CrossMark

Utilitas (2021), **33**, 480–487
doi:10.1017/S0953820821000200

CAMBRIDGE UNIVERSITY PRESS

REPLY

Reply to Bykvist and Campbell on Possible Beings

Ingmar Persson*

Oxford Uehiro Centre for Practical Ethics, University of Oxford, Oxford, UK
*Corresponding author. Ingmar.persson@philosophy.ox.ac.uk

This article replies to Krister Bykvist and Tim Campbell's criticisms of my arguments in *Inclusive Ethics* for the claim that beginning to exist can be better or worse for someone than never existing. In response to their contention that my leading argument needs revision, I propose a revision which by contrast to theirs is satisfactory. They claim that for non-existence to be evaluatively comparable with the positive or negative value of existence, it is insufficient that non-existence lacks positive and negative value for possible beings. It supposedly requires an unexplained, 'neutral' value in its own right. But I set out conditions under which lack of value suffices for such comparisons. I also show that my position does not assume that possible beings exist in any controversial sense. Lastly, my view makes better sense of our having reasons for causing someone to exist.

I

In 'Persson's Merely Possible Persons'[1] Krister Bykvist and Tim Campbell critically address my arguments in *Inclusive Ethics* for the view that 'never having existed is worse for possible beings than an existence in which things are overall intrinsically good for them'.[2] They formulate my leading argument more formally than I do in the following way (which I have slightly modified):

(1) for a being who has never existed nothing is either intrinsically good or bad,

and

(2) the fact that nothing is either intrinsically good or bad for a being is worse for it than the fact that things are overall intrinsically good for it;

therefore,

(3) never having existed is worse for a being than an existence in which things are overall intrinsically good for it. (p. 480)

[1]Krister Bykvist and Tim Campbell, 'Persson's Merely Possible Persons', *Utilitas* 32 (2020), 479–87. Unprefixed page references are to this article.

[2]Ingmar Persson, *Inclusive Ethics* (Oxford, 2017), p. 61.

© The Author(s), 2021. Published by Cambridge University Press

https://doi.org/10.1017/S0953820821000200 Published online by Cambridge University Press

They object that 'the two facts mentioned in (2) are incompatible' (p. 480). Now it is true that if you *assert* both of these facts, you are guilty of a contradiction. But in order to make an evaluative comparison between them, you need only *suppose* first that it is a fact that nothing is either intrinsically good or bad for a being, and then that it is a fact that things are overall intrinsically good for it, and judge what is better or worse for it. This means that (2) could be understood as saying: *if* it is a fact that nothing is either intrinsically good or bad for a being, this is worse for it than *if* it is a fact that things are overall intrinsically good for it. It seems to me perfectly true that what is neither intrinsically good nor bad is worse than what is overall intrinsically good, just as it is better than what is overall intrinsically bad.

However, Bykvist and Campbell also object that 'it is unclear what it means to say one evaluative fact is worse than another for a being' (p. 480). So, let us skip reference to facts and take (2) to say simply:

(2*) it is worse for a being if nothing is either intrinsically good or bad for it than if things are overall intrinsically good for it.

Another remark they make is that 'readers should interpret (1) and (3) as talking about merely possible beings', by which they mean 'beings that never *have* existed and never *will* exist' (p. 480). In *Inclusive Ethics*, however, I used 'merely possible beings' to mean beings who have not existed *so far* but who might exist in the future. Noting this, (1) and (3) should rather be understood as claiming:

(1*) for a being who has not existed so far nothing is either intrinsically good or bad as long as it remains non-existent, and
(3*) for a being who has not existed so far to remain forever non-existent is worse than (to begin) an existence in which things are overall intrinsically good for it.

Importantly, (1*) differs from Bykvist and Campbell's reading by not ruling out that the possible being begins to exist in the future. This difference is crucial because (3) evaluatively compares the non-existence of a being with a possible existence of it. Their proposed replacement for (3) is instead:

(6) For a merely possible being, non-existence is overall worse than an existence that is overall good for it.

But it makes no sense to speak of an existence that is overall good for a merely possible being in their sense of a being who never has existed and never will. To make sense, (6) must be revised to something like: for a merely possible being, non-existence is overall worse for it than it would be were it not *merely* possible but at some time had an existence that was overall good for it. Anyway, it is the argument from (1*) and (2*) to (3*) that represents my position.

II

To move onto something more substantial, Bykvist and Campbell object to (2), even if it were revised according to their liking. They object that it implies that 'something that is overall good for a being is better for it than something that *lacks* overall positive or negative value for it. But something is better for a being than something else only if

https://doi.org/10.1017/S0953820821000200 Published online by Cambridge University Press

both things have value for the being' (p. 481). The value that something that lacks positive and negative value must have to be evaluatively comparable to what is good or bad is 'neutral value'. According to them, this is 'a *value* that is neither positive nor negative, but which can be compared to positive and negative values' (p. 482). The emphasis on 'value' is meant to signal that neutral value is not merely an absence of positive and negative value, but is 'a value in its own right' (p. 482). This is confirmed when they later remind us that '*being neutral for* is not just a lack of the relations *being good for* and *being bad for*' (p. 484). Thus, in contrast to me they evidently hold that there is a sense of having neutral value in which this implies positively having some sort of value which excludes having positive and negative value.

This may explain why they are drawn to the view 'that standing in a value relation requires being concrete in some way' (p. 484): if neutral value is regarded as a value in its own right like positive and negative value, it may seem plausible that things can have neutral value only for something that has concrete existence, just as things can be good or bad only for something with such an existence.

However, I have no idea what this third value in between positive and negative value could be, and Bykvist and Campbell do nothing to clarify it. Nor do they enlighten us concerning what properties something must have in order to be something for which things can have neutral value. They tell us that in virtue of 'being a concrete being and an animal', an anencephalic infant 'seems a better candidate for a being for which things can be neutral' (p. 484) than a merely possible being. But they do not explain what it is about neutral value that makes a concrete being seem a better candidate. It could hardly be because its concreteness is a necessary condition for having a capacity for consciousness, since this capacity is surely not necessary for a life being comparable with and found worse than a good life, as for example the case of anencephalic infants shows. They leave us in the dark, then, both about what conditions are necessary and sufficient for being a subject for whom things can be of neutral value and about what this value could be in addition to an absence of positive and negative value.

Although this is nothing I have worked out in detail, instead of postulating a value in its own right, I am inclined to put constraints on the relevant class of entities for which there are things lacking intrinsic positive and negative value. Roughly speaking, they must either be an *actually existing* specimen of a kind such that things can conceivably be intrinsically good or bad for such a specimen, or be a *possible* specimen of such a kind. The rationale for such a constraint is that it identifies a class of entities with respect to which occasions could arise for making evaluative comparisons with what is good or bad for them.

To exemplify, although I maintain that nothing is intrinsically good or bad for existing beings who never have consciousness, such as anencephalic infants, they are of a kind for which there could conceivably be things that are intrinsically good or bad (*Inclusive Ethics*, p. 61). For it is conceivable that they were subjected to a neurosurgery that endowed them with consciousness. These infants are living organisms, but I do not rule out that even inorganic entities, such as computers, could conceivably have minds and, thus, be subjects of a kind for which things are either intrinsically good or bad. Consequently, not only possible conscious beings, but also possible inorganic entities could be subjects for which nothing's being good or bad falls into the category of being open to evaluative comparisons if, when they begin to exist, they could conceivably be conscious, perhaps if they underwent some treatment.

With respect to anencephalic infants in their untreated state, it seems entirely natural to hold that it is neither better nor worse for them to exist than to have remained in non-existence and, thus, that things lack intrinsic goodness and badness for them in

https://doi.org/10.1017/S0953820821000200 Published online by Cambridge University Press

the same sense whether or not they begin to exist. For in both cases this is due to their never having been conscious (*Inclusive Ethics*, p. 61). The class of entities for which the absence of both positive and negative intrinsic value satisfies the requirement I have laid down is obviously vast, but it does not include 'everything'. For instance, it does not include contradictory or necessary states of affairs, marshalled against my position by my critics (p. 481).

They endeavour to buttress their rejection of my view of evaluative comparisons by claiming that it is disanalogous with comparisons involving other properties: 'One thing is taller than another only if both have height; one thing is heavier than another only if both have weight; one thing has higher temperature than another only if both have temperature' (p. 481; cf. p. 482). But I think my view of value is in fact analogous to *tactile* or *felt* temperature, which is not the same as what is measured by thermometers. Just as there are positive and negative values, there are the felt qualities of warmth and coldness. However, there is no third felt thermal quality; 'felt thermal neutrality' would simply be a feeling lacking both warmth and coldness.

Therefore, something that feels warm feels warmer than something that feels neither warm nor cold, that is, feels as having none of these qualities. This in turn feels neither warmer nor colder than feeling nothing, that is, being numb. Certainly, we may not want to say that what feels warm feels warmer than entities of a kind that cannot conceivably feel warm or cold, such as sounds. So, we may reach for a restriction corresponding to the one I proposed for value. Granted, this restriction permits odd-sounding statements, for example, that actual red-hot iron feels hotter than merely possible red-hot iron. But I think such statements are true, indeed so obviously true as to go without saying – except in philosophical contexts (*Inclusive Ethics*, p. 64).

Weight and length are different from secondary thermal qualities felt. Being heavy is not a different *quality* from being light, nor is being tall a different *quality* from being short, and high temperature is not a different *quality* from low temperature. These differences are purely *quantitative*, a matter of more or less, whereas the difference between felt warmth and coldness is *qualitative*. It appears plausible to say the same about the difference between positive and negative intrinsic value. Notice that paradigm exemplars of what is intrinsically good and bad, namely pleasure and pain, are felt qualities; and what is felt as neither good nor bad is neither better nor worse than feeling nothing. Moreover, the idea that value is analogous to secondary qualities is a familiar one.

III

Bykvist and Campbell interpret me as claiming that

A. there are merely possible beings

follows from

B. it is possible that some beings will begin to exist in the future.

They find this claim 'puzzling' (p. 482). So, let me reformulate my point, which I take to be more trivial than puzzling. It is true that (a) some beings will begin to exist in the world in the future. There is an everyday sense of 'exist' in which this is equivalent to (b) there exist/are (some) beings who will begin to exist in the world in the future.

https://doi.org/10.1017/S0953820821000200 Published online by Cambridge University Press

This equivalence is borne out for example by the fact that in predicate logic the existential quantifier would be used in formalizing both (a) and (b). The beings in question are merely possible beings in my favoured sense, namely possible beings that have not actually existed *up to now*. (By contrast, in Bykvist and Campbell's sense merely possible beings do not actually exist at *any* time.) Therefore, (c) there exist merely possible beings who will begin to exist in the world in the future. My intended claim was that a statement like (c) follows from (a).

The 'existential expansion' which leads from (a) to (b) merely makes explicit something to which (a) is already committed. To my mind, it is trivial that this expansion is acceptable. What is not trivial is what it means to say that abstract entities like possibilities, properties, propositions, etc. exist. As far as I know, there does not exist (!) an adequate account of this existence. It is a form of existence different from existence in the world, that is, in space and time, as (c) evidently implies.

In *Inclusive Ethics* I dealt with the issue of there being more than one sense of 'existence' quite briefly for two reasons. First and foremost, I take the existential expansion mentioned to be a rather trivial operation. Secondly, I intended to refer to Derek Parfit's longish defence of the claim that there is more than one sense of 'exists',[3] to which I have nothing significant to add. But, unfortunately, making this reference slipped my mind.

Another more controversial issue concerning possible beings than their existence is their relation to actual beings. This is relevant to my view that it could be worse for a merely possible human being, Tommy, to have existed, since he would then have had some very painful disease or disability. Bykvist and Campbell wonder how this judgement can be about Tommy, since he 'would not have been around, things would not have been worse for him, for the relation "worse for" requires a subject for whom things are worse' (p. 483). On my view, things can be intrinsically *bad* (or good) only for subjects who have actually existed (with consciousness) at some time, but for reasons that should now be clear this is false for the comparative 'worse for'. For merely possible beings, nothing is either intrinsically good or bad, and an existence that is overall intrinsically bad for them is worse for them than this.

But if we assume, quite reasonably, that a 'merely possible being is not identical to Tommy' (p. 483), my critics ask how what things are like for this merely possible being could support the claim that things would have been worse for the actual Tommy. The problem here is, however, a general one about how we should understand statements about possible beings which entail that they come into existence. It is a problem that arises not only when we make evaluative comparisons between existence and non-existence. If we utter 'If Tommy begins to exist, he will start to scream', it seems we cannot be talking about the same subject, Tommy, in both the antecedent and consequent. For the consequent makes a prediction about an actually existing being, Tommy, which is falsified if this being does not start screaming, but the antecedent cannot sensibly mean 'If the actually existing being Tommy begins to exist . . .'. Since this is a more general problem which arises in other contexts than the present, it is not incumbent on me to solve it fully.

Nonetheless, I believe that the answer in outline is that, although the possible being that is the subject of the antecedent is distinct from the actual being in the consequent, it is not *wholly* distinct from the actual being, for the possible being is *actualized* in the actual being. My critics complain that 'it sounds like a category mistake to say that

[3]Derek Parfit, *On What Matters* (Oxford, 2011), vol. 2, Appendix J.

https://doi.org/10.1017/S0953820821000200 Published online by Cambridge University Press

things can be neutral for property-like entities' (p. 483n.), but to this my reply is that this depends on what the property-like or possible entities are for which things are said to lack value. It is only if they are abstract entities of a kind for which things could conceivably be intrinsically good or bad should they be actualized that things' lacking intrinsic goodness and badness for them fulfil the requirement of giving rise to possible evaluative comparisons, as proposed in section II.

To anticipate a matter further discussed in section IV, let me propose the following partial analogy to the relation between possible and actual beings. Suppose that when a pre-conscious foetus acquires consciousness, a mental entity distinct from the human organism starts to exist and that, *strictly speaking*, it is only for this entity that things can be intrinsically good or bad, but that in a *derivative* sense this could also be said of the human organism in which the mental entity is embodied. Then I think we can truly say of the pre-conscious organism that it can be better or worse for it to acquire consciousness, for even though it is distinct from what is strictly speaking the subject of value, it is identical to the organism for which things derivatively have value in virtue of embodying this subject. The parallel between pre-conscious beings and possible beings for which nothing is good or bad, but for which things can be made better or worse if they are concretely actualized, is, then, that neither of them is wholly distinct from but is rather an element of the beings for whom things can be intrinsically good or bad.

At a rather late stage of their article, Bykvist and Campbell suddenly declare that they 'need an argument for the controversial claim that things can be better or worse for non-concrete beings' (p. 484). This request is surprising not only because right from the outset they have been busy criticizing my arguments for this contention, but also because they do not appear to recognize any need on their own part to produce positive arguments for their contrary contention. They claim things cannot have 'neutral' value for merely possible beings, but they fail to substantiate this claim with an account of what this alleged neutral value is and what is required for being a subject for which there can be this value.

They speculate that my response to their criticisms might be to maintain that 'states of affairs can have value for a being if that being has a *capacity* for being concrete' (p. 484). But, as should be clear from my treatment of potentiality (for consciousness) in chapter 2.1 of *Inclusive Ethics* (esp. pp. 53–4), I could have no sympathy for the idea that the possibility of something is sufficient for a capacity or potential for it. The idea that something as abstract as a merely possible being could have something like a capacity or potential (which may be a species of capacity) strikes me as patently absurd.

IV

Bykvist and Campbell also question whether the existence of an anencephalic infant could be made better for *it* by an acquisition of consciousness. This brings us back to the distinction between primary and derivative attributions of properties alluded to in section III. 'Suppose our identity is essentially tied to our capacity for consciousness', they hypothesize, '[t]hen a *new being*, one of us, comes into existence' if 'the anencephalic somehow acquires a fully functional brain' (pp. 484–5). If this occurs, they conclude, it is the new being and not the anencephalic, a certain human organism or animal, who is the 'welfare subject'. However, even if we adopt a 'Lockean' view to the effect that we are identical to our minds, as my critics apparently are inclined to do, we cannot plausibly deny that human (and other) animals also feel pleasure, for

https://doi.org/10.1017/S0953820821000200 Published online by Cambridge University Press

instance when they chew their food or lie down when tired. Such claims imply that the (whole) human organism can feel pleasure in some sense.

But if we make this concession, my critics contend, we fall victims to the objection that 'there are *too many* welfare subjects' (p. 485): a certain mental entity as well as a living human organism. No, we do not, as I have argued elsewhere.[4] For, as adumbrated in section III, the two subjects are not *independent* welfare subjects: the mental entity is the *primary* welfare subject, and the human organism is a *secondary* or *derivative* welfare subject because it is such a subject *in virtue of embodying the primary welfare subject*. This distinction between primary and secondary subjects of attribution applies not just when mental properties are involved. If *my hand* touches a wall, *I* too can be said to touch the wall, but this does not imply that two things touch the wall independently of each other, for I touch it in virtue of the fact that my hand touches it.

The 'body-switch' that my critics imagine (p. 485) does not invalidate this distinction. They imagine someone's brain 'is transplanted into a different human body that somehow developed without a brain'. If the 'new body is superior in ways conducive to her well-being', the primary welfare subject – the mind sustained by the brain – is likely to receive a welfare boost, but the derivative welfare subject – the whole organism which receives the brain – will most likely gain an even greater boost, since before the transplant it did not derive welfare from anywhere, being brainless. My critics claim that 'it seems implausible that there are *two beneficiaries* here' (p. 485). No, it is not implausible, since the beneficiaries are not independent of each other; for instance, if we satisfy one, we satisfy the other.

There is a connected aspect of the mind–organism problem which is more pertinent to the chief topic of whether it can be better for someone to begin to exist. Suppose we start existing when our minds start existing. Then, if Bykvist and Campbell are right that beginning to exist cannot be better for anyone, our minds beginning to exist at a foetal stage cannot be better for *us*, even though what our minds experience is overall intrinsically good. But this *can* be better for the *derivative* welfare subject, the human organism, since *it* does not begin to exist when its mind does (*Inclusive Ethics*, p. 52). It is, however, somewhat odd that the derived welfare can be better for someone when it derives from the welfare of a primary subject that is not better for anyone. If we believe we have a reason to create welfare only if someone benefits from it, we are in a quandary here. It seems we do not have any reason to create welfare for the primary subject, and yet the welfare the human organism derives from this creation of welfare seems to provide a reason for creating it. This quandary is of course eliminated on my view that beginning to exist can be better for someone.

V

In *Inclusive Ethics* I suggested that to 'have a reason do an action is to have a reason to do it *rather than* not to do it' (p. 58) and that, therefore, the outcome of doing the action must be *better* (in relevant respects) than the outcome of not doing it. Imagine that if we perform an action, we cause to exist and benefit some beings in addition to benefiting some beings who exist independently of our action. If we do not perform it, we benefit only the independent beings, but we benefit them slightly more than

[4]Ingmar Persson, Our Identity and the Separability of Persons and Organisms, *Dialogue*, 38 (1999), 519–33. But, as I indicated in this article and have contended in other publications, including *Inclusive Ethics*, ch. 3.1, I do not think a Lockean or any other theory of our identity is correct.

https://doi.org/10.1017/S0953820821000200 Published online by Cambridge University Press

doing the action would, though not so much more that the sum of benefits will be as big as it would have been had we performed the action. Then we would have a reason to perform the action rather than not to perform it if the benefits to those who could be brought into existence count whether or not we bring them into existence. But this is not so if the missed benefits of these possible beings do not count if our omission of the action ensures their non-existence forever, and such beings cannot have been harmed. We would then have done what is best for everyone who counts even if we do not perform the procreative action. That is, we have no reason to perform this action *rather than* not to perform it. So far as I can see, Bykvist and Campbell do not address this objection.

I also contended that the fact that 'an action would provide someone with a non-comparative benefit is . . . not sufficient to show that . . . there is a reason to perform the action', for this is compatible with for example 'an even greater benefit' resulting from not performing the action (*Inclusive Ethics*, p. 58). My critics protest that we do have a reason to give the individual a non-comparative benefit, such as a pleasure, 'rather than do what results in her not having this benefit' (p. 486). Not so, because the result of her missing this benefit could be her getting an even greater benefit or preventing her from an even greater burden. In other words, to have a reason to give her the benefit we need to know that getting the benefit is *better* for her than not getting it, for instance by not existing.

They might retort that we have a reason to do what results in the individual having the benefit, even though this is a reason that could be outweighed by a *stronger* reason to do what results in her not having it. While we often talk about reasons in this fashion, we do not always do so. For instance, if both doing an action and not doing it result in someone getting a benefit, we would say that this benefit provides no reason to perform the action, not that it provides an equally strong reason to perform it as not to perform it. The more important question is, however, whether we have *stronger* reason to perform an action than not to perform it, and to decide this, we need to determine whether the outcome of performing the action is *better* than the outcome of not performing it. But the topic of reasons for action cannot be given its due here, since this would involve intricate issues inessential to the main issue of this reply, which is whether it can be better or worse for somebody to start existing.[5]

Competing Interests. The author declares there are none.

[5]Thanks to Bykvist and Campbell for having prodded me to clarify my position on the matters discussed. Thanks also to Suzanne Uniacke and two anonymous reviewers.

Cite this article: Persson I (2021). Reply to Bykvist and Campbell on Possible Beings. *Utilitas* **33**, 480–487. https://doi.org/10.1017/S0953820821000200

https://doi.org/10.1017/S0953820821000200 Published online by Cambridge University Press