required for its practical relevance. What is required is a model that can help us think through these issues and get us to approximate answers. And of course, no such model can be one that also (a) appears to predict that there are no answers or (b) engineers an ‘answer’ by betraying our ideals. What we appear to arrive at by Rethinking the Good’s end is choice between options (a) and (b). Perhaps, there are other relevant alternatives.\footnote{A further, less cursory, defence and introduction to the Transitional View, including its unique advantages, possible demerits and how it aggregates, can be found in Coons and Weber (unpublished data).}

Bowling Green State University
Bowling Green, OH 43403-0001, USA
clcoons@bgsu.edu

References


Temkin’s essentially comparative view, wrongful life and the mere addition paradox

M.A. Roberts

Larry Temkin’s new book displays two features that, though otherwise rare, are always found together in any great work of philosophy. Rethinking the Good is both ground-shaking and crafted with near-perfect precision. Moreover, it is fun to read and, as one of the most important works in ethics to appear in recent decades, sure to be studied closely by philosophers and non-philosophers alike for decades to come.

One of the best things about Rethinking the Good is the account of the person-affecting approach in ethics it offers. Temkin’s account avoids the conceptual issues that many competing person-affecting accounts have stumbled into. It’s true that the root of such conceptual issues is often that they are divisional in nature. It’s true as well that Temkin divides people into
two classes — *independently* existing people and *dependently* existing people. But for Temkin the distinction between classes of people operates only as a useful organizational structure. Temkin thus rejects the notion that, for purposes of our moral assessments, the plights of the members of one class of people *matter morally* while the plights of the members of another class of people do not. The person-affecting account that Temkin describes, instead, considers *both* classes of people to matter morally.

Moreover, Temkin manages to avoid the conceptual issues *without* losing sight of the core intuition that has made the person-affecting approach so interesting to so many for so long. Thus, on Temkin’s account, whether one outcome is morally worse in person-affecting terms than another is to be determined by reference not to the features those outcomes happen to have that *matter to no one* — that make no *existing or future* person worse off — but rather to ‘how … *particular* people are affected for better or worse in those outcomes’ (417). The upshot is that, where our question is whether one outcome X is worse than an alternate outcome Y and where, say, you exist in X but not in Y and an alternate (nonidentical) person exists in Y but not in X, the fact that Y makes that alternate person better off than X makes you will not in itself be sufficient to determine that X is worse than Y. In other words, *identity matters*; if Y doesn’t make you (or if not you than someone who does or will exist in X) better off, it may be that X is not going to come out so badly in comparison to Y after all.

Despite the fact that Temkin has made enormous progress in developing a cogent, intuitive person-affecting account, I want to suggest in this present paper two junctures at which Temkin may have unnecessarily pulled his punches. I want to suggest, in other words, that Temkin’s account can do a better job than he thinks it can in addressing two important problems in population ethics.

One of these is the problem of *wrongful life* — the life that is so anguished that it is *less* than worth living. We need to be able to explain why it often

---

1 One example of the divisional approach is the ‘prior existence view’, decisively refuted by Singer 2011: 8–89. For discussion of that view and other versions of the person-affecting approach that also seem immediately vulnerable to charges of inconsistency or incoherence, see Roberts 2011a, 2011b: ch. 2, 2011c. See also Holtug 2010: ch. 9.

2 Nils Holtug develops a view that may be called ‘person-affecting’ but that rejects this core person-affecting intuition. Following Parfit, Holtug describes the view he develops as a ‘wide’ person-affecting account (Holtug 2010: 156–62). In contrast, Temkin’s account evaluates outcomes in terms of how a ‘narrow group of particular people are affected for better or worse’ (Temkin 2012: 417) (emphasis added). It should be noted, however, that Temkin’s reference to a ‘narrow group’ does not signal that his account insists that the people who matter morally are only those who *actually* exist (exist, that is, at the uniquely *actual* world) or only those who exist in each of the outcomes we aim to compare or only those who exist under whatever act we aim to evaluate. It is those sorts of accounts that have led many theorists to reject the person-affecting approach as inconsistent or incoherent. See note 1 above.
makes things worse – and is indeed wrong – to bring such a life into existence. Temkin thinks his person-affecting account cannot provide that explanation. I think, however, that it can, given just a minor adjustment in his view – and indeed an adjustment I think we are going to feel some pressure to make in any case. The second is the mere addition paradox. Here, I believe we can mine from his person-affecting account as it stands – no adjustment necessary – a more intuitive story than the one Temkin himself explicitly recognizes. Specifically, it’s a story that doesn’t give rise to the intransitivity results that Temkin has famously defended in the past and defends again in this new work as well (416).³

The import of these two points is clear. After all, it’s the deficiencies in the person-affecting account – which Temkin regards as a kind of essentially comparative view – that compel him to take the position that we must combine that account with an impersonal approach – an internal aspects view – to produce a correct account of all-things-considered worseness-between-outcomes.

I should acknowledge, however, that Temkin explores deficiencies in the person-affecting approach that go beyond those I try to address here (422–25). Hence – even if the intermediate conclusions I reach are correct and we can address two specific problems within (roughly) the scope of Temkin’s own person-affecting account – we can draw no ultimate conclusion whether a complete account of worseness-between-outcomes will require an impersonal element or not.

1. Some background

A significant portion of Temkin’s book is devoted to what may initially seem a ‘winner-takes-all’ rivalry between the essentially comparative view (ECV) and the internal aspects view (IAV). ECV and IAV provide us with very different conceptions of how outcomes are to be assessed in terms of their moral value – and thus ultimately how they are to be ranked vis-à-vis one another in terms of their moral worseness. But each view has its advantages. Thus, Temkin considers ECV intuitively compelling (416–56). But he also thinks IAV addresses certain deficiencies in ECV – and, specifically, in the person-affecting account – that we cannot otherwise remedy.⁴

Temkin’s own conclusion is that ‘reasonable advocates of the [person-affecting view] should be pluralists’. They might, more generally, find ECV

³ See also Temkin 1987: 138–87. An important caveat: in this present work, Temkin advances a revised conception of what his arguments relating to the Mere Addition Paradox (and many other problem cases) establish. The claim is no longer that those arguments establish – that is, prove – that the better than relation on outcomes is intransitive. What is established, rather, is that one cannot ‘consistently maintain each of several positions’, one of which is transitivity (Temkin 2012: 9).

⁴ Many theorists would consider the non-identity problem to show that ECV itself needs to be discarded in favour of, or (somehow) supplemented by, IAV (Temkin 2012: 424).
compelling but they should recognize that ‘more than one moral factor or ideal’ – and, here, the one ideal that Temkin is referring to is the person-affecting ideal; the view that what makes an outcome worse is to be determined by reference to how particular people are affected – ‘is relevant for assessing outcomes’ (425).

According to IAV, the value of a given outcome X is in part determined by those features of X we can glean through an examination of X and X alone – that is, X’s internal features – without examining the corresponding internal features displayed by still other outcomes. For example, X’s containing a certain quantity of utility (or happiness or wellbeing) in the aggregate or displaying a certain degree of equality or human flourishing will be part of what determines an all-things-considered value for X – a value that will remain fixed whether or not there exists an alternate outcome Y that contains even more aggregate utility or displays an even higher degree of equality or human flourishing.

In contrast, ECV insists that, for at least some outcome X, the value of X is determined by reference not just to X’s own internal features but also to the internal features of ‘other outcomes with which [those features] might be compared’ (229). But obviously the ‘other outcomes’ against which the features of a given outcome X might be compared often – not always, but often – number more than one. X can be compared against an outcome Y – or X can be compared against a distinct outcome Z. That point in mind, Temkin describes ECV as taking the bold view that the value of X may shift depending on just which of the ‘other outcomes’ we compare against X. Thus, X may have one value if X is considered in itself, another value if X is compared against Y, another value if X is compared against Z and still another value if X is compared against both Y and Z. As Temkin himself puts the point, for at least some distinct outcomes X, Y and Z, the value of X varies depending on whether X is ‘considered by itself’, or X’s ‘alternative is Y’, or X’s ‘alternatives are Y and Z’ (229).

But it should be noted that ECV isn’t a theory of how it might happen that we come to perceive, or understand, X to have a higher, or a lower, value, depending on whether the alternate outcome Y we happen to be comparing X against is itself worse, or better, in some way than X is. On such a view, it would be quite easy to say as well that the value of X may seem to shift depending on whether we happen to focus on how X compares against Y alone or how X compares against Z alone or how X compares against both Y and Z. ECV, instead, takes the stronger position. ECV is not a view about the psychology of the process of evaluation or even its epistemics; it’s not an account of apparent value – of how we might be swayed to evaluate a particular outcome in a particular way. It’s rather a theory of the factors that may contribute to making the value of a given outcome what it in fact is.

Nor does ECV assert that the true value of a given outcome itself depends on its apparent value – on, that is, whether we happen to focus on X alone or
on how X compares against Y or on how X compares against Z or on how X compares against Y and Z. That's just not a plausible view. We (sadly) can't increase the value of the outcome X that we have carefully compared against Y by refraining from comparing X against Z. (The most we can do is delude ourselves into thinking X is not so bad by fixating on how X compares against one alternative and putting blinders on as to how X may compare against still other alternatives.) If Z is an alternative to X, then the comparison of X against Z may be pertinent to the evaluation of X whether we happen to attend to that comparison or not.

Thus, according to ECV, the value of X is not a matter of which alternatives to X we happen to focus on but rather what the alternatives to X in fact are. Obviously, then, a critical component of ECV is going to be a criterion for just when a given outcome Y counts as an alternative to X.

This issue is not one that Temkin explicitly addresses. It is clear, however, that he does not intend what we can call the comparison set (my term, not Temkin’s) for a given outcome X to include all logically possible outcomes – to include, that is, all those outcomes that consistent with the demands of logic might have unfolded in place of X. If that were the view, then the value of X would never shift, and it would make no sense to say that X might have one value when X is compared against Y but another value when X is compared against both Y and Z since there would be no such thing as the comparison set consisting of just Y.

Rather, Temkin’s conception must be that the comparison set for a given outcome X is itself case-dependent and, specifically, is a matter of the various ‘options’ or (Temkin sometimes says) ‘alternatives’ that are ‘available’ to agents in the particular case, where the alternatives available to agents in a particular case include only those outcomes agents have the ability and resources to bring about. Alternatives to X thus are, for Temkin, outcomes agents really, not just logically, could have brought about in lieu of X.5

It seems, then, that we can say this:

For purposes of applying ECV, the comparison set for any given outcome X in any given case is the set of just those outcomes Y that are available as alternatives to X for agents in that case, where alternatives are available to agents in a given case only if agents have the ability and resources to bring about those alternatives in that case.6

5 The outcomes Temkin considers alternatives are, in other words, not simply logically or metaphysically possible worlds, but accessible worlds (relative to a certain time and history and collection of agents). See Feldman 1986.

6 I don’t think in the end we can be happy with this way of understanding ‘comparison set’ unless we relativize to times; the outcome an agent has the ability to bring about in 2000 may not be an outcome that same agent has the ability to bring about in 2013. In this article, however, we shall not bother relativizing to times since the issues that arise here can be addressed without that work.
Clearly, then, there is here no suggestion that, according to ECV, a particular outcome X both has a particular value of n and does not have a value of n. The view, rather, is that X has a particular value n in a case in which the comparison set for X is identified by the case as consisting (consists ex hypothesi) of just (say) Y, while X may have a value distinct from n in a distinct case in which the comparison set for X is identified as consisting of (say) both Y and Z.

The traditional view, of course, is that facts about what alternatives happen to be available to agents at a given time and in a given case have to do with evaluating acts rather than outcomes. Temkin thus here departs from tradition. The alternative outcomes agents have the ability and resources to bring about may help to define, not just the permissibility of their choice of X or of their performance of an act that at the world in question gives rise to X, but also the value of X itself.

But how could this work? How could the value of an outcome change depending on which outcomes happen to be available to agents in a given case? Theorists who assume that IAV provides an exhaustive account of outcome goodness are going to say that it can’t. Temkin, however, sees things differently. The reason that X’s value may shift is that different ‘factors’ for evaluating X may come into play as the comparison set for X itself shifts. Thus: the ‘relevance and significance of the factors for determining an outcome’s value will vary depending on the alternatives with which it is compared…’ (229). A factor – a consideration, an element of the axiological analysis – that plays no role in the evaluation of X in a case in which the comparative set for X consists just of Y may play a critical role – may, that is, be both ‘relevant’ and ‘significant’ – in a case in which the comparative set for X consists of not just Y but Z as well.

What sorts of ‘factors’ have this characteristic? Clearly they do not include such things as total utility or degree of equality or human flourishing. Limiting our purview to just such factors as those, we may well consider IAV obviously correct and ECV obviously misguided. Temkin’s subsequent discussion, however, at least suggests that the sorts of ‘factors’ that may be relevant and significant in evaluating a given outcome relative to one comparison set but not relative to another are (at least) often person-affecting in nature. ECV, he says, ‘reflect[s] a general conception to which many are attracted. We can call this conception… the Narrow Person-Affecting View’ (416).

Temkin’s critical insight is this. Consider two cases, one in which the comparison set for X includes just Y and the other in which the comparison set for X includes both Y and Z. Suppose that in both cases our aim is to evaluate how X compares against Y. In both cases, Y exists as an alternative to X for purposes of our evaluation of X. But in the second case alone Z exists as an alternative to X as well. In both cases, to evaluate how X compares against Y for purposes of the person-affecting view we shall need to ask
(among others) the following question: does Y make any particular person p worse off than p is in any other member of the comparison set (which includes X)? (Is a person p who exists in Y better off in any non-Y outcome, including X, than p is in Y?) In the first case, where just Y exists as an alternative to X, our answer will depend on no more than how p fares in X and Y. As long as p is no worse off in Y than p is in X, we can stop worrying about p in connection with our evaluation of X. (We can, in other words, ignore p.) But in the second case, where both Y and Z exist as alternatives to X, it isn’t going to be enough to determine that \( p \text{ is no worse off in } Y \text{ than } p \text{ is in } X \). We can’t answer the question until we have considered as well how p fares in Z. In the second case, in other words, we can’t stop worrying about p in connection with our evaluation of how X compares against Y until we have determined, in addition, that \( p \text{ is no worse off in } Y \text{ than } p \text{ is in } Z \).

This is Temkin’s welcome twist on – indeed, his correction of – many statements of the person-affecting view: that Z, and how p fares in Z, can matter morally to how X is itself to be compared against Y.

Now, if my understanding of how the person-affecting analysis Temkin is suggesting is to proceed is correct, then it is not clear that Temkin needs as a technical matter to insist that ‘the factors that are relevant and significant for assessing an outcome can vary depending on the alternatives with which it is compared’ (434). We could instead say that the factors in determining how X is to be compared against Y remain constant from one case to the next but that the answers we give to the question of whether those factors are at play will vary depending on the case. So: in both cases, the ‘factor’ of whether there is an alternative in which p is better off is one we need to take into account in evaluating how X compares against Y. But in one case our evaluation is that there isn’t such an alternative and in the other case our evaluation is that there is.

Regardless, we get the idea. The fact that a person p has a wellbeing level of n in Y may be completely unimportant to our evaluation of how X compares against Y in one case – in a case where, that is, there is no Z and p’s wellbeing level in Y has been maximized – while that very same fact – that p has a wellbeing level of n in Y – may function in a critical way to our evaluation of how X compares against Y in another case – may, that is, count heavily against Y and hence in a roundabout way in favour of X in the case in which there is such a Z and p is worse off in Y than p is in Z.

According to ECV, then, the value of an outcome X can shift depending on X’s comparison set, which itself may shift from case to case. And we can now be more specific: according to ECV, the value of an outcome X can shift depending on how particular people fare in X and in each other member of the comparison set for the particular case.
So far, I have underlined the *essentially comparative* nature of Temkin’s person-affecting account. I now turn to the part of the view that is distinctly person-affecting in nature.

At first inspection, Temkin’s view may seem in line with earlier, *divisional* accounts, accounts that insist that people in one class (for example, the class of all *actual* people) matter morally whereas people in some other class (for example, the class of all *merely possible* people) do not. Thus Temkin’s account divides people into two classes, *independently* existing people – that is, where X is an outcome that is being evaluated in terms of its worse-ness relative to certain alternate outcomes, those who exist independently ‘of one’s choices’ (those, we can say, who exist in X and in each member of the comparison set for X) – and *dependently* existing people, those ‘whose existence … depends on the choices one makes’ (those, we can say, who exist in X or in at least one member of the comparison set for X and fail to exist in X or in at least one member of the comparison set for X) (417).

But at this point Temkin then distinguishes his own approach immediately departs from the divisional approach. Thus, in the case of independently existing people, we ‘should’, he writes, have ‘the aim of wanting them to be as well off as possible’ – we should (we might say) strive to maximize wellbeing for them. And, in the case of dependently existing people, we ‘should’ ‘ignore’ their ‘status’ – with the very important exception that we should also ‘avoid harming them as much as possible’ (417). Why say we should ‘ignore’ them at all, when we hurry on then to say that we should take care not to harm them? Because ‘ignore’ is exactly what the person-affecting approach calls on us to do in the classic two-option case where the question is limited to whether bringing an additional person p into an existence worth having makes things better: in determining whether the outcome X that excludes p is worse than the outcome Y that includes p, we should screen out as morally insignificant to our assessment the fact that p has an existence worth having in Y.

Because Temkin asserts that, for *independently* existing people, we should worry about *maximization*, while for *dependently* existing people, it is enough to worry about *harm*, it may still seem that Temkin means to suggest two distinct standards for how we ‘should’ conduct ourselves with respect to these two classes of people. Closer inspection, though, shows that that’s not the case. According to Temkin, a person is *harmed* at an outcome X only if ‘there is at least one available alternative outcome in which that very same person exists and is better off …’ (417). Of course, a necessary condition on

---

7 See note 1 above.

8 Note that here Temkin abandons the simple counterfactual test for harm which philosophers would have done well to abandon (as the law did) decades ago. According to that test, an act harms a person only if, had that act not been performed, that person would have been better off.
harm isn’t a definition of harm. But the particular necessary conditionTemkin proposes is perfectly consistent with a broad, maximizing account of harm, an account according to which, if agents have the ability and resources to make things better for a person than things are in the outcome under scrutiny – if, that is, the comparison set for that outcome includes at least one alternate outcome that is better for that person – then that person is harmed. Whatever additional necessary conditions Temkin might in the end find suitable for a complete account of harm, it seems that we will end up with one – quite stringent – standard for how we ‘should’ treat both independently and dependently existing people alike.

One other note. Temkin’s remarks on harm do not explicitly provide that a person can be harmed at an outcome X only if that person exists in X. (What he explicitly says is, as noted above, just that a person can’t be harmed in one outcome unless that same person exists at some other and is there better off.) It is, however, almost certain that Temkin means for his reader to assume that further necessary condition is also in place. However, since independently existing people are guaranteed to exist in all the outcomes under scrutiny, this second necessary condition, just like the first, does not serve to make any interesting distinction between independently and dependently existing people.

The upshot is that Temkin’s view on harm at least provides him with the space he needs to avoid the conceptual issues that have afflicted many prior person-affecting accounts. According to his view, we should strive to maximize wellbeing for independently existing people, where the question of bringing them into existence is off the table in view of the fact that they exist in each of the alternative outcomes. As to dependently existing people, where the question of bringing them into existence is very much on the table, we should avoid harming them – we should, that is, maximize wellbeing for them – in any scenario in which they do (or will) exist. That is: that a person exists and is harmed will count against the outcome in which that particular scenario arises. But ‘harm’ without existence won’t, given the concept of harm Temkin seems to be working with here: ‘harm’ without existence is something that can’t happen and hence can’t count against any given outcome.

It might be objected that the broad, maximizing account of harm Temkin seems to adopt in this context distorts our ordinary use of the term ‘harm’. My own view is that this issue is of little interest for Temkin’s project. His goal is not to provide an ordinary language or commonsensical analysis of how we humans use the term ‘harm’ in the many contexts in which we use (or misuse) it (16). I thus propose that we put that objection aside. Theorists unable to do that are, of course, free to think of Temkin’s ‘harm’ as mere

9 Thus: ‘[Y]ou can only harm someone by bringing her into existence, if there was another alternative available which you might have brought about instead…’ (422).
shorthand for the idea that a person has less wellbeing in one outcome in which that person does or will exist than that same person has in an alternate outcome in which that person does or will exist. (A fuller account of harm would need to incorporate notions of causation and perhaps probability. But it’s not clear to me that that fuller account would abandon the basic maximizing idea that Temkin’s explicit necessary condition on harm contemplates.)

In any case, however we choose to phrase things, Temkin is making an important point here: the fact that $Z$ is better for a given person $p$ than $Y$ is fair game for the person-affecting theorist to bring to bear in his or her person-affecting assessment of $X$ (or $Y$ or $Z$).

Temkin’s intent is to outline the person-affecting view, not to complete the account. He does not try to say what we should do when we can’t make each and every independently existing person ‘as well off as possible’ or how much ‘harm’ a dependently existing person must incur before we are justified in failing to make an independently existing person ‘as well off as possible’. Still, the outline that Temkin sketches does seem – for the most part, with one critical exception we will come back to in what follows – to be exactly right.

2. Wrongful life

The standard Temkin sets for dependently existing people – that we should avoid harming them – is quite stringent in light of the fact that the account of harm Temkin seems to be working with is itself broad and, indeed, maximizing. But is that standard stringent enough? Is Temkin’s account of harm broad enough?

According to Temkin, for a person $p$ to be harmed in a given outcome $X$ there must be ‘at least one available alternative outcome [Y] in which that very same person exists and is better off…’ (417 (emphasis added)). Presumably the requirement that the person exist in the alternative outcome $Y$ reflects Temkin’s own allegiance to the idea that one cannot be better off in $Y$ than one is in $X$ unless one is there, in $Y$, to be better off. But we should ask whether there really is no cogent way of constructing the intuition that a person whose life is nothing but anguish from beginning to end would have been better off never having existed at all.\(^\text{10}\)

After all, surely it can happen that a person whose life is and will continue to be anguished is better off dying now than living another 20 years. And we can surely imagine that whatever causes that anguish might have had its effect from the very moment that person commenced existing. Consider, then, the statement that the person $p$ who suffers the anguished existence

\(^{10}\) I have argued for this view (which I call Comparability) in Roberts 2003a, 2011b and 2011c.
in X would have been better off never existing at all – the statement, that is, that the outcome Y in which p never exists at all is better for p than X is. Where is the failure of cogency in that statement? Why should we assume that there exists no plausible metaphysics (whether actualist or not) that can provide a way of constructing that statement that renders it both meaningful and true?

The assumption is one that Temkin himself provides no argument for. That assumption in place, Temkin’s necessary condition on harm – that the anguished person p both is better off and exists in some alternate outcome – remains unsatisfied in the context of wrongful life. But if bringing the anguished person p into existence does not harm p, then Temkin’s person-affecting account is left with no basis on which to generate the result that X is morally worse than Y is. Temkin thus concludes that a complete account of outcome-worseness will involve, not just the essentially comparative view that the person-affecting account itself represents, but an internal aspects view – an impersonal account – as well (422–23).

I suggest, however, that – absent further argument, and given that the claims that run counter to Temkin’s assumption seem both natural and intuitive – we would do well to set that assumption aside. I suggest that we broaden Temkin’s account of harm to recognize that, in the rare case in which the person’s existence itself is anguished – is, that is, less than worth having, or genuinely wrongful – an alternate outcome in which that person never exists at all is better for that person. The person-affecting account can then nicely address the problem of wrongful life. We can say that the anguished person p is harmed in outcome X in which p suffers the anguished existence; that that harm is something we should have avoided; and that X is indeed morally worse than the outcome Y in which p never exists at all.

Now, taking the position that a person can be harmed even if there is no outcome in which that person is both better off and exists may seem a distinctly non-person-affecting strategy. If, after all, never existing can make an anguished person better off, then surely it can also make a happy person worse off. If existence can harm an anguished person, surely never existing can harm a happy person. But if we should avoid harm when its victim is a dependently existing anguished person and the harm is done when we bring that person into existence, it may seem that we are compelled to say as well that we should avoid harm when its victim is a dependently existing happy person and the harm is done when we leave that person out of existence.

11 Peter Singer suggested this argument in discussion following a DeCamp seminar (approximately) three years ago. I have elsewhere argued that we have no basis for excluding from our analysis on grounds of cogency person-affecting claims of the form this person would have been better off never existing at all. See Roberts 2003a.

12 Sider 2002; McMichael 1983.
The problem is that that final result would instantly sever Temkin’s (adjusted) person-affecting account from its intuitive roots, including Narveson’s witticism that ‘we are in favor of making people happy, but neutral about making happy people...’ (Temkin, 418). I believe, however, that we can avoid that result. Consistent with Temkin’s (adjusted) person-affecting account, we can instead adopt that I have elsewhere called Variabilism. According to that view, some differences in wellbeing level for a person between one outcome and another – for short, some harms, or losses – have moral significance while others do not, with the difference between the two being a function of whether the harmed person does or will exist in the outcome in which he or she is accorded the lower rather than the higher of the two wellbeing levels. On this view, the harm the anguished person suffers in the wrongful life context has full moral significance. But the harm of never existing at all when the alternative is an existence worth having has no moral significance whatsoever. The morally significant harm makes an outcome morally worse; it counts against that outcome and indeed against the choice that brings that outcome about. The harm that is devoid of any moral significance is morally neutral; it may as well never have taken place at all.

We thus can adjust Temkin’s person-affecting account in a way that enables it to give a plausible account of wrongful life. And we can do so without abandoning the original Narvesonian insight.

3. The mere addition paradox

3.1 Broome’s objection

His own account of the person-affecting account in hand, Temkin easily disarms Broome’s inconsistency objection against the person-affecting approach. My goal in this section is to argue that he could just as easily have disarmed the mere addition paradox.

Broome thinks that what he calls the ‘neutrality intuition’ lies at the heart of the person-affecting approach. According to the neutrality intuition, the addition of a person to an outcome is (other things equal) morally ‘neutral’ – it makes things, that is, neither better nor worse – provided that that person’s wellbeing level falls into what Broome calls the ‘neutral range’. Dissecting this seemingly simple thought just a bit, we can easily identify two distinct propositions. First, additional people existing in the neutral range do not make a given outcome better. And, second, additional people existing in the neutral range do not make a given outcome worse.
Broome’s objection to the neutrality intuition is based on a case involving three distinct outcomes (427):

- Outcome I contains a ‘large group of people [a1, a2…an] all of whom are very well off at level 1,000’ (420). Call them the A people.
- Outcome II contains all the A people at that same level 1000 plus an additional large group of people b1, b2…bn all of whom have lives that are ‘well worth living, but only at level 250’. Call them the B people.
- Outcome III contains all the A people still at level 1000 as well as all the B people now at level 750.

Schematically, where boldface indicates that the indicated person does or will exist and italics that the indicated person never exists, and where it is assumed that never existing at all is worth 0 units of wellbeing to the indicated person:

Broome’s argument is as follows. Suppose that wellbeing levels of both 250 and 750 fall within the neutral range. The neutrality intuition then implies that the addition of the B people to II makes II neither better nor worse than I; I and II are equally good. By the same reasoning, so are I and III. By symmetry and transitivity of the equally as good as relation, II and III are equally good. But any plausible person-affecting view will contain principles that go beyond the neutrality intuition and generate the result that III is better than II. (Here, a basic Pareto principle would do that work.)

Broome concludes that we must reject the neutrality intuition. Now, we might, alternatively, question – or, given his history, at least expect Temkin to question – whether the difficulty is in not the neutrality intuition per se but rather our assumption that the equally as good as relation is transitive. But that is not in fact the move Temkin makes. Rather, he shows how his

---

Figure 1. Broome’s case.
particular account of the person-affecting approach can nicely avoid Broome’s objection without rejecting transitivity.

Thus, according to that account, in evaluating I – and, in particular, how I compares against II – we must consider not just the features of I and II but the features of III as well. When we do that, we note that III is better for the B people than II is and indeed that the B people are harmed in II.

Temkin’s insight into how the person-affecting approach should be understood is here on display. Thus harm to the B people in II is not something we can discern by simply comparing I against II. If that’s all we do, we will miss something very important – something ‘significant’ and ‘relevant’ – about II. We’ll miss the fact that the B people need not have had a wellbeing level of (only) 250 and could just as easily have had a much higher wellbeing level. The comparison against III, in other words, shows that a fact about II that might initially have seemed morally neutral – that the B people have a wellbeing level of (only) 250 – is actually morally critical.14 We have no objection to – no less than stellar evaluation of – the outcome in which, after a terrible accident, the surgeon has saved a child’s life but not his ability to walk – until we learn that the surgeon could just as easily have done both.

14 ‘It is true that on [ECV’s account of the person-affecting approach] II is at least as good as I if those alternatives are considered alone…. But, on [that same account] the factors that are relevant and significant for assessing II are very different depending on whether or not II is compared with I alone, III alone, or both I and II together. Correspondingly, advocates of the [person-affecting view] can deny that there is a failure of the transitivity of ‘as good as’ in Broome’s example, and that this is so whether or not the different alternatives are considered pairwise or all at once’ (433). Notably, even in the case where III itself is an option, Temkin need not be taken to suggest that the comparison we make between I and II is anything other than a pairwise comparison. What Temkin is suggesting, rather, is that relevant and significant to that pairwise comparison between I and II is the fact that III itself exists as – but check with author. an option that has features that make certain features of II relevant and significant in a way they are not in the case in which I and II alone are options.

Temkin’s account of the person-affecting approach is similarly successful in avoiding charges of intransitivity in another important puzzle case, Parfit’s Tom, Dick and Harry case (428–32). All three options, A, B and C, seem equally good according to IAV. But it has been widely thought that the person-affecting view is compelled to say that they are not and instead that B is better than A, that C is better than B – and, still worse, that A is better than C. Since transitivity implies as well that C is better than A, the case has been taken to show that person-affecting approaches are compelled to reject transitivity. Temkin’s account of the person-affecting approach gives us, however, another take on the case. Thus: ‘[W]hen all three alternatives are considered at once, there are no independently existing people, so comparisons of the different alternatives will turn on the extent to which, if any, the dependently existing people are harmed by the different alternatives. But, as should be plain, from the standpoint of [the person-affecting approach] there will be a single dependently existing person in each outcome who will be harmed, and to the very same extent, no matter which outcome is brought about…. So the advocate of the [person-affecting view] can rightly deny that his view yields an intransitive ranking of [Parfit’s case]’ (432).
But having discerned this fact about II, we can then conclude that II is not after all equally as good as I is; rather, II is worse than I. More generally, it’s true that adding happy people can’t make an outcome better. But it can make an outcome worse.

Now, this result immediately needs to be squared against a seemingly compelling objection. If the B people are harmed in II, then aren’t they if anything harmed more in I? And if they are, then how can we even begin to think that II is worse than I? As noted earlier, Temkin himself almost surely assumes that the B people can’t be harmed in I because they never exist in I – and hence that it makes no sense to say that they are harmed more in I. My own reply to this objection would appeal to variabilism. We should concede that the B people are harmed in I and indeed are harmed more in I than in II but also take the position that that harm has no moral significance whatsoever. (This distinction is represented in Diagram 1, where all downward arrows indicate harms but broken downward arrows indicate harms devoid of any moral significance whatsoever.)

Either way – whether we say ‘no harm done’ in I or no morally significant harm done in I – what Temkin has shown is that we can give a person-affecting analysis of Broome’s case without relying on the neutrality intuition and without abandoning transitivity.15

3.2 Nontransitivity issue

Interestingly, Temkin himself finds one dark spot in this rosy picture. Even if Temkin’s account of Broome’s case is entirely consistent with the equally as good as relation being transitive, it does show, according to Temkin, that the relation, though transitive, should still be regarded as what he (non-standardly) calls nontransitive.16

To see what the concept of nontransitivity involves, consider a stripped down version of Broome’s original case in which III does not exist as an alternative outcome – a case, that is, in which the comparison set for I is just II. Temkin’s person-affecting account can be expected in this two-alternative case to support the result that I and II are equally good. But as we’ve

15 See Broome 2004: 145. My reply to Broome’s objection against the person-affecting view, though focused on the evaluation of acts rather than outcomes, closely tracks Temkin’s. I conclude that Broome’s argument against the neutrality intuition is effective but that the neutrality intuition does a poor job capturing the person-affecting approach. See Roberts 1998, 2003b and 2011a.

16 433. As he himself notes, Temkin’s use of the term ‘nontransitive’ is clearly not standard. The term ‘nontransitive’ is normally used – if used at all – to mean just what Temkin means by ‘intransitive’ (R is intransitive if for some a, b and c, Rba, Rcb and ¬Rca). Temkin, in contrast, explains ‘nontransitive’ as follows: R is nontransitive if either R is in fact intransitive or (we might add: R is transitive and) there’s a case in which Rba and another case in which Rcb and ¬Rca. It seems that a relation R would meet the second disjunctive condition just in those cases in which R doesn’t necessarily hold – so that we can’t infer, from the fact that Rba holds in one case, that it holds as well in the other.
already seen in the original case – the three-alternative case, where the comparison set for I includes both II and III – Temkin’s account also supports the result that I and II aren’t equally good but rather that II is worse than I. Moreover, according to Temkin, we have some sense that the results that pertain to the two-alternative case ‘should apply’ ‘across [both] sets of alternatives’ (433). ‘[M]ost would have initially thought that if they knew that II was as good as I, and that I was as good as III, then they could infer, by the transitivity of ‘as good as’, that II is as good as III’ (433). Temkin thus wants to say that, though Broome’s case doesn’t show the equally as good as relation to be intransitive, it does show that the relation has a certain paradoxical (my term, not Temkin’s) aspect to it. We are drawn to make a certain cross-case transitivity judgment that, on analysis and in the end, we understand to be mistaken. ‘It is only when we understand the nature and structure of Essentially Comparative Views that we understand that such inferences are illegitimate’ (433). We resolve the paradox when we realize that our results regarding the two-alternative case cannot be pooled with our results regarding the three-alternative case to obtain the still further result that II and III are equally good.

Temkin’s person-affecting account must, in other words, insist that the claims about outcome value it sanctions are relativized to the particular case, that is, to a particular comparison set. I and II are equally good, given a comparison set consisting of just II. And I and II aren’t equally good, given a comparison set consisting of II and III. That’s the hallmark of the essentially comparative analysis: ‘the relevance and significance of the factors for determining an outcome’s value will vary depending on the alternatives with which is compared, so that the value of an outcome X might be...o, when its alternative is Y, p, when its alternatives are Y and Z, and so on’ (229). A way, then, of putting Temkin’s point is just to say that certain transitive relations are to be understood as simultaneously nontransitive when we think the pertinent claims needn’t be relativized when in fact they must be relativized.

What isn’t clear is why we should consider the fact that equally as good as is nontransitive (in Temkin’s nonstandard sense) counts as an objection against Temkin’s person-affecting account or anyone else’s – or even as a worry or concern. After all, it seems that being equally tall as is also a nontransitive relation in Temkin’s sense. One could be drawn to make the mistake that, just because Tom and Dick happen to be the same height in one case, and just because Tom and Harry happen to be the same height in another case and it also happens in that other case that Harry is taller than Dick, that Harry must be taller than Tom. One might, that is, fail to keep in mind that there’s no guarantee that Dick is not six-feet tall in the one case and – perhaps because he drank too much coffee in his middle school years – a mere five-foot-ten in the other case. But we certainly don’t think that fact shows that we should have any lingering concern about the equally
as tall as relation or the claim that it is transitive. The nontransitivity of being equally tall as is not even worth mentioning.\textsuperscript{17}

Of course, in the case of some transitive relations, we really don’t need to worry about keeping our cases straight. Thus, being equal to, defined on a set of natural numbers, is not nontransitive, nor is having more utility in the aggregate, defined on a set of possible worlds. It would be highly mentionable if we identified a case, for example, in which a possible future X had exactly the same amount of utility in the aggregate as another possible future Y in one case but not in another.

Do we have reason to think that being equally as good as should be like height equality rather than like numerical equality? Should we think that the at least as good as relation between outcomes is necessary – that, if I and II are equally good in any case, then I and II are equally good in every case? Possibly, theorists for whom IAV is ingrained will think we should. Even theorists, like Temkin, who question whether IAV on its own can provide a complete theory of value outcome, might still find it psychologically difficult to wean themselves from that IAV-related theorem.

But none of that can count as an objection against the person-affecting approach. Moreover, it is not as though all theorists share Temkin’s sense that the pooling of our equally as good as results from distinct cases should be allowed – share Temkin’s sense, that is, that the relation ‘should apply’ ‘across certain sets of alternatives’ (433). For my own part, both the result that I and II are equally good in the two-alternative case and that I and II are not equally good in the three-alternative case strike me as highly intuitive – and indeed a tribute to Temkin’s own account of the person-affecting approach. Accordingly, I’m not in the least drawn to the notion that the result that I and II are equally good in the two-alternative case can be pooled with the results that I and III are equally good and that III is better than II in the three-alternative case to generate any troubling results at all.\textsuperscript{18}

\textsuperscript{17} In contrast, the unexpected logical features of the betterness relation that Temkin’s spectrum arguments seem to demonstrate are very much worth mentioning since in each of those arguments the anomalous facts all appear within a single case (see 45–52).

\textsuperscript{18} Now, as Temkin notes, there are ways of understanding the cases such that, even under Temkin’s own person-affecting account, we will find that the betterness relation is not only not intransitive but also not nontransitive. One such path would be to take what Temkin describes as a fine-grained view of how the various outcomes are themselves to be individuated (457–465). We can then say that if Y bears R to X in one case, and seems not to bear R to X in another case, what is really going on is that the ‘Y’ in the one case and the ‘Y’ in the other case are distinct possible outcomes. Thus, Y may bear R to X in one case, but it is Y’ that fails to bear R to X (or X’) in the other.

I elsewhere have elsewhere argued that that strategy is available to us in the face of the objection that the person-affecting approach fails to comply with the Axiom of the Independence of Irrelevant Alternatives. Roberts 2003b: §1.5. (For still other replies to that objection, see Roberts 1998, 74–77.) I just want to note here that surely the view one has to take in order to comply with the Independence Axiom really isn’t particularly...
3.3 A person-affecting resolution of the mere addition paradox

Temkin concedes that we should accept the transitivity of the *at least as good as* relation if we think outcome value can be fully accounted for within the confines of IAV. On that view, he writes, ‘the relevance and significance of the factors for determining an outcome’s value will not vary depending on the alternative with which the outcome is compared, and the ‘all-things-considered betterness than’ relation will be transitive’ (229). But he doesn’t think that outcome value can be fully accounted for within the confines of IAV. Rather, he thinks ECV plays a role as well. And he thinks, further, that that is going to mean that we must make room for the view that the *at least as good as* relation is not just nontransitive but intransitive as well (230).\(^{19}\)

Now, as discussed just above, Temkin does not think that Broome’s case, analysed in the way that Temkin’s own account of the person-affecting approach suggests, gives us any basis to think that the *at least as good as* relation is intransitive. But the mere addition paradox, according to Temkin, is a different story entirely. The question I want to raise here is why. If we say, with Temkin, that adding happy people makes things worse in Broome’s case, why not say that adding happy people makes things worse in the mere addition paradox as well?

Schematically:

Once Temkin concedes that the person-affecting approach will be at least part of our account of outcome value, there is an important question whether this particular case makes a plausible case for intransitivity. If the person-affecting account has the resources to analyse Broome’s case in a way that

![Figure 2. The Mere Addition paradox.](https://academic.oup.com/analysis/article-abstract/74/2/306/208135)

fine-grained at all, at least where what one is comparing for betterness in one possible world, or future, against another. Thus, it seems obvious that, on any plausible account of identity between worlds, the world where agents do have the option of bringing a particular child into a better existence and the world where they don’t have that option (and are able only to bring that child into a lesser existence or none at all) are distinct worlds.

\(^{19}\) See note 16 above.
preserves transitivity – and, as we’ve seen in Section 3.1, it seems that it does – we should expect it to have the resources to analyse the mere addition paradox in a way that preserves transitivity as well.

If the case were such that the only alternative to A is just A+, we might well say that A+ is at least as good as A. But it isn’t; B is an available alternative outcome – a member of the comparison set for A – along with A+. And it’s the availability of B that makes the existence of b1, b2, ..., bn at the wellbeing level of (only) 250 significant and relevant. On Temkin’s account, it’s the availability of B that shows that the B people are harmed in A+. And that fact, according to the person-affecting approach, must be taken into account in determining outcome betterness.

The person-affecting approach insists, in other words, that we take into account all three alternative outcomes – that we refrain, that is, from focusing just on A and A+ and putting blinders on for B. We then find support for the following results. B is better than A+ (it makes, that is, the tradeoff in wellbeing between a1, a2 ... an and b1, b2 ... bn in a way that is morally better than the way A+ does); and A is better than B (taking into account the person-affecting aim of wanting ‘independently existing people ... to be as well off as possible’ and worrying about dependently existing people only to the extent that they are harmed. Transitivity then implies that A is actually better than A+. 20

Now, it is this final implication that many theorists think gives rise to paradox. That’s so, since many earlier accounts of the case assume that ‘mere addition’ can’t make things worse – that is, that A+ is surely at least as good as A is.

Temkin himself has accepted that result – and has long argued that our best response to the paradox is to understand the betterness relation as intransitive. It’s to understand, in other words, that B is better than A+ and A is better than B but that A isn’t better than A+. And – he continues to endorse that position in this new book as well (416).

But I think that may be a mistake. As we’ve just seen, Temkin’s own account of the person-affecting approach clearly opens the door to a quite distinct solution. We can instead say, per the above analysis, that A+ isn’t as good as A is and is rather worse than A. 21

---

20 I am taking for granted here that Temkin assumes that the dependently existing people in B are not ‘harmed’ at all in A since they never exist in A (417). As noted earlier, I would put the point differently; I would concede that those people are harmed in A but argue that they are nonetheless harmed in a way that has no moral significance. Hence, in Figure 2, just as in Figure 1, downward arrows indicate harms, but broken downward arrows indicate that the particular harm has no moral significance whatsoever. See Section 2 above; see also Roberts 2011a and 2011c.

21 I have argued elsewhere in favour of this particular way of resolving the mere addition paradox, that is, in favour of the view that A+ is actually worse than A (Roberts 2003b: §1.6).
Why does Temkin deny himself this desirable fruit of the person-affecting approach? According to Temkin,

[O]n [ECV], A+ is not worse than A regarding utility, as it would have to be if ‘worse than regarding utility’ were a transitive relation. This is because...the extra people in A+ have lives that are well worth living, and their presence does not adversely affect the utility of the people who exist in both A and A+. That is, given that the people who exist in A are just as well off in A+, and given that the additional people in A+ have lives that are well worth living, there is no reason, on [ECV], to regard A+ as worse than A regarding utility. (416)

But his comparison between A and A+ at this particular juncture seems motivated not by ECV but rather by IAV. Temkin has perfectly described the internal features of A and A+. But he has – for reasons not made clear – refrained from noting how the presence in the case of B as an alternative available outcome changes – in his own terms – the ‘relevance’ and significance’ of the fact that b1, b2...bn have a wellbeing of (only) 250.22

This point can be put another way. In reaching the result that A+ is at least as good as A is, Temkin has not uniformly taken into account all the comparisons that his account of the person-affecting approach itself seems to contemplate. In putting blinders on with respect to at least some of those comparisons, he fails to deploy the full power of his own person-affecting approach.

This should be good news. A second look at the mere addition paradox has shown that we may well be able to avoid paradox and preserve transitivity. As noted earlier, Temkin’s account of the person-affecting approach in ethics is one of the best things about his book. The upshot of my discussion here is just that this brilliant and exciting book may be even better than Temkin thinks it is.

The College of New Jersey
Ewing, NJ 08638
robertsmt@tcnj.edu

22 Now, Temkin might respond that his account of the mere addition paradox is driven, not by ECV alone, but rather by a mix of ECV and IAV. Such a response would be credible since Temkin elsewhere very clearly categorizes himself as a pluralist. At the same time, pooling the results of different theories does not make for a very fair argument against transitivity. Of course, transitivity may fail – indeed consistency may fail – whenever we pool the results of unreconciled theories. And just how Temkin will go about reconciling his own allegiances to ECV and to IAV is yet to be determined. But what that means is that – at best – it’s yet to be determined whether we must in the end understand transitivity to fail or not.
References

Roberts, M.A. 2003a. Can it ever have been better never to have existed at all? Person-based consequentialism and a new repugnant conclusion. *Journal of Applied Philosophy* 20: 159–85.