



A fixed-population problem for the person-affecting restriction

Jacob M. Nebel¹

© Springer Nature B.V. 2019

Abstract According to the person-affecting restriction, one distribution of welfare can be better than another only if there is someone for whom it is better. Extant problems for the person-affecting restriction involve variable-population cases, such as the nonidentity problem, which are notoriously controversial and difficult to resolve. This paper develops a fixed-population problem for the person-affecting restriction. The problem reveals that, in the presence of incommensurable welfare levels, the person-affecting restriction is incompatible with minimal requirements of impartial beneficence even in fixed-population contexts.

Keywords Person-affecting restriction · Population ethics · Incommensurability · Anonymity

Under what conditions is one distribution of well-being better than another?¹
According to

¹ A distribution of well-being is a function from individuals to welfare levels, where each individual's welfare level represents how good that distribution is for her. I consider such distributions rather than total outcomes or possible worlds to bracket the truth or falsity of *welfarism*, according to which the goodness of an outcome depends only on its distribution of welfare. I assume that nonwelfarists would judge some welfare distributions better than others, while maintaining that outcomes with better welfare distributions need not be better overall. We could, alternatively, compare total outcomes or informationally richer distributions while imaginining that all axiologically relevant considerations other than welfare are equal, or by restricting our attention to the goodness of outcomes with respect to the value of well-being.

✉ Jacob M. Nebel
jnebel@usc.edu

¹ University of Southern California, Los Angeles, USA

The person-affecting restriction: For any welfare distributions X and Y , X is better than Y only if there is someone for whom X is better than Y .²

This principle has a ring of plausibility. Narveson (1967) seems to find it an obvious constraint on utilitarian moral thinking. But many philosophers, following Parfit (1984), reject it because of its implications in variable-population cases—i.e., cases in which different people, or different numbers of people, would exist in different distributions.

The most important challenge for the person-affecting restriction is the nonidentity problem, a version of which is illustrated in Table 1.

In Table 1, there are two distributions, A and B . In each distribution, only one person would exist: Ann in A , Beth in B . Ann's life in A would have positive welfare x . Beth's life in B would have higher welfare $x+$.

Many people believe that B is better than A —and, more generally, that if the same number of people would exist in each of two distributions, and if every person in one distribution would be better off than every person in the other distribution, then the one distribution must be better than the other. But this is hard to square with the person-affecting restriction. For there seems to be no one for whom B could be better than A . It cannot be better for Ann because Ann doesn't exist in B . And it cannot be better for Beth because, many people believe, it cannot be better for a person to exist than not to exist.

The person-affecting restriction seems to imply that B cannot be better than A , contrary to intuition. Therefore, many people believe, the person-affecting restriction must be rejected.

In response to the nonidentity problem, some philosophers say that it can be better for a person to exist with a happy life than not to exist (Holtug 2001; Roberts 2002; Arrhenius and Rabinowicz 2015; Fleurbaey and Voorhoeve 2015). On this view, the person-affecting restriction is compatible with the intuition that B is better than A , because there is someone—namely Beth—for whom B would be better. Others deny that B would be better than A , on the grounds that we cannot compare distributions in which different people would exist (Heyd 1994). This is a radical view, with highly counterintuitive implications. But every view in population ethics appears to have highly counterintuitive implications (Arrhenius 2000). So we should perhaps not dismiss it out of hand.

There are other objections to the person-affecting restriction. But most of them appeal to variable-population cases.³ This common feature is—dialectically, at

² By *better* (and *better for*) I mean better (for) *all things considered*, not just in some respect. This distinguishes the person-affecting restriction from what Temkin (1993, 256) calls *the slogan* (see note 3). The principle considered here is, in Parfit (2017)'s terminology, the *narrow telic* person-affecting restriction. I set aside its deontic analogue, as well as the so-called *wide* person-affecting restriction.

³ The person-affecting restriction, as understood here, is not committed to welfarism (see note 1). Objections to welfarism do not challenge the person-affecting restriction, understood as a restriction on the goodness of distributions of welfare. Radical egalitarians who embrace leveling down would reject the restriction. But moderate egalitarians, who reject leveling down, can accept the restriction while rejecting what Temkin (1993, 256) calls *the slogan*. The slogan says that a distribution cannot be better in *any respect* unless it is better in some respect for someone. Unlike the slogan, the person-affecting restriction

Table 1 The nonidentity problem

	Ann	Beth
A	x	
B		$x+$

least—a limitation, since such cases are notoriously controversial and difficult to resolve. It also leaves us in a position to wonder whether the person-affecting restriction would be plausible if confined to *fixed*-population cases—i.e., cases in which the same people would exist no matter what. According to

The fixed-population person-affecting restriction: For any welfare distributions X and Y in which the same people would exist, X is better than Y only if there is someone for whom X is better than Y .

I suspect that most philosophers would want to accept the fixed-population person-affecting restriction: it seems to capture a necessary condition for betterness with respect to welfare, and no one, to my knowledge, has identified any distinctive or interesting implications of the person-affecting restriction in fixed-population cases. Arrhenius and Rabinowicz (2012), for example, say that in fixed-population cases, the person-affecting restriction is “quite straightforward” and surmise it to be “widely accepted by theorists with welfarist inclinations.”

This paper advances the dialectic by identifying a problem for the fixed-population person-affecting restriction. My preferred response to the problem is to reject the fixed-population person-affecting restriction, so I frame it as an argument against the restriction, but I do not insist on this response. I suspect that many will share my preference, but there will remain some wiggle room for staunch proponents of the restriction. My primary aim here is just to lay out the problem and explain what we might learn from it.

The problem involves incommensurability in welfare. Two lives are incommensurable in welfare just in case neither is at least as good as the other. For example, the life of a great philosopher might be incommensurable with the life of a great artist. We might think neither of these lives to be better or worse than the other, nor to be equally good. For if they were equally good, then slightly improving or worsening one of these lives would make it better or worse than the other. Imagine, for example, that our great philosopher has asthma. This would make her life slightly worse. But, intuitively, it would not be enough to make her life worse than the life of a great artist without asthma. This is perhaps the most distinctive feature of incommensurability: slightly improving or worsening one of two incommensurable items is generally not enough to make it better or worse than the other (Raz

Footnote 3 continued

considered here is compatible with moderate egalitarianism. And, unlike other objections to the person-affecting restriction, the argument of this paper is compatible with the slogan.

1985; Chang 2002). The lives of great philosophers and great artists are so evaluatively different that asthma is not enough to break their incommensurability.

Incommensurability is generally thought to arise when there are tradeoffs between different dimensions of value and more than one rationally permissible way to weigh between those dimensions. Many theories of welfare will countenance such tradeoffs. This is most plausible for pluralistic theories, on which there are multiple ingredients of well-being. But it may even be plausible for monistic theories, such as hedonism: the relative importance of intensity and duration may be imprecise, and hedonic intensity might itself have multiple dimensions, instantiated in different ways by pleasures of different kinds.

I myself am not certain that there are incommensurable welfare levels. But I think there probably are, and I don't want to get into my reservations here. Here I simply assume that there are incommensurable welfare levels. I take it that this assumption would be shared by many proponents of the person-affecting restriction, and that those who would reject this assumption do so for reasons unrelated to the person-affecting restriction.

More specifically, I will assume that some possible lives have welfare levels x , y , $x+$, and $y+$, with the following evaluative structure: $x+$ is better than x ; $y+$ is better than y ; x is incommensurable with y and $y+$; and y is incommensurable with x and $x+$. For example, $x+$ might be the welfare of our great philosopher's life, $y+$ that of our great artist's, and x and y their welfare levels with asthma.

Now consider the distributions in Table 2. All three distributions share the same population of two people, Cat and Dana, whom I will imagine to be identical twins. In C , Cat's life has value x , and Dana's has y . In C' , they switch. Imagine, for example, that Cat's life in C' is, qualitatively, exactly (or almost exactly) like Dana's in C , and vice versa. In D , both lives are improved relative to C' , so that Cat is at $y+$, Dana at $x+$.⁴

It seems to me that D is better than C . To use our toy example, it would be better if Cat were a great philosopher and Dana a great artist than if they switched careers, and all other evaluatively relevant properties, but both had asthma.

This is incompatible with the fixed-population person-affecting restriction. By hypothesis, $x+$ isn't better than y , and $y+$ isn't better than x . There is no one for whom D would be better than C . Thus, the fixed-population person-affecting restriction implies that D cannot be better than C , contrary to intuition.

This seems to me a counterexample to the fixed-population person-affecting restriction. My intuition can be supported by the following argument.

First, C and C' seem equally good. These distributions of welfare are *permutations* of each other: that is, they assign the exact same welfare levels, but

⁴ The choice between C and D is inspired by Hare (2010) problem of "opaque sweetening" in the theory of rational choice. Distributions like C and C' are briefly considered by McCarthy et al. (2016) in another context; they do not consider implications for the person-affecting restriction.

Table 2 Problem for the fixed-population person-affecting restriction

	Cat	Dana
C	x	y
C'	y	x
D	$y+$	$x+$

to different people.⁵ And in neither distribution is anyone better or worse off than anyone else. According to

Minimal anonymity: Any welfare distributions X and Y in which the same people exist, in which no one is better off than anyone else, and that are permutations of each other, are equally good.

This principle seems to me a requirement of impartiality. The intuition behind minimal anonymity is that, from an impartial perspective, we should care equally about each person, and therefore ought to be indifferent between distributions that are permutations of each other—at least, in fixed-population cases in which no one would be better off than anyone else. Some person-affecting theorists (e.g., Parsons 2002) would deny that we must be impartial between, say, actual and merely possible people, and would therefore question the extension of anonymity to distributions with different populations. Others might have qualms with the anonymous evaluation of distributions in which some people are better off than others.⁶ Minimal anonymity raises none of these problems. I suspect that many person-affecting theorists would want to accept it (although, ultimately, I doubt they can).⁷

Second, D seems better than C' . This is because D is better than C' for both Cat and Dana. And, again, in neither distribution is anyone better or worse off than anyone else. According to

⁵ To be more precise, let \mathcal{I} be the set of individuals. Welfare distributions X and Y are permutations of each other just in case, for some bijection $\sigma(\cdot)$ from \mathcal{I} to \mathcal{I} , X assigns to each $i \in \mathcal{I}$ the same welfare level that Y assigns to $\sigma(i)$. Those who prefer to compare total outcomes or informationally richer distributions (as mentioned in note 1) can instead consider correspondingly richer permutations, which assign the same welfare levels *and* axiologically relevant nonwelfare properties (e.g., levels of desert) to each i and $\sigma(i)$. See, e.g., the weak anonymity axiom proposed by Blackorby et al. (2005) and the “desert-modulated anonymity” considered by Adler (2018).

⁶ The cases I have in mind involve (1) infinite populations in which one distribution is better for each person than a permutation of that distribution (as in Liedekerke 1995), (2) prospects that distribute chances of being better or worse off between people in different ways, but guarantee anonymously identical distributions (as in Diamond 1967), and (3) choices in which some people face much greater burdens than others (as in Brown 2019). I am inclined to maintain anonymity in all three cases, but others may disagree; minimal anonymity avoids these problems.

⁷ Some have suggested that minimal anonymity is flatly inconsistent with the spirit of the person-affecting restriction, which they take to include the claim that X and Y can be equally good only if there is someone for whom they are equally good. But many person-affecting theorists would want to say, for example, that $(x, x+)$ and $(x+, x)$ are equally good even though there is no one for whom they are equally good. So I do not take that claim to be implied by the spirit of the person-affecting restriction.

Minimal Pareto: For any welfare distributions X and Y in which the same people exist, and in which no one is better off than anyone else, if X is better than Y for each person, then X is better than Y .

I see this principle as a minimal requirement of beneficence. It would, I think, be accepted by all proponents of the person-affecting restriction. It seems bizarre to think that a distribution must be better for someone in order to be better, but to deny that a distribution that is better for everyone would be better.

If D is better than C' , and C' and C are equally good, then D must be better than C . This follows from the transitivity of *at least as good as*.⁸

I therefore conclude—from minimal anonymity, minimal Pareto, and transitivity—that D is better than C , contrary to the person-affecting restriction.

This argument shows that the implausible consequences of the person-affecting restriction are not confined to variable-population cases, such as versions of the nonidentity problem. It is incompatible with plausible fixed-population principles.

How might proponents of the person-affecting restriction respond?

Rejecting minimal Pareto is, I assume, off the table. Minimal Pareto seems to be a core commitment of any view that gives pride of place to the interests of people affected.

Some might question the possibility of incommensurable welfare levels. Although this is a respectable position, there is no particular reason to expect proponents of the person-affecting restriction to find it independently attractive. So I won't consider this possibility further here. It would be a surprising result if proponents of the person-affecting restriction were forced to reject the possibility of incommensurability in welfare.

Another option would be to deny the transitivity of *at least as good as*. After all, some person-affecting principles do seem to yield violations of transitivity (Temkin 1987). And there are independent arguments against transitivity (Temkin 1996; Rachels 1998). In my view, these arguments fail, and transitivity is sacrosanct.⁹ And, although some person-affecting principles straightforwardly violate transitivity, the person-affecting restriction (in either its fixed- or general-population form) is not one of them. If it were, I think that would be a decisive reason to reject it.

Rejecting minimal anonymity seems to me the least implausible option, and one that is continuous with at least some person-affecting thinking about the nonidentity problem. Some proponents of the person-affecting restriction are willing to deny, in Table 1, that B is better than A . They refuse to compare distributions in which different people would exist. These theorists might be skeptical of anonymity principles in general, since they give moral weight to the identities of particular people. The intuition in cases like Table 1 is that it doesn't matter which people exist, so we should want there to be better lives, rather than worse lives, even if they

⁸ Assume that D is better than C' and that C' and C are equally good. Then D is at least as good as C' , which is at least as good as C . So, by transitivity, D must be at least as good as C . Now suppose that C is at least as good as D . Then C' must be at least as good as D , by transitivity. But that is impossible, because D is better than C' . So C cannot be at least as good as D , but D is at least as good as C . Therefore, D is better than C .

⁹ For responses to the arguments of Rachels and Temkin, see, e.g., Pummer (2017) and Nebel (2018).

would be lived by different people. The intuition behind minimal anonymity is that it doesn't matter which of some fixed, existing population of people lead which particular lives, as long as there is no objectionable inequality. Perhaps the lesson of my argument is that the person-affecting restriction is in tension with even this weak requirement of impartiality restricted to fixed-population contexts.

Even if person-affecting theorists would be willing to reject minimal anonymity, that would be a surprising and important move in the development of person-affecting moral theories. Surprising, because the person-affecting restriction was initially proposed—by Narveson, at least—in a broadly utilitarian spirit, and minimal anonymity seems required by the utilitarian commitment to impartiality. Important, because it would drive a wedge between two welfare-related aims. It is standardly assumed that, in fixed-population contexts, there is no difference between maximizing aggregate well-being and maximizing the balance of benefits minus harms—understood as *differences* (i.e., gains and losses) in each person's well-being (Parfit 1976; McDermott 2018). That is why classical utilitarianism has been assumed to obey the person-affecting restriction in fixed-population contexts. It has also been claimed that, if a person can be benefited by coming into existence, then there would be no difference between maximizing aggregate well-being and maximizing the balance of benefits minus harms even in variable-population contexts (Holtug 1999). That is why classical utilitarianism has been assumed to *entail* the person-affecting restriction on the assumption that a life worth living is better for a person than her nonexistence (Ross 2015). The argument of this paper, however, shows how the aims of maximizing aggregate well-being and maximizing the balance of net benefits can diverge, even if we regard it as better for a person to exist. For a distribution's aggregate well-being can be assessed anonymously; it depends only on which lives are lived and how much well-being they contain, not on who lives them. The balance of net benefits, by contrast, depends on which people live which lives. These aims can come apart in the presence of incommensurability.¹⁰

However person-affecting theorists respond to this tension, the argument of this paper has an important methodological upshot: the person-affecting restriction—perhaps the most important issue in the debate between person-affecting and impersonal approaches to moral theory—cannot be evaluated, as it largely has been for the past fifty years, solely on the vexing terrain of variable-population ethics.

Acknowledgements Thanks to Cian Dorr, Laura Franklin-Hall, Johann Frick, Arden Koehler, Rob Long, Adam Lovett, Michal Masny, Sam Scheffler, Jake Zuehl, and two anonymous reviewers for helpful comments.

¹⁰ On how to make sense of maximizing aggregate well-being in the presence of incommensurability, see Sen (1970, 1980). A related problem arises for expected utility maximization; see, e.g., Dubra et al. (2004).

References

- Adler, M. D. (2018). Prioritarianism: Room for desert? *Utilitas*, 30(2), 172–97. <https://doi.org/10.1017/S0953820817000164>.
- Arrhenius, G. (2000). An impossibility theorem for welfarist axiologies. *Economics and Philosophy*, 16(02), 247–66.
- Arrhenius, G., & Rabinowicz, W. (2015). The value of existence. In I. Hirose & J. Olson (Eds.), *The Oxford handbook of value theory*. Oxford: Oxford University Press.
- Arrhenius, G., & Rabinowicz, W. (2012). Better to be than not to be? *Discusiones Filosóficas*, 13(21), 65–85.
- Blackorby, C., Bossert, W., & Donaldson, D. (2005). *Population issues in social choice theory, welfare economics, and ethics*. Cambridge: Cambridge University Press.
- Brown, C. (2019). “Is close enough good enough?” *Economics & Philosophy*. <https://doi.org/10.1017/S0266267119000099>.
- Chang, R. (2002). The possibility of parity. *Ethics*, 112(4), 659–88. <https://doi.org/10.1086/339673>.
- Diamond, P. A. (1967). Cardinal welfare, individualistic ethics, and interpersonal comparison of utility: Comment. *Journal of Political Economy*, 75(5), 765–66. <https://doi.org/10.1086/259353>.
- Dubra, J., Maccheroni, F., & Ok, E. A. (2004). Expected utility theory without the completeness axiom. *Journal of Economic Theory*, 115(1), 118–33. [https://doi.org/10.1016/S0022-0531\(03\)00166-2](https://doi.org/10.1016/S0022-0531(03)00166-2).
- Fleurbaey, M., & Voorhoeve, A. (2015). On the social and personal value of existence. In I. Hirose & A. Reisner (Eds.), *Weighing and reasoning: Themes from the philosophy of John Broome*. Oxford: Oxford University Press.
- Hare, C. (2010). Take the sugar. *Analysis*, 70(2), 237–47. <https://doi.org/10.1093/analys/70.2.237>.
- Heyd, D. (1994). *Genethics: Moral issues in the creation of people*. Berkeley: University of California Press.
- Holtug, N. (1999). Utility, priority and possible people. *Utilitas*, 11(1), 16.
- Holtug, N. (2001). On the value of coming into existence. *The Journal of Ethics*, 5(4), 361–84.
- McCarthy, D., Mikkola, K., & Thomas, T. (2016). *Utilitarianism with and without expected utility*. MPRA Paper No. 79315.
- McDermott, M. (2018). Harms and objections. *Analysis*. <https://doi.org/10.1093/analys/any062>.
- Narveson, J. (1967). Utilitarianism and new generations. *Mind*, 76(301), 62–72.
- Nebel, J. M. (2018). The good, the bad, and the transitivity of better than. *Noûs*, 52(4), 874–99. <https://doi.org/10.1111/nous.12198>.
- Parfit, D. (1976). Rights, interests, and possible people. In S. Gorovitz, A. L. Jameton, R. Macklin, J. M. O'Connor, E. V. Perrin, B. P. St. Clair, & S. Sherwin (Eds.), *Moral problems in medicine*. Englewood Cliffs, NJ: Prentice-Hall.
- Parfit, D. (1984). *Reasons and persons*. Oxford: Clarendon Press.
- Parfit, D. (2017). Future people, the non-identity problem, and person-affecting principles. *Philosophy & Public Affairs*, 45(2), 118–57. <https://doi.org/10.1111/papa.12088>.
- Parsons, J. (2002). Axiological actualism. *Australasian Journal of Philosophy*, 80(2), 137–47. <https://doi.org/10.1093/ajp/80.2.137>.
- Pummer, T. (2017). Spectrum arguments and hypersensitivity. *Philosophical Studies*. <https://doi.org/10.1007/s11098-017-0932-3>.
- Rachels, S. (1998). Counterexamples to the transitivity of better than. *Australasian Journal of Philosophy*, 76(1), 71–83. <https://doi.org/10.1080/00048409812348201>.
- Raz, J. (1985). Value incommensurability: Some preliminaries. *Proceedings of the Aristotelian Society*, 86, 117–34.
- Roberts, M. A. (2002). A new way of doing the best that we can: Person-based consequentialism and the equality problem. *Ethics*, 112(2), 315–50. <https://doi.org/10.1086/324321>.
- Ross, J. (2015). Rethinking the person-affecting principle. *Journal of Moral Philosophy*, 12(4), 428–61. <https://doi.org/10.1163/17455243-01204004>.
- Sen, A. (1970). Interpersonal aggregation and partial comparability. *Econometrica*, 38(3), 393. <https://doi.org/10.2307/1909546>.
- Sen, A. (1980). Plural Utility. In *Proceedings of the Aristotelian society*, New Series, 81 (ArticleType: research-article/Full publication date: 1980–1981/Copyright 1980 The Aristotelian Society) (pp. 193–215). <https://doi.org/10.2307/4544973>.

- Temkin, L. S. (1987). Intransitivity and the mere addition paradox. *Philosophy & Public Affairs*, 16(2), 138–87.
- Temkin, L. S. (1993). *Inequality*. New York: Oxford University Press.
- Temkin, L. S. (1996). A continuum argument for intransitivity. *Philosophy & Public Affairs*, 25(3), 175–210.
- Van Liedekerke, L. (1995). Should utilitarians be cautious about an infinite future? *Australasian Journal of Philosophy*, 73(3), 405–7. <https://doi.org/10.1080/00048409512346741>.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.