

The Harmlessness of Existence

Per Algander¹ 

Accepted: 23 June 2017

© The Author(s) 2017. This article is an open access publication

Abstract Can existence benefit or harm a person? I argue that it cannot. In order for existence to harm a person it has to be the case that existence is worse for the person than never existing. This claim could only be true if it is understood as a claim about the actual, extrinsic value of existence for a person. However, understanding harm (and benefit) in terms of actual extrinsic value comes at the cost of depriving benefits and harms of their normative relevance. I show that a person who is guided by promoting actual extrinsic value can face situations where an outcome is extrinsically better for her but where the same outcome would be extrinsically worse for her were it to obtain. A person who is guided by promoting extrinsic value will in such situations not be able to deliberate about what she should do, prudentially or morally. I conclude that extrinsic value is therefore not something we should be guided by when deliberating about what we should do, and that if harm and benefit is understood in terms of extrinsic value, then we should not be guided by these notions either.

Keywords Harm · Benefit · Existence · Extrinsic value

1 Introduction

Can it be better, or worse, for a person to exist than to never have existed? Intuitively, it seems so. It is not difficult to come up with real or imagined cases where a person's life is so wretched that she thinks, and we with her, that it is worse for her that she exists than to never have existed. Similarly, a person who lives a very happy life may think, and we with her, that it is better for her to exist than to never have existed.

This philosophical puzzle is not without practical relevance. For example, can we *benefit* or *harm* a person by bringing her into existence? A common view is that in order for an

✉ Per Algander
per.algander@filosofi.uu.se

¹ Department of Philosophy, Uppsala University, Box 627, 751 26 Uppsala, Sweden

event to benefit (harm) a person it has to be the case that the person would have been worse (better) off had the event not occurred.¹ According to this view, bringing a person into existence therefore benefits or harms that person only if existence is better or worse for that person than non-existence.

The intuitive judgement that existence can be better or worse for a person than non-existence is not without its problems. First, if it is better (worse) for a person that she exists than that she never exists, then presumably it would have been worse (better) for this person had she not existed. But, had this person not existed then nothing—not even existence—would have had any value for her. The claim that it can be better or worse for a person that she exists than that she never exists therefore seems to commit us to the implausible claim that existence can be better or worse for *non-existent* people.

Second, if it is better (worse) for a person that she exists than that she never exists, then it seems plausible that the person would have been worse (better)-off had she not existed. However, had a person not existed then she would not have been well-off to any extent at all. Here, the problem is that we are forced to saying that non-existent people are well-off to some degree, when the more plausible view is that non-existent people are not well-off to any degree at all simply because they do not exist.

In this paper I will develop these two problems for the view that existence can be better or worse for a person than non-existence. I will argue that existence can be, in the actual world, *extrinsically* better (worse) for a person than non-existence but not *intrinsically*. By appealing to actual extrinsic value it is possible to avoid the two common objections to the view that existence can be better or worse for a person than non-existence mentioned above. However, I will also argue that this conclusion comes at a high price since actual extrinsic value is irrelevant to practical reasoning. This solution is therefore not very helpful as an answer to the practical puzzle about the harms and benefits of existence.

Before proceeding there are a couple of clarifications and assumptions to be made. I will understand notions such as “a person’s good”, “good for a person”, “prudential value” and the like in terms of well-being. What is good (bad) for a person should not necessarily be understood in terms of what that person *believes* to be good for him or her (she may be mistaken), nor should it be understood as necessarily saying something about what is good (bad) *simpliciter*.² Rather, I will assume that prudential value is, roughly, a kind of agent-relative value which we are typically guided by when pursuing our own interests or acting with someone else’s self-interests in mind.

A further clarification concerns the relata of the better for-relation. I will assume that the better for-relation is a three-place relation relating two states of affairs and a subject. In order for this relation to be instantiated in a world, all its relata must exist in that world. I will assume that states of affairs are abstract entities that can exist without obtaining, so even though the states of affairs *a exists* and *a does not exist* cannot, for obvious reasons, both obtain in the same world they can still exist in the same world. I will assume possible worlds to be maximally consistent sets of states of affairs thus conceived. Finally, the question I am discussing is whether it can be better or worse for a person to exist than to never have

¹This counterfactual analysis of harm and benefit has been suggested by Feinberg (1986) and Parfit (1984), among others. More recently the view has been defended by Bradley (2009), Klocksiem (2012), Feit (2013) and Purves (2015).

²The view about personal good which I am assuming can be contrasted with Moore’s view (1993, p. 150). According to Moore, the only thing that “my good” can mean is either that something is good (impersonally) and I possess it, or that it is good (impersonally) that I possess it.

existed *at all*. That is, I am not concerned with whether it can be better or worse for a person to exist at some time in a possible world, rather than to not exist at that time. The state of affairs *a does not exist* (in *w*) should be understood as the state of affairs that a particular person *a* does not exist at all in *w*, not that the person exists at some times but not others in *w*. In order to simplify the discussion, I will leave out the reference to *w*, unless it is required for clarity.

2 Axiological Invariance

The first problem for the view that existence can be better or worse for a person than non-existence has been clearly formulated by Broome (1999). According to Broome,

it 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. (Broome 1999, p. 168).

This argument rests on two premises that are especially noteworthy. First, Broome claims that if it were the case that existence is better for a person than non-existence, then non-existence would have been worse for the person than existence, had the person not existed. This is a claim to the effect that value relations are invariant across possible worlds. For example, if it is better for *a* to exercise rather than to not exercise, then it *would be* better for *a* to exercise rather than to not exercise, regardless of whether *a* does in fact exercise. Call this premise *Axiological Invariance*:

Axiological Invariance: If *S* is better (worse) for *a* than $\neg S$, then *S* would be better (worse) for *a* than $\neg S$ regardless of whether *S* or $\neg S$ obtains.³

From this premise it follows that if existence is better (worse) for a person than non-existence, then this relation holds regardless of whether the person exists or not.

The second noteworthy premise in Broome's argument is that if a person does not exist then nothing can be better or worse for that person. Furthermore, if a person would not have existed, then nothing would have been better or worse for that person. These claims are fairly straightforward consequences of the following condition:

Existence Requirement: If *a* exemplifies properties or stands in relations in a possible world *w*, then *a* exists in *w*.

It is easy to see how it follows from Axiological Invariance and the Existence Requirement that existence cannot be better (or worse) for a person, *a* than non-existence. If existence is better for *a* than non-existence, then non-existence would have been worse for *a* than existence if *a* had not existed. This claim is true if and only if, in the closest possible world where *a* does not exist, it is the case that existence is better for *a* than non-existence. However, in the closest possible world where *a* does not exist the 'better for *a*'-relation cannot be

³See Bykvist (2007, p. 348) who calls this principle *Accessibility*. The way Broome's argument is formulated he is in fact also assuming that if *S* is better for *a* than $\neg S$, then $\neg S$ is worse for *a* than *S*. I take this to be a trivial assumption.

instantiated because of the Existence Requirement (one of the relata, a , does not exist in that world). Therefore, existence cannot be better (or worse) for a person than non-existence.

Axiological Invariance has recently received a fair amount of criticism from a number of philosophers.⁴ A common theme among this criticism is that it does not follow from the claim that existence *is* better than non-existence for a person that non-existence *would have been* worse for that person. What follows is that non-existence *is* worse for the person, but this claim can be true without ascribing properties to things in worlds where they do not exist. For example, Arrhenius and Rabinowicz (2010) claim that a:

triadic relation consisting in one state (having a certain life) being better for a person [a] 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 [a] exists. However, [...] this relation could not hold if [a] weren't alive, since the third relatum, [a], would then be missing. (Arrhenius and Rabinowicz 2010, p. 405).

Arrhenius and Rabinowicz also note that Axiological Invariance is clearly false for certain other concepts. For example: *considered better by* and *preferred by*. If a person prefers S to $\neg S$, then it does not follow that the person would prefer S to $\neg S$ regardless of whether S is the case or not. Perhaps a would have different preferences depending on whether S or $\neg S$ is the case. Because Axiological Invariance is not a plausible restriction on such concepts, Arrhenius and Rabinowicz suggest that it should be rejected as a constraint on better (worse)-for as well.

This argument is not very persuasive. The concepts which Arrhenius and Rabinowicz pick out as similar to better-for are not as similar as some other concepts. To see this, note that it seems possible for a person to prefer things which are not good for her, and to judge things to be better for her even though they are in fact worse. A more striking resemblance when it comes to better for is not with *preferred by* but with *preferable*. However, Axiological Invariance looks at least as plausible regarding *preferable* as it does regarding *better for*.

A further objection against Axiological Invariance has been raised by Johansson (2010). Johansson argues that Axiological Invariance is not true of *extrinsic* value and that existence can be extrinsically better (or worse) for a person than non-existence. Extrinsic value is a broad category of value which covers different kinds of value which a thing has in virtue of its relational properties. This includes, for example, instrumental, signatory and contributive value. On Johansson's view, the state of affairs a exists is *extrinsically* better for a than the state of affairs a does not exist just in case the intrinsic value for a of the closest possible world where a exists is greater than the intrinsic value for a of the closest possible world where a does not exist.⁵ If a exists with a life worth living, then the closest possible world where the state of affairs a exists obtains, namely the actual world, has positive intrinsic value for a . What should we say about the value for a of the closest possible world where

⁴See Rabinowicz (2009, fn. 2), Adler (2008, pp. 1505-6), Arrhenius and Rabinowicz (2010), Johansson (2010) and Holtug (2010, pp. 140-42).

⁵In its general form the view is that the extrinsic value of S for a is the difference in intrinsic value for a of the closest possible world where S obtains and the intrinsic value for a of the closest possible world where S does not obtain. This analysis of extrinsic value resembles the analysis of extrinsic value suggested by Bradley (1998, pp. 115-20) and Zimmerman (2001, pp. 252-6).

the state of affairs *a does not exist* obtains? According to Johansson this world has zero intrinsic value for *a*:

In order to calculate the intrinsic value of a possible world for [*a*] – i.e. the intrinsic value it in fact has for [*a*] – we should look at the basic intrinsic values for [*a*] of the states of affairs that obtain in that world (i.e. the states of affairs that would obtain if that world were actual). It is hard to see why we should not say that a world has zero intrinsic value for [*a*] if all the state of affairs that obtain in that world have zero basic intrinsic value for [*a*]. Moreover, it is hard to see any reason to deny that every state of affairs that does not have positive or negative basic intrinsic value for [*a*] [...] has zero basic intrinsic value for [*a*]. (Johansson 2010, p. 295).

To see how Johansson's view works, suppose that *a* exists and that the intrinsic value of the actual world is positive for *a*. If the intrinsic value for *a* of the closest world where the state of affairs *a exists* obtains (the actual world) is greater than the intrinsic value for *a* of the closest world where the state of affairs *a does not exist* obtains then *a exists* is extrinsically better for *a* than *a does not exist*. If the latter has zero intrinsic value for *a*, then existence is extrinsically better for *a* than non-existence (given that the intrinsic value for *a* of the actual world is positive).

Exactly what the intrinsic value for a person *a* of a possible world where *a* does not exist is may of course depend on what theory of prudential value is true. For our purposes however it is sufficient to note that on most theories it will be either positive, negative, or neutral. According to hedonism, for example, the basic prudential value-bearers are hedonic states. A world which does not include *a* cannot include any hedonic states of *a*. Johansson's conjecture that such a world has zero intrinsic value for a person seems plausible on this view. According to other theories of prudential value however the value for *a* of a world where *a* does not exist need not be zero; it may be positive or even negative. Consider for example a preference-theory which holds that the basic bearers of prudential value are the objects of that person's actual preferences or pro-attitudes.⁶ If *a* actually has a preference for *S*, and *S* obtains in world *w* where *a* does not exist, then *w* contains something which is intrinsically good for *a* and is, to that extent, intrinsically good for *a*.

A possible objection to Johansson's view is that states of affairs in general do not have intrinsic value for a person, only obtaining states of affairs; i.e. facts. If facts are the only primary bearers of intrinsic value then an existing person's non-existence cannot have any value for that person, not even zero, since it is not a state of affairs that obtains. However, this argument is not conclusive. First, note that all Johansson's view requires is that non-obtaining states of affairs can have *extrinsic* value in virtue of the difference in intrinsic value between the closest possible world where the state obtains and the closest world where the state does not obtain. The view is therefore compatible with facts being the primary bearers of *intrinsic* value. Second, assuming facts to be the primary bearers of value, whether intrinsic or extrinsic, seems methodologically problematic since it might be impossible to formulate certain versions of preferentialism with this assumption. Taking states of affairs as the primary bearers of value is a more ecumenical approach since it does not seem to make a difference for other axiologies, such as hedonism.

⁶This kind of desire-theory should be distinguished from the view that, roughly, it is satisfied and frustrated preferences which are the basic bearers. See Rabinowicz and Österberg (1996) for an enlightening discussion of these two versions of preferentialism.

Johansson's view has several virtues. First, it can be used to give a straightforward explanation of the connection between a life worth living, a life worth not living, and the value of existence. A person's existence is on this view better for that person than her non-existence just in case the actual world has positive intrinsic value for the person. This allows us to give the following plausible characterisation of a life worth (not) living: a 's life is worth (not) living iff the intrinsic value for a of the actual world is positive (negative). Second, it does not follow from Johansson's view that existence can be better or worse for people who do not actually exist since it satisfies the Existence Requirement. Johansson's view therefore promises to deliver the intuitive verdict that existence cannot be better or worse for non-existent people, simply because they do not exist.

The objections raised by Arrhenius and Rabinowicz as well as by Johansson are therefore sufficient to cast serious doubt on Axiological Invariance. In particular, their objections show that Axiological Invariance should be rejected with respect to extrinsic value and, consequently, that Broome's argument does not show that existence cannot be extrinsically better (or worse) for a person than non-existence. We should therefore turn to the second problem for the claim that existence can be better or worse for a person than non-existence.

3 "Better for" and "Better off"

The argument based on Axiological Invariance is as we have seen not sufficient to show that existence cannot be extrinsically better (or worse) for a person than non-existence. However, it can be argued that the extrinsic value of existence runs into problems because of the relation between prudential value and well-being. Roughly, the idea is that existence cannot be extrinsically better (or worse) for a person than non-existence because, if it were, then the person would have been *better off* had she not existed. But, a person cannot be better-off not existing. Therefore, existence cannot be extrinsically better (or worse) for a person than non-existence. I will refer to this argument as "the argument from well-being".

Here is a way of spelling out this argument in more detail. Consider the claim S is *extrinsically better for a than* $\neg S$. In the introduction I stipulated that prudential value should be understood in terms of well-being. Given this stipulation, it is an attractive view that comparative prudential value can be understood in terms of degrees of well-being in the following way:

The Prudential Value Condition: S is extrinsically better for a than $\neg S$ iff a would have had more well-being if S were the case than if $\neg S$ had been the case.

That is, in order for the claim to be true we have to be able to compare two magnitudes of well-being: the well-being a would have had were S to obtain and the well-being a would have had were $\neg S$ to obtain. However, according to the Existence Requirement a person can only exemplify properties in worlds in which the person exists. It is therefore false that a person who does not exist in a world w has any amount of well-being—even zero—in w . The Prudential Value Condition together with the Existence Requirement therefore imply that existence cannot be extrinsically better for a person than non-existence.

Why is this relation between extrinsic value and well-being so attractive? One reason is that it coheres with a more general analysis of comparatives. Consider for example a is *taller than* b . A plausible view of the truth-conditions for this proposition is that it is true iff a has length m , b has length n , and m is greater than n . The view described above is attractive because it makes the better-for relation analysable in a similar way.

There are mainly two objections to the argument from well-being. The first objection is that this argument fails because a person who does not exist in a world w has zero well-being in that world.⁷ This is not a very promising reply however. A problem for this view is that it seems plausible that in order for a person to have any amount of well-being at a world w , even zero, that person has to instantiate some property in w . But, if this claim is true then the view that non-existent people have zero well-being is ruled out by the Existence Requirement.

A further problem is that this view would make well-being quite different from other properties in a number of ways. Consider what we would say about other comparatives such as length or weight. We would not say, for example, that a non-existent object is lighter than a feather. In order for it to be true that a non-existent object is lighter than a feather it has to be the case that the non-existent object has a certain weight, the feather has a certain weight, and that the former is less than the latter. It seems clear however that non-existent objects do not have any weight, *not even zero*. However, the view under consideration holds that this is not the case when it comes to well-being. For example, it is true that I am better off than my non-existent younger brother, on this view, because both me and my non-existent younger brother are well off to a certain degree, and the degree to which I am well off is greater than the degree to which my non-existent younger brother is well off. This difference between ordinary comparatives and well-being comparisons should be resisted. At the very least, we should require an excellent reason for thinking that well-being comparisons differ from ordinary comparatives.

The second objection to the argument from well-being is that value-comparatives should not be analysed in terms of degrees of goodness but rather in terms of comparative attitudes. According to one such attitudinal view prudential comparative value should be understood in terms of what it would be fitting, or appropriate, to prefer *for that person's sake*.⁸ On this view, S is extrinsically better for a than $\neg S$ iff it would be fitting to prefer S over $\neg S$ for a 's sake. Such an attitude, it might be claimed, can be held even if one of the state of affairs is the person's non-existence. For example, it seems fitting to prefer an actual person's existence over her non-existence just in case that person has positive well-being.

The success of the attitudinal approach to comparative value hinges on whether the notion of preferring something for someone's sake can be sufficiently clarified. One worry is it that in order to clarify this notion we have to make similar assumptions as the view which analyses the better-for relation in terms of degrees of prudential value. More precisely, we need some explanation of why it is fitting to prefer a 's existence over her non-existence just in case a has a life worth living. A plausible answer to this question is that it is fitting to prefer one state of affairs to another for a person's sake just in case the person has more well-being in the one state than in the other. But, this reply just brings us back to the conclusion that existence cannot be better for a person than non-existence because the person would not have any amount of well-being were she not to exist.

A more effective reply to the argument from well-being is to consider the relation between the value of a possible world w for a person and the amount of well-being that person has in w . That is, we should make a distinction between *the value of a possible world for a person* on the one hand and *the well-being the person would enjoy were that possible world actual* on the other. The idea is that possible worlds, even worlds where a particular person a does not exist, can still have value for a given that a actually exists. By weakening

⁷This has been argued by Roberts (2003) and Holtug (2010, pp. 131–4).

⁸This attitudinal approach is suggested by Arrhenius and Rabinowicz (2010).

the connection between the value of a world for a person and the amount of well-being the person has in that world we can then say that comparative prudential value should be analysed in terms of degrees of prudential goodness without ascribing any amount of well-being to non-existent people.

In order spell this reply out in more detail, recall the view of extrinsic value mentioned earlier. According to that view, a state of affairs S is extrinsically good (bad) for a person a iff the value for a of the closest possible world where S obtains is greater than the value for a of the closest possible world where S does not obtain. For any existing person, the extrinsic value of the closest world where that person exists, i.e the actual world, is the aggregated value for the person of all the states of affairs which obtain in the actual world. Because I have assumed that prudential value should be understood entirely in terms of well-being, this amounts to the claim that the extrinsic value of the actual world for any existing person is positive (negative) iff the person has overall positive (negative) well-being in the actual world.

The extrinsic value for a person of the closest possible world where a person does not exist, on the other hand, is defined in terms the *actual* intrinsic value for the person of the closest possible world where that person does not exist, not the value this possible world *would have* for the person. That is, when determining the value of a possible world other than the actual one there is, according to this view of extrinsic value, a difference between the intrinsic value of that world for the person and the well-being that person enjoys in that world.⁹ Most importantly, as was noted above, there are good reasons to think that the actual value for a person of a world where that person does not exist is positive, negative or neutral. According to hedonism, for example, the value of such world is zero, while on preferentialism the value might be positive or negative.

The way to defend the value of existence from the objection raised in this section is therefore, in short, to make a distinction between the value of a possible world for a person and the well-being the person would enjoy were that possible world actual. This distinction is especially important when we are dealing with possible worlds where a person does not exist. By defining the extrinsic value of a state of affairs for a person in terms of the actual intrinsic value for the person of possible worlds one avoids having to say that the person has any well-being, even zero, in worlds in which the person does not exist.

4 Extrinsic Value and Practical Relevance

As we have seen, neither the argument from Axiological Invariance nor the argument based on a connection between better-for and better-off succeed in showing that existence cannot be extrinsically better for a person than non-existence. We have also seen how existence could be better or worse for a person than non-existence on different theories of well-being. While this certainly supports the conclusion that existence can be extrinsically better than non-existence, I will in this section argue that it does little to solve the practical problems connected with the harm or benefit of coming into existence.

In the previous section I argued that a crucial feature of an axiological theory which allows for existence to be extrinsically better (worse) for a person than non-existence is that

⁹A similar reply to the objection has recently been suggested by Feit (2016). However, Feit's view makes use of an "inner" and an "outer" notion of truth. The reply I sketch here does rely on this distinction, though it is compatible with it (as far as I can see).

extrinsic betterness is defined in terms of the *actual* intrinsic value of possible worlds. The problem with this feature is that in certain cases an agent who is guided by extrinsic value will have an alternative, ϕ , which is extrinsically better for her than not ϕ -ing, but were she to ϕ then it would have been extrinsically better not to have ϕ -ed. In situations such as these it seems impossible for a rational and well-informed agent to deliberate about what to do, in so far as she is guided by the extrinsic value of ϕ -ing. However, if the extrinsic value of existence has any practical relevance then it should be possible for a well-informed and rational agent to be guided by this value in deliberation. Since this is not possible in certain cases, the extrinsic value of existence lacks practical relevance.

In order to see what such a case might look like, suppose that prudential value depends on attitudes. More precisely, suppose that S is intrinsically good (bad) for a iff (and because) a has a pro (con)-attitude of the right kind towards S for its own sake. Consider a person, a , who has a stronger pro-attitude of the right kind towards the states of affairs which would obtain were she to exercise than she has towards the states of affairs which would obtain were she not to exercise. It follows from the analysis of extrinsic value that exercising is extrinsically better for a than not exercising. However, suppose that if a were to start exercising then her attitudes would change. In fact, if she were to start exercising then she would have a stronger pro-attitude towards the states of affairs which would obtain were she not to exercise. This means that in the closest possible world where she exercises, not exercising is extrinsically better for her than exercising.

This puts a in a peculiar situation. In so far as a is guided by the extrinsic value of exercising, it is reasonable of a to start exercising. However, if she were to start exercising, then it would be extrinsically better for her not to exercise. It would therefore be reasonable of her to stop exercising. Whatever a does, she will end up with something which is extrinsically worse for her, and whatever she does it would have been extrinsically better for her not to have done it. Note also that this result is not due to any shortcoming on a 's part and that the same situation arises even if we assume that a is a well-informed, rational and conscientious agent who is guided by promoting extrinsic value. It is at least unclear what kind of shortcoming this would be since the mere fact that a person's preferences can depend on the circumstances does not seem to be evidence of irrationality or unreasonableness.

If extrinsic value is to have any practical importance, whether morally or prudentially, then this is a serious problem. In situations where one will end up with a state of affairs which is extrinsically worse whatever one chooses, it seems impossible to deliberate about what to do, in so far as one is guided by extrinsic value. This suggests that the person in the example should not be guided by the actual extrinsic value of exercising, but rather something else.¹⁰

A possible objection is that it is too hasty to discard the practical relevance of extrinsic value on the basis of this example. It might be claimed that it is the theory of prudential value, not the practical relevance of extrinsic value, which is the source of the problem in the example above. However, note that the problem above only arises for preferentialism in so far as preferentialism is combined with the claim that one ought to be guided by actual intrinsic value. An alternative version of preferentialism according to which one ought to be guided by the extrinsic value the outcomes *would have* does not imply that there are

¹⁰See Bykvist (2006) for a careful discussion of how best to interpret preferentialism in the light of preference change. The problem presented here owes much to Bykvist's paper. As I argue below however, similar problems arise for other theories of well-being, and is therefore not peculiar to preferentialism.

dilemmas of this kind. In the example with shifty preferences above, such a view would recommend that the person ought to start exercising iff the pro-attitude she would have for not exercising, were she to start exercising, are weaker than the pro-attitude she would have for exercising, were she not to start exercising. This is one way to formulate the preferentialist theory such that it avoids these problems, but note that it does so by denying that actual extrinsic value is practically relevant.

Even though certain forms of preferentialism will avoid the problem of shifty preferences, it might be argued that the example above still reveals a flaw in preferentialism since the same problem does not seem to arise for other theories. In the case of hedonism, for example, it might be thought that this theory is not vulnerable to the same kind of problem since it does not make prudential value dependent on anything as shifty as a person's attitudes.

However, it is possible to construe similar cases even for hedonism. By hedonism I mean, roughly, the view that a state of affairs S is intrinsically good (bad) for a iff (and because) S is an instance of a experiencing pleasure (pain). Suppose now that a creature, a , lacks the capacity for experiencing pleasure or pain. We can give a this capacity, but if we were to do so then a would only experience pain.

It is plausible to assume that nothing is good or bad for a in the actual world, assuming hedonism. The intrinsic value for a of the actual world is therefore neither good nor bad, and likewise for the closest possible world where a experiences only pain. Having the capacity for pleasure and pain is therefore neither extrinsically good nor bad for a . It is an open possibility that it is not neutral either: it may be that it is undefined for example. However, it is clear that if we are guided by the extrinsic value for a of having the capacity for pleasure and pain, then there is no relevant difference between giving a this capacity and not giving a this capacity. But note that if we were to give a the capacity for mental states then it *would be* extrinsically worse for a to have this capacity. Therefore, in so far as we are guided by promoting extrinsic value for a , we will not be able to rule out the alternative of giving a the capacity for mental states. However, in so far as we are concerned about a it is clear that we should not give a the capacity for mental states, despite the fact that this world *is* not extrinsically worse for the person.

We can even construe deliberative dilemmas for hedonism, similar to the one about exercising. Suppose for example that there are two creatures, a and b , which lack the capacity for pleasure and pain and that we can give one of them this capacity. If we give it to a then a will only experience pain, and same is true of b . In this case if we give the capacity to a then we ought to have given it to b and vice versa, in so far as we are guided by the actual extrinsic value for a and b of having this capacity. If we give it to a , then having the capacity is extrinsically worse for a . However, the extrinsic value for b of having the capacity is neutral since nothing has any value for b when b lacks the capacity for pleasure and pain. It is easy to see that this leads to the conclusion that if we give the capacity for mental states to a then we ought, in so far as we are guided by promoting extrinsic value, to have given it to b , and vice versa.

I have argued that both a certain form of preferentialism and standard hedonism are vulnerable to deliberative dilemmas if we assume that actual extrinsic value is practically relevant. We should therefore conclude that actual extrinsic value lacks practical relevance. More precisely, the examples above support the conclusion that what matters to practical deliberation is not the value, intrinsic or extrinsic, which a non-obtaining state of affairs *has* for a person, but rather the value it *would have* for the person, were it to obtain. One possible interpretation of this result is that the actual value of non-obtaining states of affairs is only practically relevant if the state of affairs would have the same value, were it to obtain. That

is, something like Axiological Invariance holds, perhaps not for every kind of value, but at least for every kind of value which is relevant to practical deliberation.¹¹

5 The Harmlessness of Existence

It is a common assumption that harm and benefit should be analysed in terms of prudential value (rather than impersonal value, for example). A characteristic feature of harms is that they make a person's life go worse, and a characteristic feature of benefits is that they make a person's life go better. These platitudes raise the question what kind of prudential value we are talking about when we talk about harms and benefits.

One possibility is that the relevant kind of value is intrinsic prudential value. On this view, in order for existence to be a benefit (harm) to a person, it would have to be possible for existence to make a person's life go intrinsically better (worse) for that person. The old arguments, reviewed at the beginning of this paper, show that this cannot be the case. In order for existence to be intrinsically better (worse) for a person, it would have to be the case that the person would have been intrinsically better (worse) off had she not existed. But this is ruled out by the Existence Requirement. Therefore, existence cannot be a benefit or a harm to a person on this view.

A recent, and more plausible, suggestion is that harm and benefit should be understood in terms of *extrinsic prudential value*.¹² On this view, an event harms a person iff the event is extrinsically worse for the person; that is, the intrinsic value of the closest possible world where the event occurs is lower than the intrinsic value for the person of the closest possible world where the event does not occur. This view has several virtues. It can explain, for example, why death typically harms a person and why a vaccine typically benefits a person. Such events are harmful or beneficial, not because they necessarily cause something intrinsically good or bad, but because they prevent something intrinsically good or bad for the person. This analysis is especially relevant to the question whether a person can be benefited or harmed by coming into existence.

Identifying the type of prudential value which is relevant to benefits and harms with extrinsic value therefore allows us to say that existence can benefit or harm a person. However, this conclusion comes at the price of making benefits and harms practically irrelevant. That is, we should not be guided by considerations of benefit and harm when deliberating about what we ought to do (either morally or prudentially), because we shouldn't be guided by actual value, whether intrinsic or extrinsic, when deliberating about what to do.

This conclusion should strike us as surprising. Surely, benefits and harms are prime candidates for something that is morally and prudentially relevant. Proponents of the analysis of benefit and harm in terms of extrinsic value therefore seem to face the following choice. They can endorse the conclusion argued for in this paper and deny that harm and benefit are practically relevant, whether morally or prudentially. This option, however, deprives the conclusion that existence can be extrinsically better or worse for a person than never-existing

¹¹Johansson (2010) draws a similar conclusion, though the arguments he considers differ slightly from the one's offered here. Johansson claims that normative principles which are formulated in terms of the value an outcome would have are more plausible than normative principles formulated in terms of the actual value of these outcomes. My argument aims to point out that one cannot defend the practical relevance of actual extrinsic value merely by accepting counter-intuitive, first-order, normative claims.

¹²See for example Bradley (2009, pp. 65-71) and Feit (2016).

of much its wider philosophical relevance, since it no longer has any bearing on what we ought to do.

Alternatively, they could revise the analysis of harm. For example, they could claim that an event e harms a person a iff e would be extrinsically worse for a , were e to occur, than $\text{not-}e$ would be for a were $\text{not-}e$ to occur. This revised view coincides with the old version in many cases. For example, to deprive a person of goods is, on both analyses, to harm that person. However, they differ in some cases, most notably with respect to existence. In order for existence to harm a person, on this view, it has to be the case that the world where that person exists would be worse for this person than the world where that person does not exist would be for this person. This claim, it seems, can only be true if the world where the person does not exist would have any value for her, were it actual. We should reject this last claim, and with it the claim that existence and non-existence can harm or benefit a person.

Acknowledgements I would like to thank Travis Timmerman, Olle Risberg, Krister Bykvist, Karl Eken Dahl, Victor Moberger, Susanne Burri and the participants at the workshop “Harm: the concept and its relevance” for their valuable comments. I would also like to thank two anonymous referees of this journal.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Adler MD (2008) Future generations: a prioritarian view. *George Wash Law Rev* 77(5-6):1478–1520
- Arrhenius G, Rabinowicz W (2010) Better to be than not to be. In: Joas H, Klein B (eds) *The Benefit of Broad Horizons*. Brill, Leiden
- Bradley B (1998) Extrinsic value. *Philos Stud* 91:109–126
- Bradley B (2009) *Well-Being and Death*. Clarendon Press, Oxford
- Broome J (1999) *Ethics out of economics*. Cambridge University Press, Cambridge
- Bykvist K (2006) Prudence for changing selves. *Utilitas* 18(3):264–283
- Bykvist K (2007) The benefits of coming into existence. *Philos Stud* 135:335–362
- Feinberg J (1986) Wrongful life and the counterfactual element in harming. *Soc Philos Policy* 4(1):145–178
- Feit N (2013) Plural harm. *Philosophy & Phenomenological Research*, pp 1–28
- Feit N (2016) Comparative harm, creation and death. *Utilitas* 28(2):136–163
- Holtug N (2010) *Persons, Interests and justice*. Oxford University Press, Oxford
- Johansson J (2010) Being and betterness. *Utilitas* 22(3):285–302
- Klocksiem J (2012) A defense of the counterfactual comparative account of harm. *Amer Philos Q* 49(4):285–300
- Moore GE (1993) *Principia Ethica* (Revised edition). Cambridge University Press, Cambridge
- Parfit D (1984) *Reasons and persons*. Oxford University Press, Oxford
- Purves D (2015) Accounting for the harm of death. *Pacific Philosophical Quarterly* n(m):nnn–nnn
- Rabinowicz W (2009) Broome and the intuition of neutrality. *Philos Issues* 19(1):389–411
- Rabinowicz W, Österberg J (1996) Value based on preferences. *Econ Philos* 12:1–27
- Roberts MA (2003) Can it ever be better never to have existed at all? Person-based consequentialism and a new repugnant conclusion. *J Appl Philos* 20(2):159–185
- Zimmerman MJ (2001) *The nature of intrinsic value*. Rowman and Littlefield, Lanham