

REPLY

## Persson's Merely Possible Persons

Krister Bykvist<sup>1\*</sup> and Tim Campbell<sup>2</sup> 

<sup>1</sup>Stockholm University and <sup>2</sup>Institute for Futures Studies, Stockholm

\*Corresponding author. Email: [krister.bykvist@philosophy.su.se](mailto:krister.bykvist@philosophy.su.se)

### Abstract

All else being equal, creating a miserable person makes the world worse, and creating an ecstatic person makes it better. Such claims are easily justified if it can be better, or worse, for a person to exist than not to exist. But that seems to require that things can be better, or worse, for a person even in a world in which she does not exist. Ingmar Persson defends this seemingly paradoxical claim in his latest book, *Inclusive Ethics*. He argues that persons that never exist are merely possible beings for whom non-existence is worse than existence with a good life. We argue that Persson's argument, as stated in his book, has false premises and is invalid. We reconstruct the argument to make it valid, but the premises remain highly problematic. Finally, we argue, one can make sense of our procreative obligations without letting merely possible beings into the moral club.

I

Many of our choices effect who will exist in the future. These include the choice of having a child, of giving priority to young people over old when saving lives, and of offering generous state-funded parental leave. In many of these cases it seems that one outcome is worse than another in virtue of the wellbeing of people who do not exist in both. For example, it seems, creating a very unhappy person makes the world worse, other things being equal. Some would also say creating a very happy person makes the world better, other things being equal. It would be easy to justify these verdicts if it can be better, or worse, for a person to exist than not to exist. But that seems to require that things can be better, or worse, for a person even in a world in which she does not exist, which sounds paradoxical.

This paradoxical-sounding claim is defended in Ingmar Persson's latest book *Inclusive Ethics*.<sup>1</sup> He argues that in a world in which a person does not exist, she is a merely possible being – a being that never exists – and that for such beings it is worse not to exist than to exist with a good life. Furthermore, he argues for this claim from what he calls 'incontestable' premises. We argue that the premises are far from incontestable. Persson's argument, as stated, has false premises and is invalid. One can reconstruct the argument to make it valid, but the premises are still highly contestable. Finally, we argue, one can make sense of our obligations to future generations without letting merely possible beings into the moral club.

<sup>1</sup>Ingmar Persson, *Inclusive Ethics: Extending Beneficence and Egalitarian Justice* (Oxford, 2017).

## II

Persson's argument is presented in the following passage:

[I]t follows that never having existed is worse for possible beings than an existence in which things are overall intrinsically good for them if it is true that:

- (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.<sup>2</sup>

Claims (1) and (2) are the premises, which are claimed to be 'incontestable', and the conclusion is:

- (3) never having existed is worse for possible beings than an existence in which things are overall intrinsically good for them.

Note that (1) talks about a being who *has never* existed. This does not rule out that it *will* exist in the future. But Persson wants to say that non-existence can be worse for *merely possible* beings, beings that never *have* existed and never *will* exist. So, readers should interpret (1) and (3) as talking about merely possible beings.

Persson finds (2) 'incontestable', but as stated it is false, since the two facts mentioned in (2) are incompatible. Necessarily, if it is a fact that nothing is intrinsically good or bad for a being, then it is *not* a fact that things are overall intrinsically good for this being. Hence, it cannot be true that *the fact* that nothing is intrinsically good or bad for a being is worse for it than *the fact* that things are overall intrinsically good for it.<sup>3</sup>

Another problem is that the facts compared in (2) are *evaluative* facts, facts about whether things are good or bad for a being. But it is unclear what it means to say one evaluative fact is worse than another for a being.

Finally, the argument is invalid: (3) does not follow from (1) and (2). According to (2), a certain evaluative fact is worse for a being than a certain other evaluative fact. But 'never having existed' in (3) refers to the *non-evaluative* state of affairs (or fact) of a being never having existed.

## III

The following reconstruction of Persson's argument, where evaluative fact-talk is replaced by non-evaluative state-of-affairs-talk throughout, avoids the problems of the originally-stated argument:

<sup>2</sup>Persson, p. 61.

<sup>3</sup>This is an instance of the 'relata problem'. See Gustaf Arrhenius and Wlodek Rabinowicz, Better to Be than not to Be?, *The Benefit of Broad Horizons: Intellectual and Institutional Preconditions for a Global Social Science*, ed. Hans Joas (Leiden, 2010), pp. 404–08; Gustaf Arrhenius and Wlodek Rabinowicz, The Value of Existence, *The Oxford Handbook of Value Theory*, ed. Iwao Hirose and Jonas Olson (New York, 2015), pp. 424–44, at 432; Krister Bykvist, Being and Well-Being, *Weighing and Reasoning*, ed. Iwao Hirose and Andrew Reisner (Oxford, 2015), pp. 90–91; and Nils Holtug, *Persons, Interests, and Justice* (New York, 2010), pp. 140–41.

- (4) For a merely possible being, non-existence is neither overall good nor overall bad.
- (5) Any state of affairs which is neither overall good nor overall bad for a being is overall worse for it than a state of affairs that is overall good for it.

Therefore,

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

This reconstruction is valid, assuming ‘existence’ and ‘non-existence’ pick out states of affairs. Moreover, it avoids comparisons of incompatible evaluative facts. Finally, it makes clear that the relevant comparisons concern the overall values of states of affairs (existence or non-existence), i.e. the total intrinsic value for the being of all the things that the states of affairs would realize, if they obtained.<sup>4</sup> We should add that the reason why (4) is true is that everything lacks positive and negative intrinsic value for the being.

But the argument is still problematic. According to (5), something that is overall good for a being is better for it than something that *lacks* overall positive and negative value for it. But something is better for a being than something else only if both things have value for the being. The same holds for any comparative notion. 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 a higher temperature than another only if both have temperature. The question, then, is what value non-existence has for a merely possible being.

Persson states that non-existence has neutral value for a merely possible being. Indeed, he claims that everything is neutral for a merely possible being. However, he characterizes ‘neutral value for a being’ as ‘being neither intrinsically good nor intrinsically bad for a being’.<sup>5</sup> This is an implausible characterization of ‘neutral value for a being’ because, plausibly, some states of affairs are *undefined* in value for a being. Candidates include contradictory states of affairs (e.g. that  $2 + 2 = 5$ ), necessary states of affairs (e.g. that  $2 + 2 = 4$ ), evaluative states of affairs (e.g. that happiness is good for people), and states of affairs that concern other people’s wellbeing (e.g. that a stranger is unhappy).<sup>6</sup>

<sup>4</sup>In conversation, Persson has said that he would accept this reconstruction if (2) were formulated subjunctively as ‘Any state of affairs which *would* be neither overall good nor overall bad for a being *would* be overall worse for it than a state of affairs that *would* be overall good for it’ just to make clear that states of affairs have value for a being only when they obtain. We shall assume this reading implicitly. He has also said he prefers that the premises be stated in terms of intrinsic goodness and badness for a being rather than in terms of the overall goodness and badness that states of affairs have for a being. This is because he thinks non-existence could be extrinsically good or bad for a being by excluding an existence that would be good or bad for that being. On our view, the overall goodness or badness a state of affairs has for a being does not depend on this kind of preventive value.

<sup>5</sup>Persson, pp. 11, 57, 61. Nils Holtug defends a similar definition of ‘zero value for a being’ in *On the Value of Coming into Existence*, *The Journal of Ethics* 5 (2001), 361–84, but he now rejects this definition because of the objection raised here. On the intelligibility of assigning neutral or zero value to non-existence, see Melinda Roberts, *Can It Ever Be Better Never to Have Existed at All? Person-Based Consequentialism and a New Repugnant Conclusion*, *Journal of Applied Philosophy* 20 (2003), 153–85; Jens Johansson, *Being and Betterness*, *Utilitas* 22 (2010), 285–302; and Krister Bykvist, *The Benefits of Coming into Existence*, *Philosophical Studies* 135 (2007), 335–62.

<sup>6</sup>On the distinction between being neutral for and having undefined value for, see Ben Bradley, *Well-Being and Death* (New York, 2009), pp. 98–104.

Furthermore, being neutral for one is different from being neither good nor bad for one. If something is neutral for one, it has a *value* that is neither positive nor negative, but which can be compared to positive and negative values. This is analogous to:

- (a) having zero temperature, which is having a temperature that is neither positive nor negative, and thus different from lacking positive and negative temperature just because one lacks *any* temperature, or
- (b) having a weight that is neither heavy nor light, which is different from being neither heavy nor light just because one lacks *any* weight, or
- (c) having a height that is neither tall nor short, which is different from being neither tall nor short just because one lacks any height.<sup>7</sup>

Persson could deny that things must have value to stand in value relations or that neutral value is a value in its own right. But such claims are highly contestable and Persson does not give arguments for why value comparisons differ so radically from other comparisons.

Premise (4) is also highly contestable. Indeed, that there are merely possible beings is one of the most contestable claims in modal metaphysics.<sup>8</sup> Persson claims not only that there are non-existent beings, but also that things are neutral for them and, given premise (5), that things can be better or worse for them.

Consider the existence claim first. It seems incoherent, since it seems to assert the *existence* of *non-existent* beings. But Persson believes that in a certain sense of ‘exist’, the claim is true:

In one sense of ‘exist’, it is true that there are merely possible beings because this follows from the clearly true claim that it is possible that some beings will begin to exist in the future. This is the sense in which I believe there to be merely possible and, thus, non-existent beings. (60–61)

Persson’s claim that

A. there are merely possible beings

follows from

B. it is possible that some beings will begin to exist in the future<sup>9</sup>

is puzzling. B states that in some possible world some beings will begin to exist in the future. Why would it follow that there are and thus exist, *in any sense*, merely possible beings *here in the actual world*? Compare: it is possible that the pope has two children in the future, and thus possible that two children of the pope will begin to exist in the

<sup>7</sup>These examples are relevant to how we should understand the nature of quantitative properties. See Yuri Balashov, *Zero-Value Physical Quantities*, *Synthese* 119 (1999), 253–86.

<sup>8</sup>For a critique of the claim, see Robert Stalnaker, *Mere Possibilities* (Princeton, 2012), chs 1 and 2.

<sup>9</sup>We interpret B as ‘it is possible that it will be the case that some beings begin to exist’, not ‘it is possible that *there are now* some beings that will begin to exist’, since the latter commits one to the possibility of merely *future* beings, beings that do not exist yet but will exist in the future, a commitment which is almost as contestable as the commitment to merely possible beings.

future. But why would it follow that there exist, *in any sense*, merely possible beings *here in the actual world*? Regardless of what sense of 'exist' is assumed, we need to know how to derive A from B. We could derive it if we accepted the Barcan formula as a bridge principle:

If it is possible that there is a being that is F, then there is a being such that it is possible that it is F.

For then we could make use of the following instance of the Barcan formula:

If it is possible that there are beings that will begin to exist, then there are beings such that it is possible that they begin to exist.

and from B derive that there are beings such that it is possible that they begin to exist, which entails A, assuming that they in fact will not begin to exist. But the Barcan formula is highly contestable, and Persson claims (in correspondence) that this is not what he intended.<sup>10</sup> So, we do not see how he can incontestably derive A from B.

Even if we could establish that A follows from B, we need to know what the *relation* is between merely possible beings in one world and beings that begin to exist in some alternative possible world. Consider Tommy, a child born with a painful condition who lived an utterly miserable life. Persson seems committed to saying that things would have been better for Tommy if he had not existed, and that this holds in virtue of the fact that if he had not existed, things would have been neutral for a merely possible being. Either this merely possible being and Tommy are identical or they are not. Both options are problematic.

If they are identical, then a merely possible being which is not concrete (i.e. lacks space-time location and causal powers) could have existed and been concrete. But being non-concrete seems essential to its bearers: if something is non-concrete, it is essentially non-concrete. Similarly, if something is concrete, it is essentially concrete.

If they are not identical, then it is unclear how Persson can establish that things would be worse for *Tommy*, if he did not exist. If we had not created Tommy, he would not have been around in any sense, since being concrete is essential to Tommy. Since Tommy 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. Even if there is a *distinct* merely possible being in the Tommy-less world and things are worse for this being, it does not follow that things are also worse for *Tommy* in that world, for we have assumed that the merely possible being is not identical to Tommy.<sup>11</sup>

Let us now turn to the crucial value claim: things can be neutral, better or worse for merely possible beings. This claim is far from obvious. After all, a merely possible being is not a concrete being, an animal, a conscious being, a human, and so on. At most, it is a *merely possible* concrete being, a *merely possible* animal, and so on. Why think that a

<sup>10</sup>On the Barcan formula, see Timothy Williamson, *Modal Logic as Metaphysics* (New York, 2013), ch. 2.

<sup>11</sup>In conversation, Persson seems to accept this horn of the dilemma. He compares merely possible beings to abstract properties and concrete beings to property bearers. Merely possible beings are 'actualized' by concrete beings, but not identical to them. However, this doesn't explain how things are worse for Tommy in a Tommy-less world. Also, it sounds like a category mistake to say that things can be neutral for property-like entities.

being that merely possibly exemplifies these features can stand in value relations? Why not think that standing in a value relation requires being concrete in some way, for example, having a mind, body, space-time location or causal powers? We need an argument for the controversial claim that things can be better or worse for non-concrete beings.

Persson offers an argument from analogy. The analogy involves a comparison between a non-existent being and an anencephalic infant that lacks the capacity for consciousness and so never becomes conscious:

It seems indisputable that, given that existing without consciousness, like anencephalic infants do, is neither intrinsically good nor bad for them, this is worse for them than having consciousness and leading a life in which things are predominantly good for them. But . . . non-conscious beings are like non-existent beings in that nothing is either intrinsically good or bad . . . for them. Therefore, non-existence is worse for a being than a predominantly intrinsically good existence, just as its existence is worse for a non-conscious being than a good existence.<sup>12</sup>

According to Persson, the crucial similarity between the anencephalic and the merely possible being is that both lack consciousness. This is supposed to show that nothing is intrinsically good or bad for them, and hence that lacking conscious existence is neutral for them. The analogy is contestable, however, for there is a clear difference between the two beings. Although both lack consciousness, the anencephalic is a concrete being and an animal. The merely possible being is only *merely possibly* concrete and *merely possibly* an animal. Thus, the anencephalic seems a better candidate for a being for which things can be neutral. Remember that *being neutral for* is not just a lack of the relations *being good for* and *being bad for*; it is an evaluative relation in its own right. Moreover, the analogy breaks down for flesh-and-blood individuals, such as the previously discussed Tommy, for he would not exist in any sense if he were not created (assuming such a concrete individual could not have been non-concrete).

Persson could reply that states of affairs can have value for a being if that being has a *capacity* for being concrete, an animal, and so on, and both the anencephalic being and the merely possible being have these capacities. This reply works at most for merely possible beings, not for the flesh-and-blood Tommy, since he does not exist at all in the Tommy-less world (assuming, again, that he could not be non-concrete), and thus cannot exemplify any capacities in that world. Moreover, the reply succeeds only if possibly having a certain property *F* is sufficient for having a capacity to have *F*. But this seems false. For example, the fact that you possibly jump to the moon, i.e. that there is a possible world in which you do this, does not entail that you have the capacity to jump to the moon.<sup>13</sup> Having a capacity to do or have something (in the ordinary sense of ‘capacity’) requires more than having the purely modal feature of being such that one possibly does or has it. Arguably, the capacity must be somehow grounded in features that are not purely modal.

Persson’s analogy also assumes that lacking consciousness is worse for the anencephalic than having a conscious good life, that the anencephalic would *benefit* from having a conscious good life. But whether this is true seems to depend on which theory of our identity is correct. Suppose our identity is essentially tied to our capacity for

<sup>12</sup>Persson, p. 62.

<sup>13</sup>Johansson (2010) makes this important observation.

consciousness.<sup>14</sup> Then none of us could exist as the anencephalic in Persson's example, since it lacks this capacity. Suppose the anencephalic somehow acquires a fully functional brain that is necessary and sufficient to realize the capacity for consciousness. Then a *new being*, one of us, comes into existence. Call this being Beth. If Beth has a good life, this does not imply that *the anencephalic* benefits. For it seems Beth is the only welfare subject in this case, and she is not identical to the anencephalic.

Persson might claim that the anencephalic benefits in virtue of standing in some intimate relation to Beth. For example, suppose the anencephalic is essentially a *human body*. Then when Beth comes into existence, this human body might constitute Beth, or it might acquire Beth as a spatial or temporal part. If Beth has a good life, then the human body might also have a good life in virtue of standing in some such relation to Beth.

But this reply is problematic. Beth is a welfare subject. If Beth's body is also a welfare subject, there are *too many* welfare subjects. To emphasize the problem, consider the case of body-switching. Suppose that Beth's brain, which realizes her capacity for consciousness, is transplanted into a different human body that somehow developed without a brain. Since Beth's capacity for consciousness is essential to her, Beth switches bodies. Now suppose the body-switch benefits Beth because her new body is superior in ways conducive to her well-being. Moreover, suppose Beth's life after the body-switch is overall good. If human bodies are welfare subjects, then Beth's new body, call it Body, should also benefit. There should be two beneficiaries – Beth and Body. In that case, Body's benefit would be distinct from Beth's. Their benefits might even be of different sizes. (For example, Beth might benefit only marginally from the body-switch while Body benefits greatly.) But it seems implausible that there are *two beneficiaries* here. Surely, Beth is *the only beneficiary* of the body-switch.

If our identity is essentially tied to our capacity for consciousness, we should deny that Body benefits from acquiring Beth's brain. But then it seems we should also deny that the anencephalic in Persson's example (a human body) would benefit from acquiring a brain.

#### IV

We conclude that Persson's argument that non-existence is worse for a merely possible being than an existence in which things are overall intrinsically good for that being fails.

Is this bad news? Not if our obligations of beneficence to future generations in non-identity cases can be explained by appealing to their existence being *good* or *bad* for them. We can have a reason to create a person if her life would be good for her. We needn't assume that her existence would be *better* for her than her non-existence.

Persson is not convinced that appealing to what would be good or bad for a person (as opposed to what would be better or worse for her) can justify creating or refraining from creating a person. He provides both a specific argument and a general argument. His specific argument is this. Suppose you can either create A with a good life or benefit a different person, B, who would exist independently of your actions.

Then if you bring the being (A) into existence, you would have done what most benefits the beings who morally count. But, on the other hand, if you do not

<sup>14</sup>See e.g. Tim Campbell and Jeff McMahan, *Animalism and the Varieties of Conjoined Twinning*, in *Animalism: New Essays on Persons, Animals, and Identity*, ed. Stephan Blatti and Paul Snowdon (New York, 2016), pp. 229–52.

bring this being into existence, you would also have done what most benefits the beings (B) who morally count, since now the being (A) that you could have brought into existence does not count morally because, being forever non-existent, it cannot be said to have been harmed by being denied existence. Therefore, you do not have a reason to bring a being into existence rather than not to bring it into existence if it can be benefited only in a non-comparative sense because if you do not act on the reason, it dissolves, and there is no reason to which you have acted contrary.<sup>15</sup>

This is unconvincing. Persson assumes that defenders of a reason to create the non-comparative benefit of a good life must be *moral actualists* – i.e. that they must say that you have a reason to confer a non-comparative benefit only if the beneficiary actually exists (or will exist).<sup>16</sup> This view is vulnerable to Persson's objection. But consider the following non-actualist view: if an action would bring about an outcome in which someone exists with a good life, then this fact provides a reason to perform the action. This view avoids Persson's objection, since on this view the existence of the reason does not depend on who does or will exist. It is enough that the relevant counterfactual is true.

Persson's general argument for there being no reason to benefit a being in a non-comparative sense unless one thereby also benefits her in a comparative sense is as follows:

To ascertain that an action would provide someone with a non-comparative benefit is . . . not sufficient to show that – as far as this individual is concerned – there is reason to perform the action. For it may be that the outcome of performing the action is not better, all things considered for the individual, than the outcome of not performing it because bestowing this benefit removes or prevents the individual's having or getting an *even greater* benefit. Thus, in order to determine that – as far as this being is concerned – you have reason to perform the action, you need to ascertain . . . that the action benefits the being in the comparative sense.<sup>17</sup>

We agree that there is no contrastive reason to give a person a non-comparative benefit rather than do what results in her having an *even greater benefit*. However, this does not show that there is no contrastive reason to give her a certain non-comparative benefit – a pleasure, say – rather than do what results in her not having this benefit. Thus, we think, Persson has not established that the only reason to create a being with a good life is that existence with a good life is *better for that being* than non-existence. We can have a contrastive reason to create a person who will have a good life rather than do what results in her not existing at all, and thus not having any non-comparative benefit.

## V

We have scrutinized Persson's argument for the claim that non-existence is worse for a merely possible being than an existence in which things are overall good for that being. Far from being 'incontestable' the argument's premises, as initially stated, are false, and

<sup>15</sup>Persson, p. 59.

<sup>16</sup>Throughout this section, when we refer to reasons, we mean non-instrumental reasons of beneficence, i.e. those grounded in considerations of what would benefit individuals. According to Persson, such reasons make up one important class of moral reasons (Persson, ch. 1).

<sup>17</sup>Persson, p. 58 (italics added).

the argument is invalid. We presented what we take to be the most promising reconstruction of Persson's argument. Although the reconstructed argument is valid, the premises are *highly contestable*. They assume that there are merely possible beings and that things can be better or worse for them. Persson's argument for the existence claim is unconvincing and his argument for the value claim is based on a questionable analogy involving anencephalic infants. Finally, we showed that Persson is not justified in his dismissal of the idea that we can explain our obligations to future generations by appealing to facts about their existence being good or bad for them.

In sum, Persson has not made a convincing case for giving merely possible persons membership in the moral club. They will have to look for a different guarantor.<sup>18</sup>

---

<sup>18</sup>Many thanks to Ingmar Persson for helpful comments on earlier drafts of this article. Funding from the Riksbankens Jubileumsfond and from the Swedish Research Council is gratefully acknowledged.