
The Value of Existence

2013-01-15

Gustaf Arrhenius & Wlodek Rabinowicz

Department of Philosophy, Stockholm University

Department of Philosophy, Lund University

Gustaf.Arrhenius@philosophy.su.se

Wlodek.Rabinowicz@fil.lu.se

“Not to be born at all is best, far best that can befall.” Sophocles, *Oedipus at Colonus*, line 1224.¹

“Gut ist der Schlaf, der Tod ist besser — freilich das beste wäre, nie geboren sein.” Heinrich Heine, *Morphine*, 1848, lines 15-16.

“Life is so terrible, it would have been better not to have been born. Who is so lucky? Not one in a hundred thousand!” Old Jewish saying.

I. Introduction

Can it be better or worse for a person to be than not to be, that is, can it be better or worse for her to exist than not to exist at all? This old and challenging philosophical question, which we can call *the existential question*, has been raised anew in contemporary moral philosophy. There are roughly two reasons for this renewed interest. Firstly, traditional so-called “impersonal” ethical theories, such as utilitarianism, have paradoxical and very counterintuitive implications in regard to questions concerning procreation and our moral duties to future, not yet existing people. Secondly, it has seemed evident to many that an outcome can only be better than another if it is better for someone, and that only moral theories that are in this sense “person affecting” can be correct. The implications of this so-called *Person Affecting Restriction* will differ radically, however, depending on which answer one gives to the existential question.

Hence, many of the problems regarding our moral duties to future generations turn around the issue of whether existence can be better or worse for a person than non-existence. Some think so, others adamantly deny it. Sigmund Freud, for instance, described the Jewish

¹ Translated by F. Storr (London, Heinemann; New York, Macmillan, 1912-13).

saying we have quoted above as a “nonsensical joke”.² Others, as illustrated by the quotes from Sophocles and Heine above, seem to have a different view. Thus, for example, Melinda Roberts (2003) and Matthew Adler (2009) have defended an affirmative answer to the existential question. Contrariwise, Derek Parfit (1984), John Broome (1999), and others have worried that if we take a person’s life to be better for her than non-existence, then we would have to conclude that it would have been worse for her if she did not exist, which is absurd: Nothing would have been worse or better for a person if she had not existed.

We shall suggest an answer to the existential question to the effect that one can claim that it is better or worse for a person to exist than not to exist, without implying any absurdities. First, however, we shall explain in more detail why this question has again moved to the forefront of moral philosophy. We shall then discuss some of the proposed answers in the literature and our own suggestion. Lastly, we shall consider and rebut some possible objections to our position.

II. The Person Affecting Restriction and the Existential Question

The Person Affecting Restriction, put as a slogan, states that an outcome can only be better than another if it is better for someone. The restriction has a strong intuitive appeal and it has been suggested that it is presupposed in many arguments in moral philosophy, political theory, and welfare economics.³ Moreover, several theorists have argued that the counterintuitive implications in population ethics of so-called “impersonal” welfarist theories could be avoided by adopting the restriction. This applies in particular to the well-known Repugnant Conclusion, which – as has been pointed out by Parfit – is entailed by classical utilitarianism.⁴

It is not easy to discern what exactly the distinction between “impersonal” and “person affecting” theories amounts to in the literature, partly because different authors have had a different take on the distinction and partly because other ideas have been conflated or mixed

² Freud (1960), p. 57 (quoted after Benatar (2006), p. 3). Freud tried to account for the nonsensicality of the joke by this observation: “who is not born is not a mortal man at all, and there is no good and no best for him.” (ibid.) His suggestion thus seems to be that the existential question requires a negative answer. However, the joke would still of course be nonsensical (and for that reason funny) even if the existential question were answerable in the affirmative.

³ See Temkin (1993a, b). The term “Person Affecting Restriction” was introduced by Glover (1977), p. 66, but see also Narveson (1967).

⁴ See Parfit (1984), p. 388. For an overview of these counterintuitive implications, see Arrhenius et al. (2006) and Arrhenius (2000, 2012). The Repugnant Conclusion is the claim that for any world inhabited by people with very high welfare, there is a possible world in which everyone has a life that is barely worth living which is *better*, other things being equal. Imposing the Person Affecting Restriction can block the derivation of the Repugnant Conclusion only if it is conjoined with a negative answer to the existential question. Then it is arguable that a world in which everyone has a life barely worth living cannot be better than a world consisting of individuals with very high quality of life, since the former is not better for anyone, not even for the people who exist in the former but not in the latter world. Since we are going to argue that the existential question should be answered in the affirmative, however, we are sceptical about this manoeuvre. Making population ethics more ‘person affecting’, so to speak, does not suffice to save it from counterintuitive implications (see Arrhenius 2009a, 2012).

with the Person Affecting Restriction. As has been shown elsewhere, one can interpret the restriction in a manner which makes it perfectly compatible with impersonal welfarist theories such as classical utilitarianism.⁵ Thus, it could be understood as an idea about what kind of facts moral goodness supervenes on, for example, that goodness is essentially related to the interests of human beings. We are, however, interested in a stronger reading of the restriction which stresses the individualist aspect of value even more by claiming that morality is essentially *person comparative*:

The Person Affecting Restriction: If an outcome A is better (worse) than B, then A is better (worse) than B for at least one individual.⁶

This is the principle that we shall henceforth refer to as the Person Affecting Restriction (or “the restriction” for short). In cases involving only the same people in the compared outcomes, this view is quite straightforward and, we surmise, widely accepted.⁷

In comparisons between outcomes involving different people, however, and in particular in cases involving people whose existence is contingent on our choices, the restriction becomes ambiguous. An outcome A is better than B for Peter if Peter has a higher welfare in A as compared to B. We can assume that much. But what if Peter exists in outcome A but not in outcome B? Is A then better than B *for Peter*? More generally, can existence be better or worse for a person than non-existence? In other words, what is the correct answer to the existential question? Hence, depending on the answer to the existential question we get different versions of the Person Affecting Restriction and very different implications regarding how to morally evaluate different possible futures.

⁵ See Arrhenius (2003, 2009, 2012).

⁶ An interesting question is whether the restriction should be supplemented with a person affecting necessary condition for outcomes being equally good. We would suggest the following condition: If outcome A is equally as good as B, then either A and B are equally as good for at least one individual, or A is better (worse) for at least one individual and B is better (worse) for at least one individual. What if both A and B are empty worlds? We think that it is in the spirit of the person affecting idea that such worlds are not ranked as equally good but rather that they completely lack value from a person affecting perspective.

⁷ The term “Person Affecting Restriction” might be misleading, since many theorists would, sensibly we think, weaken the restriction to also include other sentient beings. Cf. Holtug (1996). Notice that since the Person Affecting Restriction is formulated without any *ceteris paribus* clause, value pluralists are not likely to accept it since it leaves little room for other values apart from welfarist ones. For instance, one might believe in some non-welfarist values such as virtue, reward in accordance to desert (cf., for example, Feldman 1995a, b, 1997), beauty (cf. Moore 1903, section 50), variety of natural species, or what have you (for a general discussion of value pluralism, see Rabinowicz & Rønnow-Rasmussen 2004; for a discussion of this issue in connection with the Person Affecting Restriction, see Arrhenius 2003, 2009a, 2012). Moreover, certain welfarist theories might also be ruled out by the restriction, such as some versions of welfarist egalitarianism (Arrhenius 2009b, 2012). However, we shall only discuss implications of the restriction in cases where one can assume that other values are not at stake. Hence, the arguments below also apply to a *ceteris paribus* version of the restriction.

III. Neither Better nor Worse to Be than Not to Be

A popular answer to the existential question is to claim that existence cannot be better or worse than non-existence for a person, nor equally as good for that matter, since existence and non-existence are, in some sense, incomparable in value for a person. David Heyd seems to endorse this position when he argues against the view that existence could be worse than non-existence by claiming that such a view “is inconsistent with a person-affecting theory as it presupposes the comparability of non-existence with life of a certain quality”.⁸

In his early pioneering work in population ethics, Narveson seems to share Heyd’s concern, although he formulates it in terms of happiness comparisons rather than comparisons in value:

If you ask, “whose happiness has been increased as a result of his being born?”, the answer is that nobody’s has. - - - Remember that the question we must ask about *him* is not whether he is happy but whether he is happier as a result of being born. And if put this way, we see that again we have a piece of nonsense on our hands if we suppose the answer is either “yes” or “no”. For if it is, then with whom, or with what, are we comparing his new state of bliss? Is the child, perhaps, happier than he used to be before he was born? Or happier, perhaps, than his alter ego? Obviously, there can be no sensible answer here.⁹

Similarly, Alan Buchanan et al. claim that “when the alternative is nonexistence, there is no individual who is made worse off by being conceived and born”¹⁰ and John Broome states that “...it cannot ever be *true* that it is better for a person that she lives than that she should never have lived at all”.¹¹

The negative answer to the existential question in combination with the Person Affecting Restriction has such counterintuitive conclusions that it is hard to believe that anyone would seriously endorse the conjunction of these two views. Consider the Future Bliss or Hell Case:

⁸ Heyd (1988), p. 161. See also Heyd (1992), pp. 124-5. Heyd states that his view is “grounded in an ‘anthropocentric’ conception of value according to which value is necessarily related to human interests, welfare, expectations, desires and wishes – that is to say to human volitions” (1988, p. 164).

⁹ Narveson (1967), p. 67 (emphasis in original). Cf. Dasgupta (1995), p. 383: “Recall our definition of the zero level of well-being. This isn’t a standard arrived at through a comparison with ‘non-existence’. Such comparisons can’t be made. The ‘unborn’ aren’t a class of people. It makes no sense to attribute a degree of wellbeing, low or high or nil, to the ‘state of not being born’”.

¹⁰ Buchanan et al. (2000), p. 234. This passage continues: “Nonexistence is not a condition that is better for an individual only in rare cases like having Lesh-Nyhan or Tay Sachs disease; it is no condition at all, and so it is not better or worse than any other condition” (ibid.).

¹¹ Broome (1999), ch. 10, p. 168 (emphasis in original). See also Parfit (1984), pp. 395, 489.

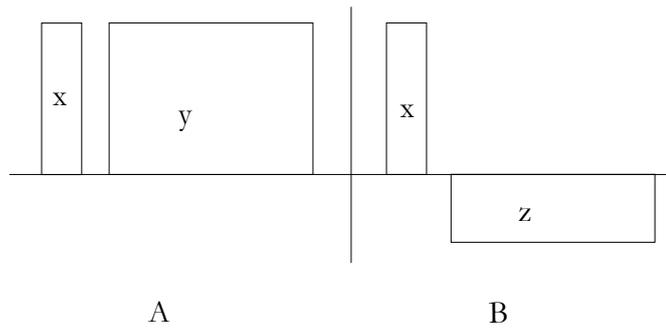


Diagram 1

The blocks in the above diagram represent populations. The width of each block represents the number of people in the corresponding population, whereas the height represents their welfare. Assume that we can either see to it that all the people in the future have excellent lives (the y -people in outcome A) or that they have hellish lives (the z -people in outcome B). Assume further that these two possible futures consist of different but the same number of people and that these two outcomes are equally good for us, the present x -people.

Most people, we presume, would consider outcome A clearly superior to outcome B and agree that we ought to realise A rather than B. However, since the y - and z -people are uniquely realisable (i.e. exist in just one of the two outcomes), the negative answer to the existential question implies that outcome A is neither better nor worse for the y - and z -people as compared to B. Moreover, the two outcomes are equally good for the x -people. Hence, according to the Person Affecting Restriction, A cannot be better than B since it is not better for any individual. Nor is of course B better than A. Consequently, if combined with the negative answer to the existential question, the Person Affecting Restriction ranks these outcomes as either equally good or as incomparable in value.¹² But that is clearly the wrong diagnosis of the Future Bliss or Hell Case.

This and other counterintuitive implications of the Person Affecting Restriction in combination with the negative answer to the existential question have led philosophers to

¹² The Person Affecting Restriction coupled with a negative answer to the existential question yields a position close to what we call *Strict Comparativism*. When comparing outcomes, one should only count the welfare of people who exist in both of the outcomes that are being compared and completely disregard the welfare of people who only exist in one of them. This seems to be, for example, Broome's take on the restriction: "Suppose [an alternative X] contains a certain number of people, and [an alternative Y] contains all the same people, and some more as well Then [the person-affecting view] is that [X] is at least as good as [Y] if and only if it is at least as good for the people who exist in both" (Broome (1992), p. 124). Broome rejects the restriction understood in this way but Heyd seems to accept it since he argues that "[e]xcluding the welfare and interest of future merely possible person ... is a necessary consequence of a coherent person-regarding theory of value" (Heyd (1988), pp. 159 – 61; see also Heyd (1992), pp. 124-5). See Arrhenius (2003, 2009a, 2012) for a discussion of different versions of Comparativism.

abandon the restriction (the majority) or to accept not only that existence can be better or worse for a person than non-existence but also that a non-existent person has a certain welfare level (namely, zero welfare) and that, consequently, non-existence can be better or worse for that non-existent person than a life at some specified level of welfare. As we shall show, both of these moves are uncalled for.

IV. The Argument from Absurdity

What is the reason behind the negative answer to the existential question? Well, one worry seems to be that if we give an affirmative answer to the existential question, that is, if we take a person's life to be better or worse for her than non-existence, then we would have to conclude that *it would have been worse or better for her if she did not exist*, which is obviously absurd: Nothing would have been worse or better for a person if she had not existed.¹³ This argument is eloquently stated by Broome:

...[I]t cannot ever be *true* that it is better for a person that she lives than that she should never have lived at all. If it were better for a person that she lives than that she should never have lived at all, then if she had never lived at all, that would have been worse for her than if she had lived. But if she had never lived at all, there would have been no her for it to be worse for, so it could not have been worse for her.¹⁴

However, this *Absurd Conclusion* (italicized above) does not follow. A triadic relation consisting in one state (having a certain life) being better for a person p than another state (non-existence) cannot hold unless all its three relata exist. Now, the states in question are abstract objects and thus can be assumed to exist even if they do not actually obtain. Consequently, the triadic relation in question can indeed hold as long as p exists. However, if a person is a concrete object, which is the received view (and, we surmise, the correct one), then this relation

¹³ Rabinowicz (2009a), fn. 2, ascribes this worry to Derek Parfit (1984), who writes: "in being caused to exist, someone can be benefited. (...) we need not claim that this outcome is *better* for that person than the alternative. This would imply the implausible claim that, if that person had never existed, it would have been worse for this person." (p. 395; cf. also p. 489). Parfit asserts, however, that causing someone to exist still can be *good* for the person in question. Good, but not better. It seems to us that Parfit needn't have been so cautious. His discussion of the matter contains all that is needed for the bolder betterness claim (see below).

¹⁴ Broome (1999), ch. 10, p. 168 (emphasis in original). Notice that this argument, if correct, would also work equally well against the idea that existence could be worse for someone than non-existence: If it were worse for a person that she exists than that she should never have existed, then it would have been better for her if she had never existed. If she had never existed, then there would have been no her for it to be better for, so it could not have been better for her. Thus, it cannot be true that it could be worse for a person to exist than not to exist.

could not hold if p weren't alive, since the third relatum, p , would then be missing.¹⁵ Consequently, even if it is better for p to exist than not to exist, assuming she has a life worth living, it doesn't follow that it *would have been worse* for p if she did not exist, since one of the relata, p , would then have been absent. What does follow is only that non-existence *is* worse for her than existence (since 'worse' is just the converse of 'better'), but not that it *would have been worse* if she didn't exist. Hence, Broome's argument is a non-sequitur and the Absurd Conclusion doesn't follow from the idea that existence can be better or worse for a person than non-existence.¹⁶

It might be that Broome assumes that the following general principle is true:

Subjunctive Connection 1 (SC1): An outcome A is better (worse, equally as good) for p than (as) another outcome B if and only if outcome B would be worse (better, equally as good) for p than (as) A if B came about. ¹⁷

Kristen Bykvist has suggested a similar principle which he calls *Accessibility*: "If A is better (worse) for S than B, then A would be better (worse) for S than B even if A obtained".¹⁸ However, as we pointed out above, it doesn't follow logically from "it is better for p to exist than not to exist" that "it would have been worse for p if she did not exist" since in the latter

¹⁵ On the other hand, if, contrary to the received view, a person were itself constructed as a collection of (abstract) states of affairs, then it would be correct to say that she would exist, as an abstract object, even if she didn't 'obtain', so to speak. Hence, one might then say that there is nothing absurd in claiming that, if she didn't obtain, this state could have been worse for her than her actual state, since all three relata would then exist as abstract objects. However, this interpretation of persons as abstract objects is a view that few philosophers would be prepared to accept.

It should be noted that in this paper we simply presuppose that relations require the existence of relata. This presupposition could, of course, be questioned. (We are indebted to Staffan Angere for pressing this point.) Thus, for example, it could be argued that intentional attitudes are best interpreted as relations between the subject and the intentional object, which would mean that a relation after all can obtain even when one of the relata (the intentional object) happens not to exist. On our view, such a relational account of intentional attitudes is unsatisfactory. An ontology that allows obtaining of the relations without existing relata leads to a whole host of problems, but the discussion of these issues would take us too far afield.

¹⁶ Rabinowicz suggested this argument already back in 2000 in personal conversation with Arrhenius, Broome, Bykvist and Erik Carlsson at a workshop in Leipzig; and he has briefly presented it in Rabinowicz (2003), fn. 29, and in more detail in Rabinowicz (2009a), fn. 2. For a similar argument, see Arrhenius (1999), p. 158, who suggests that an affirmative answer to the existential question "only involves a claim that if a person exists, then she can compare the value of her life to her non-existence. A person that will never exist cannot, of course, compare "her" non-existence with her existence. Consequently, one can claim that it is better ... for a person to exist ... than ... not to exist without implying any absurdities." In fact, even though he accepted the negative answer to the existential question (and instead went for the view that it can be good but not better for a person to exist than not to exist), Parfit (1984) came very close to making the same point as we make when he observed that there is nothing problematic in the claim that one can benefit a person by causing her to exist: "In judging that some person's life is worth living, or better than nothing, we need *not* be implying that it would have been worse for this person if he had never existed. --- Since this person *does* exist, we can refer to this person when describing the alternative [i.e. the world in which she wouldn't have existed]. We know who it is who, in this possible alternative, would never have existed" (pp. 487-8, emphasis in original; cf. fn. 13 above). See also Holtug (2001), Bykvist (2006) and Johansson 2010.

¹⁷ As we have seen, the only-if part of this principle is assumed by Parfit (1984), pp. 395, 489.

¹⁸ Bykvist (2007), p. 348. Symbols have been changed in this quotation, for the sake of consistency.

case one of the relata, p , would be absent. Moreover, it seems clear that SC1 is false, *mutatis mutandis*, for related concepts, such as “considered better by” / “would be considered worse by” and “preferred by” / “would be dispreferred by”.¹⁹ So it is not clear to us why one should go for SC1 rather than for the following connection between “better for” and “would be worse for”:

Subjunctive Connection 2 (SC2):

- (i) If a person p exists in both outcomes A and B, then A is better (worse, equally as good) for p than (as) B if and only if B would be worse (better, equally as good) for p than (as) A, if B obtained.
- (ii) If a person p exists in A but not in B, then A can be better (worse, equally as good) for p than (as) B although B would not be worse (better, equally as good) for p than (as) A, if B obtained.

Of course, one might find SC1 more attractive than SC2, perhaps because one finds it more in line with common language use: If we consider one outcome as being better for someone than another outcome, then we are normally prepared to conclude that the other outcome would be worse for that person (and not just that it *is* worse).²⁰ So we acknowledge that there might be a price to pay here, in terms of departure from common usage and our common way of thinking, for our answer to the existential question. However, as long as no other reason for SC1 has been brought forward, we find the price worth paying. Moreover, if SC1 is accepted, then one would have to give up the idea that existence can be better or worse than non-existence for a person, since with SC1 as an extra premise, Broome’s argument would be valid and the Absurd Conclusion would follow from an affirmative answer to the existential question.

Notice that our argument is not based on any revision of the logic of “better for” and “worse for”. In one of his earlier contribution to this topic, Nils Holtug seems to suggest such a revision to avoid the Absurd Conclusion:

¹⁹ This is acknowledged by Bykvist (2007), p. 349.

²⁰ Indeed, as Johan Brännmark has pointed out (in private communication) the connection between indicative ascriptions of properties and relations and subjunctive claims about what would happen under hypothetical circumstances is a pervasive feature of our language and our way of thinking about the world. When I say, for example, that the medicine I take decreases blood pressure or that exercise keeps me fit, I imply something about what would be the case if I didn’t take my medicine or if I didn’t exercise. We of course agree with this observation.

There is a clear sense in which existence can be better for a person than nonexistence, even if nonexistence is not worse for her (a person can have no properties in a possible world in which she does not exist).²¹

Holtug thus seems to suggest that one can avoid the Absurd Conclusion by revising the logic of “better for”: One can hold that it can be better for a person to exist than not to exist, but deny that the opposite is worse for her. It is clear that a state X is better than a state Y if and only if state Y is worse than state X (this seems to us to be a conceptual truth, if any). Holtug seems to deny that this logic also holds for “better for”, that is, that a state X is better for a person than another state Y if and only if state Y is worse for the person than state X. His reason is that “better for” and “worse for” are only applicable when a person to which the “for” refers to exists.

See also his (1996), p. 77:

When saying that a person has been benefited by coming into existence, I mean that this person is better off than if he had never existed. Of course, normally, if a person is better (worse) off in a situation X than in a situation Y, he is worse (better) off in situation Y. While this is normally true, it is not true when Y involves his nonexistence. And there is a perfectly natural explanation for that. The property of “being worse off”, like other properties, does not apply to people in worlds in which they do not exist.²²

In our view, there is no need for a revision of the logic of value comparisons. If A is better for p than B, then it trivially follows that B is worse for p than A. What does not follow is that B *would* be worse for p if it obtained, for p might then be missing. Revising the logic of “better” and “worse” is unnecessary unless one conflates “worse for” with “would be worse for”.²³

However, our proposal confronts another, general problem which is not as easy to deal with. We have assumed that the better-for relation obtains between abstract states and a person. But an obvious objection is the following: What is better, or good, for a person or

²¹ Holtug (1998), p. 171.

²² On another reading of this passage, which is closer to its actual wording, Holtug here only denies that being *better off* in X than in Y entails being *worse off* in Y than in X. However, even this suggestion seems to us unmotivated: It is strictly incorrect to say that a person who in X has a life worth living is better off in that state than she is in the state Y in which she does not exist (it is another matter that X is better for her than Y). Comparisons of how “well off” a person is in two different states do seem to presuppose that she exists in both states that are being compared. See next section for a further discussion of this point.

²³ It should be mentioned that in his (2001), Holtug gives up on his earlier proposal and instead moves to a position similar to the one defended here, referring to personal communication with Rabinowicz.

impersonally, is not an abstract state but the *obtaining* of that state. To assume that states and not just their obtainings are good or better seems counter-intuitive, since the existence of such abstract states does not make the world better in any way. It is only their obtaining that can make the world better. Now, it is clear that we cannot claim that the obtaining of the state of p's existence is better for p than the obtaining of p's non-existence. The two obtainings in question in question and the person under consideration do not co-exist, indeed they cannot co-exist, and so they cannot stand in any relation to each other. As we have assumed throughout, a relation cannot exist without the existence of all the relata.

Actually, this problem is much more general and fundamental. It concerns not only betterness for a person, but also betterness period. We quite often want to say that one state, S, is better than another state, T, when S and T are incompatible with each other. But we cannot reduce this relation between abstract states to a relation between the *obtainings* of those states. We cannot say that if S were to obtain this would be better than if T were to obtain, if this is meant to be a relation between the obtainings of S and T, respectively. For if S and T are mutually incompatible, then if one of them obtains, the other does not. So the two obtainings cannot co-exist and consequently cannot stand in any relation to each other.

As this is a general problem, which does not especially arise for the comparisons of existence and non-existence, we might simply ignore it in this context. But perhaps we should say something about how we believe it could be dealt with. One might simply take the view that not only obtainings of states of affairs can be good or better, but also those very states themselves. On this reading, S is literally better than T (impersonally, or for a person). We have no need to try to reduce this relational statement to some statement about obtainings of these states. Similarly, on this view, abstract states are literally good or bad (again, impersonally or for a person). But how can we then deal with the objection that the world is not made any better by the existence of such states? Well, the obvious answer is to say that abstract states can be valuable but they lack contributive value. As Michael Zimmerman (REFERENCE) puts it: "One might ... contend that the way to avoid having to declare that our world is better simply for the existence in it of the state of affairs of everyone being happy is to say that what counts in the evaluation of a world is not what valuable states of affairs exist in it but what valuable states of affairs obtain in it."

Zimmerman himself rejects the view we have just suggested. He thinks it is inadequate: "Perhaps it is reasonable to say that a world's parts consist of those states of affairs that obtain in it, rather than of those that exist in it, so that it's no surprise that its final value derives only from the former. But this still leaves us with the problem of saying that states of affairs (including worlds) themselves have final value, even when they don't obtain." He doesn't explain however, why the latter problem he mentions is a problem to begin with. It seems to us that the view in

question is tenable provided we allow for the possibility of valuable entities that do not contribute in value to the worlds in which they exist.

V. The Argument from Welfare Level Comparisons

To save the Person Affecting Restriction from cases like the Future Bliss or Hell case, Melinda Roberts has suggested that we should accept not only that existence can be better or worse for a person than non-existence, but also the apparently absurd conclusion that, in cases like this, non-existence would be better or worse for a person. The reason is that according to Roberts a non-existing person has a certain welfare level, namely, zero welfare:²⁴

...Nora does not have any properties at all at any alternative at which she does not exist and ..., where Nora has no properties at all, all the properties that she does have — that empty set — add up to a zero level of wellbeing. --- It would have been better for Nora not to have any wellbeing at all — to have zero wellbeing — than to have the negative level of wellbeing that she in fact has. It would have been [better] for Nora ... never to have existed at all than it is for Nora to exist.²⁵

However, in our view it is quite nonsensical to ascribe any wellbeing level at all to a person in a state in which she does not exist. Wellbeing presupposes being. Having a zero degree of wellbeing is a property and property instantiations require the existence of property bearers.²⁶ Moreover, as we have shown above, one can endorse an affirmative answer to the existential question without being committed to affirming that non-existence could have been better or worse for a person and without assigning any welfare levels to persons who don't exist.

However, one might insist that the suggestion we make still doesn't make sense: that we cannot make sense of one state, A, being better for p than another state, B, if we cannot

²⁴ Adler (2009), p. 1506, tentatively embraces a similar position: "Existence can be better or worse for an individual than nonexistence. Nonexistence can be better or worse for an individual than existence. Where an outcome set contains potential nonexistents, their interests should be taken into account by assigning them a utility level of zero in the outcomes where they do not exist".

²⁵ Roberts (2003), pp. 168-9. Moreover, Roberts (1998), p. 64, writes that "I am thus supposing that it is at least possible that s has more well-being in a world in which s does not exist than s actually has. Suppose s 's existence in X is unavoidably *less* than one worth living ... and that s has, in any world in which s does not exist, a zero level of well-being. Under these conditions, s 's level of well-being at zero is actually *greater* than s 's well-being in X..." (emphasis in original). On the other hand, she there also claims that "[t]here is no need, ever, on my account of either wrongful life or the non-identity problem to assign a value, *even a value of zero*, to nonexistence for an individual who never exists in the world that is subject to appraisal" (Roberts (1998), p. 174-5, emphasis in original).

²⁶ This is a bit dogmatic, of course. Here, we gloss over the controversial issue of property ascriptions to fictional objects and an even more controversial issue of tropes, i.e., concrete property-like particulars whose existence is supposed to be ontologically independent of the existence of property bearers.

compare the wellbeing levels of p in the two states in question. This might be what Heyd and Narveson have in mind in the quotes above.²⁷ Likewise, when Bykvist claims that SC1 (his Accessibility principle) is true about any “interpretation of ‘better for’ that is conceptually linked to well-being...”, it seems that his idea is that “better for”-claims are analyzable in terms of comparisons between well-being levels possessed by a given individual in different outcomes. This would entail SC1 given that no individual has any level of wellbeing in an outcome in which she does not exist. The idea is that there is a necessary connection between “better for” and “has a higher welfare than”:²⁸

Welfare Level Connection (WLC): An outcome A is better for a person p than another outcome B if and only if p has higher welfare in A as compared to B.

But again, p would not have any welfare level at all in a state in which she would not exist.²⁹ However, it seems to us that “better for” comparisons can be made without comparisons of welfare levels. Consequently, one should reject the suggested tight connection between “better for” and comparisons of welfare levels as expressed by WLC.

VI. Guardian Angels and Fitting Attitudes

Instead of relying on WLC, one might explicate “better for” in terms of what a benevolent impartial observer would choose for a person when she is only considering what is in the interests of the person in question, or – better – in terms of what that person’s guardian angel would be willing to choose.³⁰ According to this view, an outcome A is better for a person than another outcome B if and only if this is what her guardian angel would choose for her sake. If a person exists in the two compared outcomes, then trivially the guardian angel will choose the state in which her charge has the highest welfare level. However, if the guardian angel has a choice between bringing her charge into existence with negative welfare or not bringing her

²⁷ See also the quote from Dasgupta in fn. 9.

²⁸ See Bykvist (2007), p. 348. Adler (2009), p. 1503, considers a similar conceptual connection, which he presents as the connection between “worse for” and “worse off than”. However, unlike Bykvist, Adler’s discussion leads him to reject, at least tentatively, this supposed conceptual link (cf. *ibid.*, p. 1505).

²⁹ This seems to hold even if we were to construct persons as abstract objects that can obtain or not obtain. A specific welfare level is something an abstract person can possess only in a world in which she obtains.

³⁰ Rabinowicz suggested the guardian angel approach in 2000 (see fn. 16) and Arrhenius (2003) proposes the benevolent impartial observer approach. See also Bykvist (2006). Broome (2004), p. 63, credits Rabinowicz with a suggestion that is simpler but less plausible: A history (or a world) X is better for p than a history Y if and only if p prefers X to Y. As Broome points out: “A person may prefer one history to another even if she does not exist in both of them” (*ibid.*). Obviously, however, this simple proposal is not satisfactory as it stands. The advantage of appealing to the preferences of the guardian angel rather than to those of the benevolent impartial observer is that the latter are supposed to track what is impersonally good (good, period) rather than what’s good for the person under consideration. A benevolent impartial observer tracks impersonal goodness even when she only focuses on the interests of that person and of no one else. The task of a guardian angel is different in that respect.

into existence at all, she would choose the latter. Moreover, if the guardian angel had the choice between bringing her charge into existence with a positive welfare or not bringing her into existence, she would choose the former. Or so it may seem, at least.

We can think of this idea of a guardian angel as just a *criterion* for the “better for”-relation. On this criterial interpretation, we can try to find out what is better for a person by putting ourselves, in imagination, in her guardian angel’s shoes and then trying to determine what our preferences would be in that hypothetical position. Additionally, on a view that is philosophically more far-reaching and radical, the idea of a guardian angel can also be seen as a metaphor for a certain *analytical* proposal. More precisely, on this reading, we should take it as an application to “better for” of the so-called *fitting-attitudes analysis of value*. Along the lines of this format of analysis, we could say that

A is better for *p* than B if and only if one ought to prefer A to B for *p*’s sake.³¹

This analytic proposal could be made to work provided we can make some sense of locutions such as “preferring A to B for *p*’s sake”.³² Again, it seems reasonable to say that in the choice between bringing *p* into existence with negative welfare or not bringing her into existence at all, one ought to prefer the latter for *p*’s sake. Likewise, in the choice between bringing *p* into existence with positive welfare or not bringing her into existence at all, one ought to prefer the former for *p*’s sake.

On both these interpretations, the criterial and the analytic one, if a person *p* has higher welfare in an outcome A as compared to another outcome B, then A is better for *p* than B, but

³¹ Cf. Darwall (2002) for this proposal. As Darwall puts it: “[W]hat it is for something to be good for someone *just is* for it to be something one should desire for him for his sake, that is, insofar as one cares for him” (p. 8). See also Toni Rønnow-Rasmussen (2007), where this fitting-attitudes account of value-for is elaborated and defended. That this account can be used to clarify comparisons between existence and no-existence has been suggested in Rabinowicz (2009a), fn. 2.

³² The challenge here is whether the “for *p*’s sake”-locution can be independently understood, without presupposing the notion of “better for” as already given. If preferring something for *p*’s sake just means “preferring it insofar as one only cares for what is better for *p*”, then the analysis becomes circular. On the other hand, if “for *p*’s sake” is given an independent interpretation, then it is not obvious that all that one ought to prefer for *p*’s sake *is* better for *p*. In particular, moral considerations (and perfectionist ideas in general) complicate matters at this point. Thus, for example, it might be argued that one ought to desire for *p*’s sake that *p* experiences sorrow at the thought of others’ suffering. But it is not obvious that such a set of mind would be good for *p*. (We are indebted to Johan Brännmark for pressing this point.) Maybe, therefore, a circular analysis of “sake” would, after all, be preferable! It should be noted that even circular analyses can be instructive to some extent: They can be used to exhibit structural connections between concepts appearing in the analysans and the analysandum. Thereby, they can provide relevant information to those who already possess the concepts involved but are not clear about their mutual relationships. Thus, to take an example, David Wiggins adheres to the sentimentalist version of the fitting-attitudes account even though he explicitly recognizes the charge of circularity. Still, as he argues, the account is informative in its “detour through sentiments” (See Wiggins (1987), p. 189. Cf. Rabinowicz & Rønnow-Rasmussen (2004, 2006). The circularity Wiggins has in mind is different from the one mentioned here, though. He thinks that it might be essential to the fitting sentiments with regard to objects that these attitudes themselves already involve evaluations.

the reverse doesn't always hold. Hence, there is a connection between "better for" and "has higher welfare than" but this connection isn't as tight as WLC would have it.

As a matter of fact, the fitting-attitudes analysis of value can be used to define the very notion of wellbeing. To say that p has a positive (negative, zero) wellbeing in a given outcome A presupposes that p exists in A and means that to exist in A is better (worse, equally as good) for p than (as) not to exist at all, with the latter claim interpreted on the lines of the fitting-attitudes account.

This format of analysis makes it also possible to account for interpersonal comparisons of wellbeing, of the form "p's wellbeing in A is greater than q's wellbeing in B ." To do it, though, we first need to re-interpret preference. Instead of treating it as a dyadic comparative attitude we need to think of it as a comparison between monadic attitudes. More precisely, that a person prefers A to B would on this re-interpretation be understood as a claim to the effect that the person in question favours A to a higher degree than she favours B .³³ This would mean that we end up with the following series of analyses:

A is better than B = A ought to be preferred to B = It is required that A is favoured to a higher degree than B is favoured.

A is better för p than B = A ought to be preferred to B for p 's sake = It is required that A is favoured for p 's sake to a higher degree than B is favoured for p 's sake.

A is better för p than B is for q = It is required that A is favoured for p 's sake to a higher degree than B is favoured for q 's sake.³⁴

p 's wellbeing in A is greater than q 's wellbeing in B = p exists in A , q exists in B , and A is better för p than B is for q .

VII. Person Affecting Restriction Revisited

As for the connection between "better" and "better for", the Person Affecting Restriction remains an attractive option. It does seem plausible to claim that, to the extent we focus on welfare, an outcome cannot be better than another outcome without being better for someone. While this restriction would lead to counterintuitive implications if combined with the negative

³³ This formulation is simplified and covers only one form of preference: when both alternatives are favoured, though to a different degrees. But preference for A over B might instead be a matter of disfavouring A to a lesser degree than B , or of favouring A but disfavouring or being indifferent to B , or – finally – of being indifferent to A and disfavouring B . Complete analysis of preference needs to cover all these four possible cases.

³⁴ It is at this point that the re-interpretation of preference as a comparison between monadic attitudes is crucial. That A is favoured for p 's sake to a higher degree than B is favoured for p 's sake is the thought that could not be expressed in terms of a dyadic comparative attitude with respect to A and B . That dyadic attitudes should be replaced by comparisons between monadic attitudes in the fitting-attitudes account of value relations has also other reasons. On this issue, see Rabinowicz (2009b).

answer to the existential question (see the case of Future Bliss or Hell above), we have argued in this paper that the existential question should be answered in the affirmative.

It should be noted, however, that even coupled with the affirmative answer to the existential question, the Person Affecting Restriction, as we have stated it above, leads to counterintuitive implications, unless it is appropriately weakened. The reason is that the betterness relation between outcomes does not require for its obtaining the actual existence of the affected persons. Persons enter as relata in the triadic ‘better for’-relation and therefore must exist for that relation to obtain, but they are not relata in the dyadic betterness relation that obtains between outcomes. This contrast between the triadic and the dyadic relations of betterness explains why the Person Affecting Restriction cannot be correct as it stands.

Thus, to give an example, consider a variant of the Future Bliss or Hell Case above in which only the x -people exist in outcome A (i.e., in this variant, outcome A does not contain any future y -people) while outcome B still in addition contains z -people that lead hellish lives, and suppose that outcome A is the one that actually obtains. The Person Affecting Restriction would imply, counterintuitively, that A is not better than B, since – as things actually are – there exists no one for whom A is better than B: The added people in the hypothetical outcome B, for whom A would have been better, do not actually exist. Intuitively, however, if A would have been better than B had B obtained (and B is not worse than A if A obtains, as in our case), then A *is* better than B irrespective of whether A or B obtains.³⁵ To solve problems like this, Holtug (2004) has argued that we should replace the restriction with a weaker version, which in our formulation runs as follows:

The Wide Person Affecting Restriction: If an outcome A is better than B, then A would be better than B for at least one individual if either A *or* B would obtain.

In the example above, it is the second disjunct of this weaker restriction that is applicable. Clearly, it is only this wide, disjunctive version of the restriction that deserves serious consideration.³⁶

³⁵ This counterfactual invariance of the dyadic betterness relation is possible only because its relata (outcomes) can be assumed to exist even if they do not obtain. By contrast, the triadic relation of ‘better for’ can only satisfy a weaker condition of counterfactual invariance: If A would have been better for p than B if B obtained, then A is better for p than B even if B does not obtain, *provided that p exists*.

³⁶ In fact, even this disjunctive version of the restriction might be too strong. In principle, it is conceivable that A is better than B because it would be better for some individual who only exists in the actual outcome but neither exist in A or B. For example, suppose that C is the actual outcome and p exists in C but not in A or B. Suppose further that p in C devotes her whole life to a certain goal x . For p , this goal is categorical: she wishes it to be realized even if it weren’t her goal in the first place. Although p would neither exist in A or B, let us suppose that x would be better realized in the former outcome than in the latter. Assume also that the realization of p ’s fundamental goal is of one of the things that determine how good an outcome is for p , and that in the case of categorical goals this applies even to possible outcomes in which p would never strive for the goal in question. It

It might seem that given the positive answer to the existential question the restriction doesn't have much bite. If one compares two outcomes A and B with non-identical populations, then the restriction won't have anything to say as soon as A contains some persons with positive well-being that don't exist in B and B contains some persons with positive wellbeing that don't exist in A.

However, the appearances are misleading. The weak restriction does have some force, irrespectively if it is coupled with a positive or negative answer to the existential question, since it, for example, rules out all welfarist theories which imply that a mere addition of lives with positive welfare can make a population worse. Prominent examples of such theories are Average and Critical-Level Utilitarianism.³⁷ To see this, consider the following case:

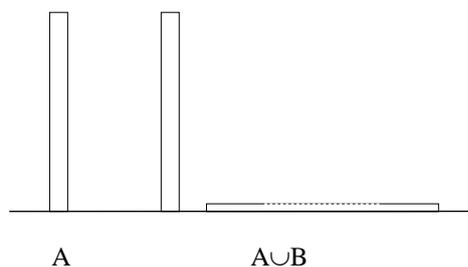


Diagram 2

According to Average Utilitarianism, $A \cup B$ is worse than A since the average welfare is lower in $A \cup B$ as compared to A. According to Critical-Level Utilitarianism, the contributive value of a person's life is her welfare minus a positive critical level and the value of a population is calculated by summing these differences for all individuals in the population. Assuming that the B-people are below the critical level, Critical-Level Utilitarianism reaches the same verdict as Average Utilitarianism. The addition of B-people has a negative contributive value given that their welfare is below the critical level.

However, $A \cup B$ would not not worse than A for anyone, irrespectively of whether A or $A \cup B$ would obtain. Consequently, the Weak Person Affecting Restriction rules out theories such as Average and Critical-Level Utilitarianism.

Before we finish, we should say more about ordinary language formulations of value comparisons between outcomes. The reader might have gotten an impression that, on our

then might well be the case that A is better for p than B, even though – as it happens – there would be no one in either A or B for whom the former outcome would be better than the latter. However, note that this would make better-for comparisons outcome relative, which one might find problematic. For example, if in outcome D p would have not- x as her fundamental categorical goal, then B would be better than A relative to D whereas A is better than B relative to C.

³⁷ For the latter theory, see Blackorby, Bossert, and Donaldson (1997, 1995, 2005) and Blackorby and Donaldson (1984). For a discussion of both theories, see Arrhenius (2013), ch. 3 and 5.

view, counterfactual claims such as “not to exist would have been better/worse for p than to exist” are absurd. But are such claims really so unpalatable? Think of Melinda Roberts’ statement about Nora: “It would have been better for Nora never to have existed at all than it is for Nora to exist.” This doesn’t sound absurd at all. But, if not, then perhaps the Argument from Absurdity doesn’t even get started?

Still, how should a counterfactual statement like the one about Nora be understood? Here is what we’d like to suggest. When we use ordinary language formulations of the form “ A would have been better/worse for p than B ” what we mean is something along these lines: (i) we state that A is better/worse than B for p , i.e., that a certain triadic relation obtains; *and* (ii) we imply that A does not obtain.

On this analysis of “ A would have been better/worse for p than B ,” or “ A would be better/worse for p than B ,” when we use such formulations, we don’t take a stand on what relation would obtain between A , B and p under counterfactual or potential circumstances, i.e. if A had obtained. It is this feature that explains the absence of absurdity in Nora-type statements. However, we do take such a stand when we expressly state that “ A would have been better/worse for p than B , *if C had been the case.*” Thereby we do state that the betterness/worseness relation would have obtained between A , B and p under the counterfactual circumstance C . Consequently, the following *is* absurd: “Not to exist would have been better/worse for p than to exist, *if p had not existed.*” It is this kind of statements that the Argument from Absurdity focuses on. Therefore, if that argument is to be invalidated, it has to be met head on, as we have done, rather than rejected on the grounds that there is no absurdity there to begin with.

VI. Summary

We have defended an affirmative answer to the existential question to the effect that one can claim that it is better or worse for a person to exist than not to exist, without implying any absurdities. Hence, not only is your existence, dear reader, better for us than your non-existence; it is also better for you.³⁸

³⁸ We would like to thank Matthew Adler, Staffan Angere, Johan Brännmark, Krister Bykvist, Erik Carlson, Marc Fleurbaey, Ingvar Johansson, and Toni Rønnow-Rasmussen for their helpful comments. We gratefully acknowledge financial support from the Bank of Sweden Tercentenary Foundation, the Swedish Research Council, and the Swedish Collegium for Advanced Study.

References

- Adler, M. D., "Future Generations: A Prioritarian View", *The George Washington Law Review*, Vol. 77, No. 5/6, September 2009.
- Arrhenius, G. *Population Axiology*, PhD -Diss., Dept. of Philosophy, University of Toronto, 1999.
- *Future Generations: A Challenge for Moral Theory*, F.D.-Diss., Uppsala: University Printers, 2000.
- "The Person Affecting Restriction, Comparativism, and the Moral Status of Potential People", *Ethical Perspectives*, no. 3-4, 2003.
- "The Repugnant Conclusion", *Stanford Encyclopaedia of Philosophy*, 2006.
- "Can the Person Affecting Restriction Solve the Problems in Population Ethics?" in M. Roberts & D. Wasserman (eds.) *Harming Future Persons*, Ashgate, 2009.
- *Population Ethics*, Oxford UP, forthcoming, 2012.
- Arrhenius, G., Ryberg, J., and Tännsjö, T., 'The Repugnant Conclusion', *Stanford Encyclopaedia of Philosophy*, <http://plato.stanford.edu/>, 2006.
- Benatar, D., *Better Never to Have Been: The Harm of Coming into Existence*, Oxford: Oxford University Press, 2006.
- Blackorby, C., and Donaldson, D. "Social Criteria for Evaluating Population Change", *Journal of Public Economics*, 25, pp. 13-33, 1984.
- Blackorby, C., Bossert, W., and Donaldson, D. "Intertemporal Population Ethics: Critical-Level Utilitarian Principles", *Econometrica* 65, pp. 1303-1320, 1995.
- "Critical-Level Utilitarianism and the Population-Ethics Dilemma", *Economics and Philosophy*, 13, pp. 197-230, 1997.
- *Population Issues in Social Choice Theory, Welfare Economics, and Ethics*, Cambridge UP, 2005.
- Broome, J., *Counting the Costs of Global Warming*, Cambridge: White Horse, 1992.
- *Ethics out of Economics*, Cambridge: Cambridge University Press, 1999.
- *Weighing Lives*, Oxford: Oxford University Press, 2004.
- Bykvist, K., "The Benefits of Coming into Existence", *Philosophical Studies* 135, 2007.
- Buchanan, A., Brock D. W., Daniels N., Wikler D. *From Chance to Choice: Genetics and Justice*, Cambridge: Cambridge University Press, 2000.
- Darwall, S., *Welfare and Rational Care*, Princeton, NJ: Princeton University Press, 2002.
- Dasgupta, P., *An Inquiry into Well-Being and Destitution*, Oxford: Oxford UP, 1995.
- Feldman, F., "Adjusting Utility for Justice: A Consequentialist Reply to the Objection from Justice", *Philosophy and Phenomenological Research*, Vol. LV, No. 3, pp. 567-85, 1995a. Reprinted in Feldman (1997).
- "Justice, Desert, and the Repugnant Conclusion", *Utilitas*, Vol. 7, No. 2, pp. 189-206, November 1995b. Reprinted in Feldman (1997).
- *Utilitarianism, Hedonism, and Desert: Essays in Moral Philosophy*, Cambridge: Cambridge University Press, 1997.
- Freud, S., *The Standard Edition of the Complete Psychological Works of Sigmund Freud*, Vol. VII, translated by James Strachey, London: The Hogarth Press, 1960.
- Glover, J., *Causing Death and Saving Lives*, Penguin Books, 1977.
- Heyd, D., "Procreation and Value: Can Ethics Deal With Futurity Problems?", *Philosophia* (Israel), No. 18, pp. 151-170, July, 1988.
- *Genethics: Moral Issues in the Creation of People*, Berkeley: University of California Press, 1992.
- Holtug, N., "In Defence of the Slogan", pp. 64-89 in W. Rabinowicz (ed.), *Preference and Value: Preferentialism in Ethics*, Studies in Philosophy, Dept. of Philosophy, Lund University, Vol. 1, 1996.
- "On the Value of Coming into Existence", *The Journal of Ethics*, Vol. 5, 2001.
- "Person-affecting Moralities", in J. Ryberg and T. Tännsjö (eds), *The Repugnant Conclusion. Essays on Population Ethics*, Dordrecht: Kluwer, 2004, pp. 129-61.

- Johansson, J., "Being and Betterness", *Utilitas* 22, 2010, pp. 285-302.
- Narveson, J. "Utilitarianism and New Generations", *Mind* 76, pp. 62-72, Jan 1967.
- Moore, G. E. *Principia Ethica*, Cambridge: Cambridge University Press, [1966] 1903.
- Parfit, D. *Reasons and Persons*, Oxford: Clarendon Press, [1991] 1984.
- Rabinowicz, W., "The Size of Inequality and Its Badness: Some Reflections around Temkin's Inequality", *Theoria*, 69, 2003, 60 - 84.
- "Broome and the Intuition of Neutrality", *Philosophical Issues* 19, 2009a, 389 - 411.
- "Values Compared", *Polish Journal of Philosophy* 3, 2009b, 73 - 96.
- Rabinowicz, W., and Rønnow-Rasmussen, T., "The Strike of the Demon: On Fitting Pro-attitudes and Value" *Ethics* 114, 2004, 391 - 423.
- "Buck-Passing and the Right Kind of Reasons", *Philosophical Quarterly* 56, 2006, 114 - 120.
- Roberts, M. A., *Child versus Childmaker: Future Persons and Present Duties in Ethics and the Law*, Rowman and Littlefield, 1998.
- "Can it Ever Be Better Never to Have Existed At All? Person-Based Consequentialism and a New Repugnant Conclusion", *Journal of Applied Philosophy*, Vol. 20, No. 2, 2003.
- Rønnow-Rasmussen, T., "Analyzing Personal Values" *The Journal of Ethics* 11, 2007.
- Temkin, L. S., *Inequality*, Oxford: Oxford University Press, 1993a.
- "Harmful Goods, Harmless Bads", pp. 291- 324 in R. G. Frey and C. W. Morris (eds.), *Value, Welfare, and Morality*, Cambridge: Cambridge University Press, 1993b.
- Wiggins, D., *Needs, Values, Truth: Essays in the philosophy of value*, Oxford: Blackwell, 1987.