much of *Rethinking the Good* is devoted to elaborating these ideas, and in particular to showing that better than is indeed multidimensional in just this way. Temkin offers a wide range of examples, many of which are intended to

help us see just how it is that given the values that most of us are strongly inclined to embrace, the specific criteria relevant to making a given comparison will shift depending on what is being compared to what, and how, accordingly, we find ourselves drawn toward judgments about comparative goodness that violate transitivity.

Again, Temkin's official purpose is not to insist that we *must*, therefore, give up on the transitivity of better than. Temkin is the first to recognize just how costly any such decision would be. But the only way to avoid it, he argues, is to give up one or more of the other value judgments to which we are also deeply attracted. Either way, then, there will be a price to pay.

Of course, just how great the price is will depend on how deeply you are attracted to the relevant value judgments, the ones that end up being incompatible with the transitivity of better than. So to do justice to the book, we would now need to consider each of the many examples that Temkin offers, so that you could ask yourself, in each case, what claim is the claim that you are prepared to abandon, and how deeply it bothers you to have to do this. Perhaps, for one or more examples, you would ultimately agree that giving up on transitivity might be the most reasonable choice to make. That would be of obvious significance, even if there were still other cases where you preferred to abandon something else.

Truth be told, however, even if one was *never* prepared to abandon transitivity, and always chose to give up something else, that might still be significant, insofar as you were prepared to admit that the particular value judgments you were giving up were ones you didn't really want to abandon. For it would still be the case that you might need to significantly rethink some of your deepest commitments in value theory, revising your views in ways you hadn't anticipated. Temkin's book, after all, is called ***Rethinking the Good***. He would be content, I think, to have us realize just how much rethinking of our views is going to be necessary.

As far as I can see, the only people who won't be troubled by Temkin's arguments will be those who think that his examples are *easily* answered. That in each case there is some assumption that Temkin thinks we are drawn to, but for which it is actually the case that the critic himself isn't drawn to the assumption at all, or is only weakly drawn. Or perhaps the critic has an argument of her own up her sleeve, to show that we have *independent* reason to abandon the belief in question, so the cost isn't really all that great after all.

A fair bit of the book is devoted to considering arguments of this last sort. Temkin has discussed his various examples with a lot of people over the years, and he has heard a lot of objections. The book considers a very large number of these objections, and painstakingly argues, in each case, why the objection in

question is far less compelling than one might initially have thought. Accordingly, Temkin concludes, while you may well be prepared to pay the price of resolving this or that inconsistency, the price is still likely to be considerably higher than such critics assume.

Still, the proof is in the details, and as I say, to do justice to the book we would now need to look at the most important examples, objections, and replies. Unsurprisingly, I don't have the space to do that here. But it does make sense to mention one of the cases that many people will find most persuasive, or, at least, deeply troubling.

Building on an example from Stuart Rachels, Temkin asks us to consider a very long life in which you would have to undergo two years of intense torture. Next, compare this to an alternative life, equally long, and similar in other respects, except that instead of having two years of torture, you have four or six years of an experience that is also torture, but *slightly* less intense. Temkin suggests, and I agree, that most of us would think the second life is worse than the first, all things considered. Admittedly, the pain is slightly less intense in the second life, but it is almost as bad, and it lasts two or three times as long. So the second life is worse, all things considered. And now consider a third life, equally long as the first two, and similar in other respects, except that now you have twelve years of an extraordinarily painful experience that is in turn somewhat less intense than, but only slightly less intense than, the pain of the second life. Again, although the pain in the third life is *slightly* less intense than that of the second life, it lasts at least two or three times as long; and so, on balance, it seems that the third life is worse than the second.

Continue constructing a series of lives like this, with each one having a pain that is slightly less intense than the one had in the life before it, but lasting two or three times as long. For each pair of "neighboring" lives, the life with the longer pain will be *worse*, since even though the longer-lasting pain is slightly milder than the pain in the earlier life, given how *much* longer the pain lasts, the later life is worse overall. If we carry this process on long enough, eventually we will end up with a life that has a pain that is very mild indeed, but which lasts for a *very* long time. In Temkin's own version of the case, the final life is one where the pain in question is equivalent to having an extra mosquito bite every month, for every single month of your very long life.

Now let us consider the judgments that we would apparently make about this series of lives. The first life (with two years of torture) is better than the second life (with six years, say, of only slightly less intense torture), which is better than the third life, which is better than the fourth, and so on, and so forth, all the way down to: ... which is better than the final life (with the very long series of months in each of which you experience the pain of one extra

mosquito bite). If the better than relation is transitive, this entails that the first life is better than the final life.

But this, Temkin argues, is very hard to believe. Think about it. You have to choose between two lives, both very long, and otherwise similar. But in one of them you will suffer two entire years of incredibly intense torture, while in the other you merely have to put up with the extra very mild discomfort of having one extra mosquito bite each month. Surely the first life is *not* better than the second. On the contrary, the second is better than the first. But if that's right, then we have a contradiction.

So something has to go: *either* we have to say that we cannot possibly construct a series of lives like this, or we have to give up the transitivity of better than, *or* we have to say that for some pair of neighboring lives in the series, it is better to have a pain that is only slightly milder, even though it will last several times as long. Or we have to say that it is better to have the first life, where we suffer torture for two years, than it is to have the final life, where we merely undergo the mild discomfort of an extra mosquito bite every month, albeit for a very long time. Temkin says, and I am sure he is right about this, that most of us will have an incredibly hard time choosing among these various unpalatable alternatives.

Temkin doesn't say which way out to take, but he does insist that we should at least be open to the alternative according to which we abandon transitivity. After all, this seems to be a case which fits the model I mentioned earlier, where shifting underlying criteria for comparisons can result in lack of transitivity.

Intuitively, the situation is this. When comparing neighboring lives, it is relevant to think about both the qualitative difference in the pain, and the duration of the pain. So that is one set of criteria that does work for us in determining many of our better than judgments. But when it comes time to compare the first life (with two years of torture) and the final life (with the ongoing series of mosquito bites), the simple fact is that the qualitative difference now seems to matter in a decisive way, while the difference in duration seems to drop away as utterly irrelevant. Admittedly, we might say, the pain of mosquito bites last for an immensely longer time than the two years of torture—but this just doesn't matter! It doesn't matter *how long* the mosquito bite discomfort goes on, this simply doesn't make a difference when we are talking about the choice between that sort of insignificant discomfort and *torture*. So in making the comparisons between neighboring lives, the criteria are duration and quality, and duration can outweigh quality, but when it comes to making the comparison between the first and the last life, the criteria shift, and now all that matters is quality alone. But as we have already seen, failures of transitivity aren't

surprising with shifting criteria. So perhaps the best way to resolve the inconsistency really is to deny the transitivity of better than.

I do think that Temkin has said enough to make this a live option, one that needs to be taken very seriously. And that is a significant achievement indeed.

As it happens, though, it is not the way out that appeals to me. I honestly don't know how a typical person—the mythical person in the street—will go here, but I imagine that at least some of the readers of this essay will prefer to say that, surprising as it may seem, the final life with all its mosquito bites really *is* worse than the first life, with its two years of torture. And this is what I am prepared to say as well.

Mind you, I don't say this easily. I want to recognize that this is an extremely unintuitive thing to say. Temkin is right, I think, when he says there are no cost-free ways out here. But that means, of course, that *some* way out is going to be the *right* way out, even if it is difficult to accept. And having thought about it, I think the best of the sorry alternatives is to just insist that the life with the mosquito bites really *is* worse than the life with the torture. I wish I saw a more appealing thing to say about this case. But I don't. So I am going to say it.

Of course many philosophers would rush to my defense at this point, and insist that this cost isn't really all that horrible. Admittedly, it seems very unintuitive that the life with the mosquito bites could be worse than the life with the torture, but we really shouldn't trust our intuitions about this comparison, since by the time we have taken all the steps needed to move us from torture down to mosquito bites, the length of the final pain will be *unimaginably large*. After all, we are starting with two years of torture, and doubling or tripling the duration as we move from step to step. To fit this all in within a single life, at a minimum the life itself must be thousands upon thousands of years long, probably much longer than that. And in the final life we will have to suffer the pain of a mosquito bite every month, year after year, for thousands upon thousands of years, if not much more. We simply have no ability to accurately wrap our head around those kinds of numbers, or pains that last this long, and so—the objection concludes—it really shouldn't surprise us that our judgment here, comparing two years of torture to this sort of unimaginably long-lasting discomfort, is so unreliable. So the "cost" of taking this way out of the inconsistency really isn't all that great after all; and we certainly needn't deny the transitivity of better than.

As I say, I am confident that many philosophers will rush to my defense with a speech along these lines (though the details might differ). And while I appreciate the help, I honestly don't know whether to accept it. At the very least, the answer is too crude as it stands. We are, after all, able to get our minds around all sorts of large numbers. And at any rate, when we consider the neighboring

pairs of lives toward the end of the series, both lives will involve unimaginably long pains, and I bet you don't find yourself at all tempted to say something like, "Oh, I can't choose between a million years of somewhat itchy mosquito bites and three million years of only slightly less itchy mosquito bites; the numbers are too large for me to grasp." So if our intuition is indeed letting us down here because of the incredibly long durations in question, we need a more detailed account of how and why it lets us down, a story I don't think we have.

Still, as I say, this does seem to me the best of the various unattractive alternatives. So for whatever it is worth, this example and others of this same basic sort don't move me to abandon the transitivity of better than.

We might still ask: how great *is* the cost of taking this way out? That's not at all clear, and I do think it is a shortcoming of Temkin's book that he doesn't say anything on this score. At one point, in discussing the various possible answers to cases like this, Temkin says that each way out will "require major revision in our practical and theoretical thinking" (p. 47), but I am not at all sure that he has made good on this charge. I imagine we can all see the cost of giving up transitivity, and I certainly can see the cost of claiming that duration cannot outweigh quality of pain, not even when the intensities are fairly similar. But what is the cost of insisting that the incredibly long life with two years of torture is—surprising as this may be—actually *better* than the incredibly long life with an incredibly long stretch of mild discomforts? I admit it seems rather unintuitive to say this, but is there any wider cost beyond this? If there is, I don't see it.

(In saying this, I don't mean to deny that there is indeed a cost involved in rejecting the intuition that the life with long-lasting but mild discomfort is better than the life with torture. And of course there are similar intuitions, involving other such cases, that will have to be abandoned as well; and this too is a cost. But I don't see how abandoning these intuitions requires the sort of "major revision" in our practical thinking that Temkin says it would. On the contrary, the required changes seem relatively straightforward and reasonably self-contained. In contrast, to abandon transitivity threatens to require a systematic overhaul of almost all of our practical deliberation.)

I find myself having a similar reaction when I consider most of Temkin's various other examples as well. In fairness to Temkin, I should explicitly note that the examples he offers are of various kinds, with different structures, so even if you aren't much troubled by the one I just discussed, there is a good chance you *will* be troubled by some of the others. I certainly do have to pause when I think them through. To be sure, I don't *always* have the intuitions Temkin is trying to elicit, but I must admit that I have them often enough. So for several of the cases he considers, the intuitions I share with him will lead to

mutually inconsistent claims, and so *something* will have to go. But for almost all of these cases I can pretty quickly identify what I want to give up, and in almost none of these cases is it the transitivity of better than. And although I don't have much of anything by way of argument that the intuitions I am prepared to give up really should never have been trusted in the first place, still, as I contemplate the prospect of abandoning the intuitions in question, although this gives me some misgivings, the truth is that the misgivings are not, in fact, all that great.

But honesty forces me to confess that there is an exception. One of Temkin's examples not only gives me pause, the simple fact is, it really troubles me. Others may find different examples of Temkin's more troubling. This is the one that does it for me.

As it happens, this particular example isn't actually given in the book itself, though it can be viewed as a sort of simplified cousin of one that is in fact given there (on pp. 437–38). At any rate, I learned this example from Temkin, and it neatly brings out the difficulty of preserving transitivity across the board.

Imagine a world, L1, with an infinite number of people, and suppose that for each integer, positive, negative, or zero, there is exactly one person in that world at the level of well-being represented by that integer. That is to say: there is one person, Larry, say, with 1 unit of well-being, and another person, Lisa, with 2 units of well-being, while Leonard has 3, Linda has 4, and so on; meanwhile, Lucile has 0, Luke -1, Lydia -2, and so on.

Now compare this to world M, which also has an infinite number of people, with one person at each level of well-being. Indeed, M is exactly like L1, except that there is a completely different set of people in the world. Thus Michael is at 1, Miranda at 2, Manfred at 3, and so on, while Monica is at 0, Matt at -1, Meg at -2, and so forth.

Thinking about these two worlds, I have a completely firm intuition that the worlds are equally good. For each integer, after all, there is exactly one person at the level of well-being represented by that integer. And since it makes no difference whatsoever *who* is at any given level, Larry or Michael, Lisa or Miranda, and so on, the two worlds are exactly equal to one another in value.

Now consider world L2. This is exactly like world L1, with the very *same* people in the world as L1 has. The *only* difference is that each person is 100 units of well-being better off in L2 than they would be in L1. Thus Larry has 101 units of well-being in L2, rather than the 1 unit he has in L1; Lisa has 102, rather than 2; Lydia has 98, rather than -2; and so forth. Similarly, Lyle (whom we haven't mentioned up till now) has -471 in L2, rather than the -571 he has in L1, while Lois has +329 in L2, rather than the +229 she has in L1. As I say, every single

person has been bumped up 100 units of well-being, in comparison to the level they would have had under L1.

When I compare L2 to L1, I have a completely firm intuition that L2 is *superior* to L1: it is tremendously better for L2 to exist than for L1 to exist. After all, each and every person is better off in L2 than they would have been under L1. L2 is clearly a better outcome.

Finally, compare L2 to M. When I think about these two worlds, it is once again completely clear to me that they are exactly equal to one another in value. After all, for both worlds there is exactly one person at the level of well-being represented by that integer. The only difference is that L2 has each slot filled by one of the L-folk, while M has that same slot filled by one of the M-folk, and it is about as clear as clear can be that this fact alone cannot make a difference to the value of the worlds. So they are equal in value.

Now put these results together: L2 is equal in value to M, which is equal in value to L1. And surely, one would have thought, the equal in value to relation is transitive. So this implies that L2 is equal in value to L1. But L2 is *not* equal in value to L1. L2 is *better* than L1.

That's a contradiction, so something is going to have to go. Either we must deny that L2 is equal in value to M, or deny that M is equal in value to L1, or we must deny that L2 is better than L1. The only alternative is to accept the non-transitivity of the equal in value to relation.

One quick word about this last alternative. Up to this point, I haven't actually mentioned the equal in value to relation at all. I have, rather, been discussing the transitivity of the better than relation. But I take it that the transitivity of these two relations is going to stand or fall together. (And the same thing is true, I presume, for various other familiar cases—involving the less good than relation, for example, or mixtures of better than and equal in value to, and so on. For example: if A is better than B, and B is equal in value to C, then A is better than C.) At any rate, anyone committed to the transitivity of better than is probably going to be committed to the transitivity of equal in value to as well.

So, what should go? I imagine that at least some people will want to insist that the claim to abandon is the judgment that L2 is better than L1. After all, one might note, the two worlds are exactly alike, in that for each integer there is one person at the level of well-being represented by that integer. The only difference between the two worlds is who is filling which slot. And since we have already admitted that this makes no difference at all when comparing the L worlds to the M world, surely we must admit that this makes no difference when comparing L2 to L1. So despite whatever intuition we might have to the contrary, L2 is not *better* than L1, it is equal in value to L1.

Well yes, one could say that, but I find that I simply cannot take this option seriously. Larry is *better off* in L2 than in L1, and so is Lisa and Leonard and Lucile and Luke, and *all* of the other people in that world as well. Every single person in L2 (and L1) is better off in L2 than in L1. I find it impossible to think that this is irrelevant. Accordingly, it seems to me extremely difficult to take seriously the suggestion that the two worlds are only equal in value.

And if you think that *you* are tempted by this suggestion, try imagining that you yourself are one of the L-folk, perhaps at -50. There you are, in L1, and God asks you if it would have been better for him to have created L2 instead—a world where you, and everyone else as well, would have been *100* units of well-being better off. Try to imagine telling God not to worry about it, that L2 isn't really any better a world. Try to imagine saying it and believing it! (I dare you.)

Admittedly, this pareto-like consideration isn't the only thing that matters to us in comparing worlds.[2] There is the same amount of well-being in both worlds, and the same amount of inequality, and so on. But none of that undermines my intuition that it also matters that everyone in L2 is better off than they are in L1; indeed it matters sufficiently so as to make it true that L2 is better than L1.

And yet, at the same time, I find myself completely confident that L2 is no different in value from M. After all, when comparing L2 with M, the pareto superiority of L2 to L1 falls away as completely irrelevant. The L-folk don't appear in M at all. So all we are *left* with is thinking about the pattern of distribution of well-being, and this, obviously enough, is exactly the same in L2 and M. So I find myself completely confident that L2 is exactly as valuable as M. And of course, for similar reasons, I find myself confident that *L1* is exactly as valuable as M.

Where does that leave me? Just where Temkin wants me to be: questioning the transitivity of the equal in value to relation! It seems that one sort of consideration—the pareto consideration—is relevant for one of the comparisons (L2 to L1), but not relevant for the others (L1 or L2 to M), and that this leads to nontransitivity.

---

2   In endorsing this pareto-like thought, I do not mean to embrace the distinct idea that there is special significance to the fact that everyone is better off and *no one* is worse off, as though the absence of someone's being made worse off is an *extra* reason to favor L2 over L1. Rather, what we have here is just the limit case of a more general principle, one which takes into account how each person is affected, for better or for worse, when comparing one outcome to another, and which then balances the comparative losses (if any) for some against the comparative gains (if any) for others. It is only this more general principle to which I am attracted.

Let me emphasize that last word. What we seem to have here is a demonstration of *non*transitivity. And here it might be helpful for me to remind you that as Temkin intends the term, the idea is that to say that equal in value to is nontransitive is to say that either it is intransitive, *or* that it fails to apply where we would have expected it to apply.

I am not at all sure about what I am about to say, but I am inclined to suspect that it is this second disjunct that applies to the case we are considering. It isn't so much that equal in value to has been shown to be intransitive, but rather, that the relation in question doesn't actually apply in quite the way that we might have thought it did.

More particularly, perhaps this is a place where we need to go "fine grained" in thinking about the appropriate specification of the alternatives. More particularly still, perhaps we need to think of each relevant outcome as partially specified in terms of what its relevant alternatives are. If we have different alternatives to what appears to be the same outcome (when it is only coarsely specified), this will actually mean that what we actually have are different alternatives to quite distinct *fully specified* outcomes. Thus, for example, it is a mistake to think that there is a single outcome, L2, that we have been comparing to its various alternatives. Rather, there is one outcome "L2 (where the only alternative is M)" and there is a quite distinct outcome "L2 (where the only alternative is L1)," and there would be yet another possible outcome "L2 (where the only alternatives are both L1 and M)."

If we distinguish our various outcomes in this more fine grained fashion, then strictly speaking the transitivity of equal in value to has been saved. For now we are not, inconsistently, committed to the claims that equal in value to is transitive, and that L2 is equal in value to M, which is equal in value to L1, and yet L2 is not equal in value to L1. Rather, we can consistently maintain that L2 (with the sole alternative of M) is equal in value to M (with the sole alternative of L2), and M (with the sole alternative of L1) is equal in value to L1 (with the sole alternative of M), and yet L2 (with the sole alternative of L1) is *not* equal in value to L1 (with the sole alternative of L2). Even if we throw in the transitivity of equal in value to, there is no contradiction, because of course we have changed middle terms in stringing together our various judgments: initially we were discussing the value of M (with the alternative of *L2*), and then we moved to discussing the quite different outcome M (with the sole alternative of *L1*). Nothing interesting follows from the transitivity of equal in value to, given this switch in the middle outcomes being compared. (And, similarly, even if we put this first point aside, the conclusion would only be that L2 (with the sole alternative of M) is equal in value to L1 (with the sole alternative of M); and this obviously isn't at all inconsistent with the claim we are genuinely committed

to, namely that L2 (with the sole alternative of L1) is not equal in value to L1 (with the sole alternative of L2).)

So we can save the transitivity of equal in value to (and similarly for better than and worse than, and so on) by going fine grained. But of course we do this by admitting that equal in value to is *non*transitive. You might have thought that there was a single, simply describable outcome, L2, for example, whose value could be compared in turn to M or to L1. But that is just a mistake. There is no such simple outcome to be evaluated.

Thus we save transitivity, and avoid contradiction, but at the cost of admitting that relations like equal in value to, and better than, and so forth, don't actually apply in all the sorts of cases where we would have thought that they apply. That, at any rate, is the way out I find myself inclined to take here.

Well and good, but what is the price of doing this? And is this, indeed, a price we are prepared to pay? Again, Temkin's claim is not that there are no solutions to his problems, only that any solution comes with a price, and we won't be happy about it. So what is the price of going fine grained in our example?

As we have already seen, going fine grained allows us to make all of the specific pairwise comparisons we are inclined to make, while preserving transitivity. So far, so good. But suppose we have a choice where all three worlds are available to us *simultaneously*, and we must choose between them. What then? Here I find myself drawn to the thought that L2 (with all three worlds available) is clearly better than L1 (with all three worlds available). But now I get troubled once more. I am tempted *both* by the claim that L2 (with all three worlds available) is equal in value to M (with all three worlds available), *and* by the further claim that M (with all three worlds available) is equal in value to L1 (with all three worlds available). Yet if we add in both of these judgments, along with transitivity, we will once again get a contradiction!

I tentatively prefer to retain transitivity and avoid inconsistency by *abandoning* the claim that M (with all three worlds available) is equal in value to L1 (with all three worlds available). Admittedly, that move seems to me to be somewhat unappealing, a genuine further cost; but it may well be the way to go: crudely speaking, when all three worlds are "contenders," it just isn't true that M is no better than L1, since L1 can be beaten by L2, while M cannot. Loosely and somewhat inaccurately speaking, M and L1 are only equal in value when L2 *isn't an alternative*; when L2 *is* an alternative, M is *better* than L1. As I say, I certainly do think that having to say this counts as an extra cost, but I think it is a cost I am prepared to pay. (Unsurprisingly, Temkin has some very helpful things to say about all of this as well.)

Unfortunately, the costs of going fine grained in our example may be greater still. After all, if a fine grained approach is appropriate here, why not else-where? Suppose we conclude that what our discussion shows is that one must go fine grained *everywhere*. That is, suppose we conclude that since the proper specification of the outcomes in our infinity case requires that an outcome be specified partly in terms of its alternatives, we must do this in all choice cases whatsoever, even finite cases. Whenever we must choose between outcomes, those outcomes must be specified in a fine grained manner.

As Temkin notes (pp. 462–64), this might well turn out to be a *horrible* cost to pay. In principle at least, it could render practical deliberation impossibly complicated in anything other than the easiest cases. Suppose for example that you have to choose between 10 candidates for a job in your department. Instead of being able to get by with a mere nine comparisons (Candidate 1 to Candidate 2, the winner of that comparison to Candidate 3, and so on), you may actually need to make 45 comparisons, since a proper specification of a given candidate will actually include information about who they are being compared to. And if the number of candidates becomes remotely realistic (in the hundreds, say), there may need to be tens of thousands of such pairwise comparisons.

Of course, we might decide that a better way to make the comparisons is by thinking of the alternatives as being the entire class of candidates, all at once. Then we can compare Candidate 1 (where the set of 10 candidates exhausts the alternatives) to Candidate 2 (where the set of 10 candidates exhausts the alter-natives). And here, happily, we will once again only need to make nine com-parisons. So perhaps the cost isn't that great after all.

But wait. What if the application of a new candidate comes in, just as we come close to finishing our work? That changes the set of alternatives, and so, given the need for fine grained comparisons, in principle it seems that all of our work goes out the window, and now we may need to start again! As I say: in principle, at least, practical deliberation could become incredibly more difficult.

At first glance it might seem that we could escape this cost by deciding that the need to go fine grained only arises when dealing with cases that involve infinity. Perhaps as long as the case being examined involves only a *finite* num-ber of people (or times, and so on) we can stick to the more familiar course grained approach, avoiding the explosion in difficulty that potentially arises when we go fine grained. After all, it is a familiar thought that cases involving infinite numbers need to be handled with different tools than finite cases. So perhaps the trouble-making move to fine grained specification only arises in a few fancy philosophical examples, nothing more.

But that thought is too quick. For all we know, the cases we deal with *every day* are themselves infinite ones! After all, for all we know the universe is infinite, with life on countless other planets; indeed, the universe might contain an infinite number of people. If so, then strictly speaking every single choice we make is a choice for which a full and complete description of the alternatives would refer to an infinite number of people. So if a fine grained approach is required for infinite cases, it may well be appropriate for every single choice we make.

A second tempting thought is that we may not need to go fine grained for every case involving infinity. Perhaps we only need to do this when the case involves infinity *and* our choice will *affect* an infinite number of people (or times, etc.), as with the example we have been discussing. If so, then even if we live in an infinite universe, it might be that for all practical purposes a coarse grained approach suffices. After all, it might be suggested, even if there are an infinite number of people, in real life cases I can only affect a *finite* number of them. If that's right, then the "cost" of going fine grained for cases where an *infinite* number of people are affected may not be all that unreasonable.

But this second thought may be too quick as well. First of all, it is far from clear whether infinite cases only push us toward a fine grained approach when an infinite number of people are affected by our choice.[3] But even if it *were* true, what of it? Even if I can affect only a finite number of people at a given time, it is at least conceivable that my actions might affect future generations as well, on and on for all eternity. If so, then for all we know we might be back to facing the relevant sorts of infinity cases all the time, with the need to go fine grained in every choice we make.

Well, perhaps we *do* face that need—on a *theoretical* level. But that still doesn't mean that we will in fact need to make some huge number of actual comparisons. It might be that we can establish some general principles that justify coarse grained comparisons in the sorts of cases we actually do face,

---

3 Consider the following case (adapted from Caspar Hare, *The Limits of Kindness* [Oxford, 2013], p. 188): Imagine that in world B1 there are an infinite number of B-folk at -100 and an infinite number of B-folk at +100. World C is exactly the same, except that all the people involved are C-folk instead of B-folk. Finally, imagine that world B2 is exactly the same as B1 except that one person who would have been at -100 in B1 is at +100 instead. Here too, it initially seems to me that B2 is equal in value to C, which is equal in value to B1, yet B2 is better than B1. So if transitivity is to be preserved, perhaps we should go fine grained here as well. Is this, then, a counterexample to the adequacy of the proposal being considered, since only a finite number of people are affected by the choice between B2 and B1? Or is it relevant that the choice between the B worlds and *C* does involve affecting the fate of an infinite number of people? As I say, the matter isn't clear to me.

even if they are (for all we know) infinite ones. That doesn't strike me as a hopeless thought. It might be, after all, that we can make sufficient progress on our value theory so as to be able to identify the sorts of factors and the precise sorts of cases that would leave coarse grained comparisons inadequate. And it might be that such cases rarely arise. Maybe they never actually arise at all, and remain mere theoretical possibilities—like our case involving the integers. If so, then despite the theoretical complexity introduced by going fine grained for infinity, the practical costs might be quite limited indeed.

I don't know if that's right. I certainly hope it is. Or perhaps you yourself have a different preference from mine concerning how best to escape the contradiction that our worry about the infinite worlds has raised.

But remember, this is just one example, and the book discusses dozens of them. Larry Temkin's *Rethinking the Good* opens up a veritable Pandora's box of puzzles and problems, challenging some of our deepest held beliefs about practical reasoning, deliberation, and value. Thinking through all these problems and issues, even if only to your own personal satisfaction, is going to take you a while. Regardless of what you think about the transitivity of better than, you have your work cut out for you.

There is, however, one thing I am sure of. The original Pandora's box of Greek myth would have been better left unopened. But as philosophers, I am confident, ultimately we will all be the richer for Temkin's having opened his.