Transformative Choice and the Non-Identity Problem

Nilanjan Das and L.A. Paul

**(N.B. This is a penultimate draft of a published book chapter.
Please cite the published version)**

Abstract

Some acts change who we are. Call them *personally transformative* acts. When an agent performs a personally transformative act, she brings into existence a future self that is radically different from who she previously was. In some of these cases, the agent may be antecedently certain that the existence of this future self, though worth having, will be unavoidably flawed, even if the future self values its existence. Now, if the agent doesn't perform the transformative act, she won't change so radically, so her unchanged future self may indeed be better off than her transformed future self. In this essay, we argue that situations of this kind raise a problem that is structurally similar to the *non-identity problem.*

Some of our acts change who we are. Call them *personally transformative* acts. When an agent performs a personally transformative act, she brings into existence a future self that is radically different from who she previously was. In some of these cases, the agent may be antecedently certain that the existence of this future self, though worth having, will be unavoidably flawed, even if the future self values its existence. If the agent doesn't perform the transformative act, she won't change so radically, so her unchanged self may indeed be better off in the future than her transformed future self. In this essay, we argue that situations of this kind raise a problem that is structurally similar to the *non-identity problem*.[1]

Here is the plan for this essay. We begin by explaining what we mean by *personally transformative acts* (§1). We next argue that transformative acts involve the creation of selves that wouldn't otherwise exist (§2). Then, we show how two plausible principles about prudence raise a problem analogous to the non-identity problem with respect to transformative acts (§3). Finally, we consider some responses to this problem (§4).

§1. **Transformative Acts**

[1] The earliest discussions of this problem occur in Narveson (1967), Kavka (1981), Woodward (1986) and Parfit (1987). For more recent work on the topic, see McMahan (1981, 2009, 2013), Bykvist (2007, 2015), Adler (2012), Arrhenius & Rabinowicz (2015), Rabinowicz (2009), and Fleurbaey & Voorhoeve (2015).

In recent work on transformative experience, L. A. Paul (2014) distinguishes two kinds of transformative experience.

Some experiences are *epistemically* transformative: before undergoing those experiences, the agent doesn't know what it would be like to undergo those experiences. Imagine a person who acquires a new sense-modality, e.g., someone who has been blind since birth, but gains ordinary vision. Such a person could now encounter qualities, e.g., luminance and colour, which she didn't see before,[2] and didn't know what it would be like to those qualities before she underwent that experience.[3]

Some experiences are also *personally* transformative: they change the core beliefs, desires, character traits, and other mental states that determine *who the agent is*. Examples include: "experiencing a horrific physical attack, gaining a new sensory ability, having a traumatic accident, undergoing major surgery, winning an Olympic gold medal, participating in a revolution, having a religious conversion, having a child, experiencing the death of a parent, making a major scientific discovery, or experiencing the death of a child" (Paul 2015, p. 16).

There are two ways of framing such personal transformations. One could take a third-person approach to them, like Edna Ullmann-Margalit (2006). She considers situations where an agent must decide amongst certain options, at least one of which, by her own lights, will change her personality, and calls decisions made in such scenarios *opting*. As Ullmann-Margalit describes it, if the agent goes for the personality-changing option in such scenarios, a *New Person* comes into existence. She says:

> New Person is now, by hypothesis, a transformed person. Opting transforms the sets of one's core beliefs and desires. A significant personality shift takes place in our opter, a shift that alters his cognitive as well as evaluative systems. New Person's new sets of beliefs and desires may well be internally consistent but the point about the transformation is that inconsistency now exists between New Person's system of beliefs and desires, taken as a whole, and Old Person's system taken as a whole. (Ullmann-Margalit 2006, p. 167)

She goes on to illustrate the idea with the following example:

[2] Ostrovsky et al (2009).

[3] Cases of this kind can be treated as analogous to the case of Mary in Jackson (1982). This in turn raises an interesting question of what the relevant kind of epistemic transformation consists in. Following the different views on the Mary example, one could flesh it out in terms of the acquisition of certain phenomenal concepts (Loar 1990), or in terms of the acquisition of certain cognitive skills (Lewis 1996), or in terms of the acquisition of self-locating information (Egan 2006).

> I was told of a person who hesitated to have children because he did not want to become the 'boring type' that all his friends became after they had children. Finally, he did decide to have a child and, with time, he did adopt the boring characteristics of his parent friends—but he was happy! I suppose second order preferences are crucial to the way we are to make sense of this story. As Old Person, he did not approve of the personality he knew he would become if he has children: his preferences were not to have New Person's preferences. As New Person, however, not only did he acquire the predicted new set of preferences, he also seems to have approved of himself having them. (Ibid.)

In this case, New Person differs from Old Person not only with respect to her first-order preferences about what to do in particular situations (e.g., about whether to stay out late or come home early), but also by her second-order preferences about which first-order preferences to have. When an agent undergoes a personal transformation, both her first- and second- order preferences change in tandem. This is a purely *third-person* way of characterizing personally transformative experiences.

In contrast, Paul characterizes personally transformative experiences by highlighting first-personally accessible changes. One of her examples is this.

> Imagine that you have the chance to become a vampire. With one swift, painless bite, you'll be permanently transformed into an elegant and fabulous creature of the night. As a member of the undead, your life will be completely different. You'll experience a range of intense, revelatory new sense experiences, you'll gain immortal strength, speed and power, and you'll look fantastic in everything you wear. You'll also need to drink blood and avoid sunlight. (Paul 2014, p. 1)

On this view, when you become a vampire, the phenomenal character of your lived experiences, i.e., what it's like for you to undergo those experiences, will change. These changes may be connected to the changes in one's preferences: assuming that our preferences are sensitive to some extent to the phenomenal character of our experiences, radical changes in the phenomenal character of certain experiences could affect the preferences of the agent. In Paul's example, you can know that your preferences will change, but you don't have first person access to the way they'll change.

This affects the way you can think about the choice. You can know on the basis of the testimony of others how your preferences will change when you become a vampire. For instance, everyone you know may have already become a vampire, and may tell you that they love it. They might even tell you that if they were offered a pill to become human again, they would reject the offer without a moment's thought. Despite having all this information, you still won't know what it is like to be a vampire, since you can only come to know what it's like to be a vampire by

becoming one. Thus, since you lack access to the phenomenal character of a vampire's lived experiences, you lack access to certain concepts, pieces of information, or cognitive skills necessary for understanding why the preferences that come with vampirehood are rational to maintain (or, at least make sense to have) for someone who is a vampire. The case is intended to illustrate ways in which radical changes in phenomenal character can lead to changes in the very core of our personality: since the beliefs, desires and commitments that make us who we are can be shaped by what it is like to be us, certain types of radical changes in the phenomenal character of our lived experiences can personally transform us. Our focus in this essay is on this latter point: on the way in which radical changes in our lived experience can lead to changes in who we are.

Let an *epistemically transformative act* be an act that brings about an *epistemically* transformative experience. Let a *personally transformative act* be an act that brings about a *personally* transformative experience. Two features of personally transformative acts are worth getting clear on.

First of all, acts that epistemically *and* personally transformative raise problems for theories of practical rationality. REFS Those problems are not our concern here. In what follows, we focus exclusively on personally transformative acts, without addressing the question of whether or how they are also epistemically transformative acts, or how epistemic transformation relates to practical decisionmaking.

Second, some personally transformative acts might be preceded by a moment of conscious choice where the relevant agent decides to perform that act. But this may not always be the case. In particular, some transformative acts are not the product of conscious choice. Or, as Callard (2018) correctly observes, sometimes, we *gradually* transform our personalities without ever directly deciding to undergo that transformation. Yet, that doesn't mean that we are thoughtlessly drifting into these new personalities. Rather, in many of these cases – e.g., in a case where one becomes a mother, an artist, a lover of classical music etc. – one actively tries on new roles and activities which enable one to navigate the world using a new set of beliefs and preferences, or to experience the world in a new way. On Callard's account, we can aspire to rational self-transformation by aiming at it indirectly, guided by proleptic reasons. Our notion of a personally transformative act covers these kinds of cases (unchosen, unconsciously chosen, aspirational). Since the agent could still have preferences about such acts (whether or not they guide her actions), her act of undertaking the project counts as a personally transformative act, even if the agent never explicitly decided to perform a personally transformative act.

§2. Self-Creation

In this section, we defend the following claim.

> **The Principle of Act-Dependence.** If an agent performs a personally transformative act, then (ceteris paribus, or keeping all else fixed), she thereby creates a new self that wouldn't exist if she hadn't performed that transformative act.

To understand what this means, start with the notion of *self* that the principle invokes. For the purposes of this essay, we'll assume that an agent's *self* at a certain time is constituted by the core values, beliefs, desires, commitments, ideals, character traits, etc., that make her who she is at that time, including the first person phenomenology that is realized by having these core values, etc. Now, a change in these needn't be so dramatic that we are no longer inclined to think that the person who existed before the change exists any more. For instance, Sue may take a college course that changes her worldview radically, thus changing her plans about what to do with her life. But that needn't mean that the person that Sue was before taking the college course has ceased to exist. Plausibly, that person still exists, but she's no longer exactly who she was earlier. On this way of thinking, a person's life may be partitioned into intervals, corresponding to a series of successive selves, that in turn are suitably extended (collections of) temporal parts. Here's a more careful definition.

> **Self.** For any person *S*, a temporal part *x* of *S* that exists during an interval of time *t* counts as a *self* iff
>
> (i) *Prudential Status. x* has prudential status, i.e., it or its well-being can be an object of prudential concern for that and other temporal parts of *S*[4];
>
> *(ii) Constancy.* S's self-identity-conferring mental states remain unchanged during *t*, i.e., during any two sub-intervals *t1* and *t2* of *t*, the mental states that make *S* who she is at *t1* are the same as those that make her who she is at *t2*.
>
> (iii) *Maximality*. There is no interval of time *t** during which *S* exists, such that (i) *t** is distinct from *t*, (ii) *t* is a sub-interval of *t**, and (iii) *S*'s self-identity-conferring mental states remain (relevantly) unchanged during *t**.[5]

[4] One worry about views like this is that assigning temporal parts of a person prudential or moral status makes it difficult to explain why practical rationality or morality seems to require a temporal part of a person to subject itself to a small harm to protect a future temporal part from a greater harm; for the later temporal part's good fortune doesn't really compensate the earlier temporal part for the harm that it is subjected to. See Miller (2015) and Johnston (2016) for different versions of this worry. As Kaiserman (forthcoming) correctly notes, such problems only arise for stage-theoretic versions of perdurantism.

[5] Note that this notion of self is importantly different from a notion that Parfit (1987, pp. 301-6, pp. 326-8) proposes and Shoemaker (1999) defends. For Parfit, selves are person-stages united by strong psychological connectedness, e.g., direct memory connections. Since relations of strong psychological connectedness aren't transitive, a person at a certain time may have two different selves that overlap with each other. However, on our view, selves don't overlap with each other. Our conception is closer to a view of the self defended by Kristjansson (2010) and Strohminger and Nichols (2014), on which certain morally assessable beliefs, desires, dispositions, etc., of an agent

Let's explore this a bit.

Start with the idea of a temporal part of a person. If Sue is a person who is now in her forties, Sue-in-her-twenties is a temporal part of Sue. Why should we think that persons have such temporal parts? It follows from a view about persistence. According to *endurantists*, objects persist through time by being wholly present at each moment at which they exist. According to *perdurantists*, they persist by having temporal parts that are wholly present at each moment at which they exist. More precisely, let an object *x* be an *instantaneous temporal part* of an object *y* at *t* iff (i) *x* exists at, but only at, *t*; (ii) *x* is part of *y* at *t*; (iii) *x* overlaps at *t* everything that is part of *y* at *t*.[6] Perdurantists just say that for any object *x*, if *t* is a time at which *x* exists, there is an intantaneous temporal part of *x* at *t*. Endurantists deny this. To take a concrete case, suppose your desk has persisted through time. Perdurantists would say that it has done so by having instantaneous temporal parts at every moment at which it has existed. According to them, however, the desk was never wholly present at any of those moments. Endurantists would say that it persisted by being wholly present at every moment at which it has existed. Similarly, if Sue has persisted through time, then, according to perdurantists, she too has done so by having instantaneous temporal parts at every moment at which she has existed; endurantists would deny this. Perdurantism is preferable to endurantism as an account of objective persistence; for the former allows us to respond to certain puzzles satisfactorily.[7] (For an account of the subjective persistence of selves that is consistent with this approach, see Paul 2017.)

Typically, perdurantists claim that persons and material objects do not just have instantaneous temporal parts but also have temporal parts that themselves are extended temporally. The most liberal version of this view would be one according to which a person or a material object has temporal parts corresponding to any period of time at which she exists. This would mean that just as Sue-in-her-twenties counts as a temporal part of Sue, so does Sue-in-October-2018 or Sue-at-this-moment.

For the purposes of this paper, we'll take a temporal part of a person to count as a self just in case it satisfies three constraints. The first is *Prudential Status*: a self or its well-being can be an object of prudential concern for that person (or other temporal parts of that person). This seems plausible: the well-being of different temporal parts of a person can be the object of prudential concern for that person. In fact, some people might say something even stronger: namely, that all instantaneous temporal parts of a person ought to be objects of equal prudential concern for that

constitutes her self. However, for the sake of avoiding confusion, it's worth pointing out that Strohminger and Nichols use the terms "self" and "person" interchangeably, which, again, we don't.

[6] Sider (2001), ch. 3.

[7] Ibid, chs 4-5.

person.[8] The second feature is *Constancy*: if *t* is the period of time during which a self exists, the agent's identity-conferring mental states---the mental states that make her who she is---cannot change. We may imagine a fictional case, where a person changes her mental states at t1 so that, immediately after t1, she is no longer who she was before t1, but then at t2, she reverts back to her old mental states before t1. In such cases, the same self doesn't persist through the continuous interval of time that includes the time before t1 as well as the time after t2. The third feature is *Maximality*: a self is the *largest* temporal part of a person unified by sameness with respect to who she is.

Here's a way of illustrating this. Suppose Sue doesn't undergo any personal transformation in her thirties: who she is at 31 is no different from who she is at 32, or 33, and so on. According to *Maximality*, this means that Sue-at-31, Sue-at-32, Sue-at-33, etc. are parts of the same self, but aren't selves in their own right. This is because Sue-at-31 exists during an interval of time that has a non-empty intersection with a larger interval of time, namely the period during which Sue is in her thirties exists, and Sue's self-identity during her thirties isn't any different from her self-identity at 31.

How does this help us with the **Principle of Act Dependence**? We'll assume that a person *S* at two times *t* and *t** is realized by the *same, temporally extended self* if and only if the self that realizers her at *t* is no different from the self that realizes her at *t*.* We are interested in cases where, in virtue of *S* performing a transformative act at t, a new self comes into existence at t*. Such cases are those where, when we hold relevant background conditions fixed[9], the following counterfactual is true: if, at time t, *S* hadn't performed that transformative act, then the new self would not have been created at t*. (And, by extension, we assume that if the act had not been performed, the same self that realizes S at t* would persist from t to t*.) So we are focusing on cases for which the **Principle of Act-Dependence** holds.

[8] It's worth addressing a worry here. Saying that selves have prudential status doesn't commit us to a view on which separateness of persons doesn't matter in prudential reasoning. For Brink (1997), the assumption that persons are metaphysically separate units allows us to preserve the hybrid structure of prudence, which allows an agent to be biased in favour of herself but does not allow her to be biased in favour of any of her temporal parts. As Brink (2011) acknowledges, the fact that an agent is required by rationality not to be biased in favour of any of her temporal parts only motivates the requirement of *temporal impartiality*, i.e., the requirement that the agent should have equal concern for all of her parts, and not the stronger requirement of *temporal neutrality*, i.e., the view that the agent should attach equal weight to the well-being of her temporal parts. Even though Brink defends this latter requirement at a number of places, we think it is implausible: it leads to the same problems that the **Totalist Theory of Prudence** (discussed in Section 3) leads to. If only the requirement of temporal partiality is true, then there remains a non-trivial question for theories of prudence to settle, i.e., the question of how to distribute benefits and harms across different temporal parts of the same person. So, if prudential reasoning is concerned with this question of distribution, then it indeed may be right to say that different temporal parts of an agent, e.g., selves, can be objects of prudential concern.

[9] In particular, for simplicity, we are ruling out the possibility of overdetermination or preemption in transformative self-creation: for discussion of problems with preemption and overdetermination for reductive accounts of dependence, see Paul and Hall 2013. We are helping ourselves to this simplification because we are interested in formulating a puzzle as it relates to Parfit's identity problem, not in giving a fully reductive, independent account of transformative self-creation.

§3. **The Non-Identity Problem for Transformative Acts**

Parfit's (1987) version of the non-identity problem is motivated by two plausible claims.

> **The Person-Affecting Principle.** An act can be morally wrong only if it makes things worse for some existing or future person.
>
> **The Comparative Notion of Harm for Persons.** Suppose an act or an event brings a person into existence such that (i) the person wouldn't have existed in the absence of the act, (ii) the person's existence is avoidably flawed, and (iii) the person's existence is still worth having for the person. Then, this act or event does not make things worse for that person.[10]

We can apply these two principles to the following scenario.

> **The 14-Year-Old Girl.** This girl chooses to have a child. Because she is so young, she gives her child a bad start in life. Though this will have bad effects throughout this child's life, his life will, predictably, be worth living. If this girl had waited for several years, she would have had a different child, to whom she would have given a better start in life. (Parfit 1987, p. 358)

The intuition is supposed to be that the girl makes the morally wrong decision. How do we explain this? According the **Person-Affecting Intuition**, this can be true only if her act makes things worse for, or harms, the child. According to the **Comparative Notion of Harm,** the girl's act doesn't make things worse for, or harm, the child. For, if the girl had waited, a different child would be born, and this child's life (despite its flaws) is still be worth living. So, either we must reject the intuition that the girl does something wrong, or we must give up the **Comparative Notion of Harm** or the **Person-Affecting Intuition**.

We can create a similar problem with respect to transformative acts.

A. Two Principles

[10] Here, we are working with a notion of harm, according to which harming a person or a self involves making things for it. If we adopt a non-comparative notion of harm, we may reject this view; for example, see Shiffrin (1999). However, we can formulate a version of the same puzzle by substituting every occurrence of "make things worse" with "harm." Note, however, that Shiffrin will reject the relevant version of **Comparative Notion of Harm for Persons**, because she thinks that inflicting a harm to confer a pure benefit is morally wrong unless the subject of the harm consents to it. But her view has the implausible consequence that even when a person's life goes extremely well but involves some unavoidable harm, even then it's morally impermissible to bring that person into question.

Once again, we start with two principles. The first is an analogue of the **Person-Affecting Principle**.

> **The Self-Affecting Principle.** An agent can rationally prefer not to perform an available act *A* only if there exist an available act *B* and a current or possible future self *x* of the agent such that the expected well-being of *x* conditional on the agent's performing *B* is greater than the expected well-being of *x* conditional on the agent's performing *A*.[11]

Consider the following pair of cases.

> **Surgery I.** On Monday, you have to schedule a surgery to have your wisdom teeth removed, painfully but safely, under a weak local anaesthetic. You are certain about the following facts. There are two surgeons who could perform the surgery. The surgery will begin exactly at the same time no matter who performs it. But the first surgeon will take more time to perform the operation, so you will be in pain for a longer period of time. Whom will you pick as your surgeon?
>
> **Surgery II.** On Monday, you have to schedule a surgery to have your wisdom teeth removed, painfully but safely, under a weak local anaesthetic. You are certain about the following facts. There are two surgeons who could perform the surgery. The surgery will begin exactly at the same time no matter who performs it. Moreover, both surgeons will take exactly the same amount of time to perform the operation. Whom will you pick as your surgeon?

If you are rational, in **Surgery I** you will prefer not to pick the first surgeon, but in **Surgery II** you will remain indifferent between the two. The **Self-Affecting Principle** explains why. Since you know that the first surgeon in **Surgery I** will subject you to pain for a longer period of time, you know that in **Surgery I** that there is at least one (existing or future) self that would be made worse off if you picked the first surgeon. By contrast, you know that in **Surgery II** things won't be better or worse for your present or future selves in either of these scenarios. That is why it is irrational for you to prefer one surgeon to another.

The second principle is an analogue of the **Comparative Notion of Harm for Persons**.

> **The Comparative Notion of Harm for Selves.** Suppose an act or event results in the existence of a future self such that (i) the future self wouldn't have existed in the absence of the act (or omission), (ii) the future self's existence is unavoidably flawed, and (iii) the

[11] Here, and everywhere else, the expected well-being or harm (or any other kind of value) is calculated according to a credence function that is rational for the agent to currently adopt.

future self's existence is still worth having for that future self. Then, this act or event does not make things worse for that future self.

Here is a way of motivating this principle. Many disabled people report that from their own first-personal standpoint, they don't experience their disability as a harm, since (a) despite its challenges, disability doesn't make their lives unworthy of having, and (b) if they didn't have that disability, they wouldn't be who they actually are.[12] Here is a revealing passage from Emily Ladau who writes:

> I can't count how many times I've been asked variations of the question: "If there was a pill that could cure your disability, would you take it?" Though the short answer is a resounding "No!" I rarely get the chance to elaborate on the complex feelings and emotions that are behind my answer.
>
> I think "cure" is actually a rather loaded term in relation to my disability, because to cure something implies that you are returning the body to its normal state. My disability *is* my normal state. To cure me in accordance with the medical definition of the word would not only give me new abilities, but also essentially transform me into a whole new person. I can't imagine myself as an able-bodied person, because I never was an able-bodied person. I've embraced my disability as a huge facet of my identity, and I take pride in it.
>
> While I don't define myself solely by my disability, having a disability has undeniably shaped who I am. Without my lived experiences as a disabled person, I would be a completely different Emily. And as tough as certain aspects of my life have been, and though I know I will continue to face disability-related challenges throughout my life, I wouldn't trade my life for a minute. My disability has given me a place in a community and a culture; it has been the reason why I've had amazing adventures and unforgettable experiences. To walk freely up and down stairs for one day would *never* measure up to the things I've done *because* I have a disability. (Ladau 2013)

Building on this idea, we can create a case like this.

> **Disability I.** Suppose you went blind as a child because of an accident. Since then, your blindness has given you the benefit of certain valuable experiences that a sighted person wouldn't be able to have. For example, you hear and feel things that sighted individuals fail to notice. You experience the world differently and discover new ways to relate to your environment. It has also allowed you to become part of a community and a culture, which sighted people do not have access to. And you know that if the accident hadn't

[12] For discussion of this point, see Saigal et al. (1996), Albrecht and Devlieger (1999), Gill (2000) and Goering (2008), and Barnes (2016).

> happened and you retained your sight, you wouldn't have enjoyed these experiences, or formed the close relationships with other members of the blind community, which have shaped the way you see yourself and have enriched your life. Thus, you have come to see your own blindness as a blessing in disguise.

In this case, as Ladau puts it, perhaps you "wouldn't trade" your life for the life of your non-disabled counterpart who somehow avoided the accident. That is, the self that you are would not trade itself in for a new self. From your (self-ish) perspective, your disabled life – despite its challenges – is fulfilling and is not to be traded for a scenario where you don't exist at all. Unless we want to treat the selves realized by disabled people like Ladau as irrational, we should grant that this preference is rational. Moreover, as Barnes (2016) convincingly argues, we have very little reason to think that this is just some sort of adaptive preference or self-deception that doesn't reflect actual quality of life. So, assuming that your preference does provide evidence for your well-being, we should conclude that your life as a disabled self is worth living. Moreover, since you see your disability as a blessing in disguise and therefore don't experience it as a harm, this suggests that the self you are now wasn't *harmed* by the accident.[13] This lends plausibility to the **Comparative Notion of Harm for Selves.**

B. The Problem

Consider a slightly different scenario.

> **Disability II.** You learn from your doctor that you will become blind soon if you don't undergo cataract surgery. If you were to become blind, there wouldn't be any serious transition costs: fortunately, your family is extremely supportive, and you live in a society that doesn't treat blind people all that differently from the sighted. However, you will inescapably lose certain capacities: you won't able to read very many books, you won't be able to paint, and you won't be able to play the sport you love the most. At the same time, your blindness will allow you to have certain experiences that a sighted person couldn't have. And it will give you a place within a community and a culture that sighted people don't have access to. Finally, you are certain that your blindness will eventually change how you see yourself: your identity will at least be partly shaped by your blindness, and you'll be glad that you became blind.

[13] Harman (2009) argues that the preferences expressed by disabled people in these cases are *strongly person-affecting*: they are happy with their disabled lives, and don't identify with the people they would have been had they not been disabled. But such preferences, according to Harman, shouldn't give us reason to think that their lives just as good as that of their non-disabled counterparts. Even if Harman is right that a life of a non-disabled person is overall better than the life of a disabled person, it can still remain true that the disability in question doesn't harm the disabled person: given that she wouldn't exist without the disability, there's no clear sense in which *she* would have been better off without her disability.

If you choose not to have cataract surgery, you will be performing a transformative act: you will be radically changing who you are. According to the **Principle of Act-Dependence**, then, your decision to go blind will bring into existence a future self of yours that wouldn't exist if you chose to have the surgery. The lived experience of this future self is unavoidably flawed. Yet, at the same time, there is no reason to think that the life your blind self would have wouldn't be worth having.

Here is the problem. On the one hand, it seems tempting in this case to say that you can rationally prefer the act of undergoing cataract surgery to doing nothing. You know that you won't be able to do lots of things that you currently care about, enjoy, and excel at: your ability to play certain sports, to paint, to read a lot of books, etc. On the other hand, this claim is incompatible with the conjunction of the **Self-Affecting Principle for Selves** and the **Comparative Notion of Harm for Selves.** According to the **Self-Affecting Principle,** you can rationally prefer not to perform an act A only if there is another available act that is expected to make things one of your present or future selves better than A does. This means that if an act maximizes the well-being of your current and future selves, it is rationally impermissible for you to not to prefer that act. But in this case, your blind future self wouldn't exist if you underwent cataract surgery. Since that self's existence is unavoidably flawed but still worth having, according to the **Comparative Notion of Harm for Selves,** not having surgery does not harm your blind future self or make it worse off. Therefore, it is indeed irrational for you to prefer undergoing the surgery to not undergoing the surgery; in fact, you should be indifferent between the two options.

The upshot is this. If we accept the **Principle of Act-Dependence**, either we have to reject the intuition that it is practically rational for you to prefer to undergo the surgery in this case, or we have to give up one of the two principles we introduced (i.e., the **Self-Affecting Principle** and the **Comparative Notion of Harm for Selves).** In this respect, the problem has a structure that is exactly analogous to the non-identity problem.

It's worth noticing how this problem of transformative choice is different from other problems raised by scenarios of transformative choice. Two such problems ought to be mentioned.

First of all, Paul (2014) argues that epistemically transformative experiences reveal a problem for standard decision theory. Consider an epistemically transformative experience like tasting the durian fruit for the first time. If an agent is given the choice of undergoing such an experience, can it be practically rational for her to take it? Paul's claim is that since an agent cannot know what that experience will be like before undergoing to the experience itself, she cannot accurately assign subjective value to the possibility where she undergoes that experience. But standard decision theory---which requires us to maximize expected value of some kind---can help us generate an ordering of preferences over options only if we can assign values to the

different outcomes of taking each option. So, standard decision theory is silent about cases like this.[14]

Personally transformative experiences, e.g., the experience of having a child, give rise to a different problem. In practical decision-making contexts, they can change the relevant agent's core values and preferences in unforeseeable ways. Writing about the transformative choice of becoming a vampire, Paul (2014) says: "So until you actually become a vampire, you cannot know if the values of any of the relevant outcomes will swamp the rest, or how to compare the subjective value of being a vampire to the subjective value of being human, or which preferences about the outcomes that you'd have as a vampire will be the same as the preferences you have as a human" (p. 44). The idea is this: when an agent makes a decision, she makes a decision on behalf of her current as well as her future selves. But if the agent can't accurately forecast the preferences of her future selves (after performing an act), she cannot take those preferences into account while making her decision.[15]

The problem we are concerned with doesn't depend on the agent's prior ignorance about what experiences she will undergo as a result of her transformative act, or what her future preferences or values will be. As we noted in **§**1, our focus is on a different issue: the way that radical changes in lived experience can create a new self. It's a problem simply about which selves will exist as a result of an agent's transformative act. For instance, in **Disability II**, we can assume that the agent somehow (by some feat of imagination or a blindness simulator) has full access to what it will be like to be her blind self, or what the values or preferences of that self will be. Still, the problem we have raised will persist. As long as the existence of her blind self is unavoidably flawed due to the lack of certain abilities that only the sighted possess, it will be rational for her current self to prefer undergoing the cataract surgery to not undergoing the surgery. And yet, given that her future blind self's existence is worth having, it will be hard to say that not undergoing the cataract surgery was bad for, or harmed, her blind self.

§4. Possible Responses

Let us consider some responses to the problem posed in the last section.[16]

[14] In subsequent discussion, some have asked whether this can be solved by appealing to knowledge norms of action: for a sample of the literature, see Pettigrew (2015, 2016), Dougherty, Horowitz, and Sliwa (2015), Moss (2016), Fraser (2018), and Isaacs (forthcoming).

[15] This problem should be distinguished from a different problem that doesn't depend on ignorance in the same way: namely, the problem of making decisions in scenarios where the agent's values change over time. For discussion of the problem raised by Paul, see Briggs (2015) and Pettigrew (ms.).

[16] An initially tempting response to the problem: if psychological continuity is what matters for survival, then we might think that a person cannot survive her personal transformation (at least if her core beliefs, desires, commitments, etc., track psychological continuity). Parfit's (1987, pp. 326-8) example of the 19th century Russian nobleman seems like a good example of this. Since it is rationally permissible for a person not to prefer her own death, it is rationally permissible for her not to perform a personally transformative act. Two responses. First, we

A. Denying the Intuition

The first response is to reject the intuition that it is rational for the agent to prefer that she not go blind. If you aren't convinced by the somewhat natural example described earlier, consider a variant of the same case.

> **Disability III.** Your ophthalmologist gives you two options. She could either painlessly blind you, or you can retain your imperfect, but fairly well-functioning, vision. The rest remains the same as in **Disability II**.

Here, it certainly seems rational for you to prefer not to be blinded by your ophthalmologist. But note that this case isn't all that different from **Disability II**. Just as you have two options in that scenario, so also you have two options in this case: either to go blind or not to. The consequences of these two options are the same in the two cases. If it is rational for you to disprefer blindness in this case, it should also be rational for you to disprefer blindness in **Disability II**.

However, someone who rejects the intuition that it is rational for you to disprefer blindness in **Disability II** may still reject the same intuition with respect to **Disability III**. They may offer the following error theory.

> **Error Theory I.** We are conditioned to believe that disability is intrinsically bad (i.e., bad not only because of the social disadvantages that it gives rise to, but bad in itself). This belief is what explains our intuitions with respect to **Disability II** and **III**. But this belief is false.[17]

But the problem we are raising doesn't really have anything to do with disability. So, consider another example:

> **Procreation**. If you have children, you will also realize a new self: your love for your children will change who you are. At the same time, your financial situation will also deteriorate: as a result, you will have to work longer hours, eat less healthy food, and will

have already argued that a person can survive personal transformations. Second, even if this diagnosis were plausible in some cases, it doesn't generalize well. It seems overly strong to claim that a person who is thinking of having a child or a similar personally transformative experience is contemplating death. Or consider examples where a personally transformative act brings about a gradual transformation. The person who exists at any stage of such a transformation may be strongly psychologically connected with the person who exists at any immediately preceding stage of that transformation, making the person who exists before the transformative act psychologically continuous with the one after the transformation is complete. Such a case shouldn't therefore be treated as a case of death either. Thanks to Joe Horton for discussion.

[17] This is connected to an error theory that Barnes (2016) defends in relation to Parfit's 'handicapped child case'.

> have much less time for the hobbies or pastimes that make your life slightly more exciting. But your life as a parent, despite being worse than your previous life, will be full of other valuable experiences that would make it worth having.

This case is analogous to **Disability I** and **II**. On the one hand, we may be tempted to say that it is rational for you to strictly prefer not to have children in this scenario. On the other hand, this claim is incompatible with the conjunction of the **Self-Affecting Principle** and the **Comparative Notion of Harm.** According to the **Self-Affecting Principle,** you can prefer not to perform an act A only if A makes at least one of your present or future selves worse off or harms them. This means that if an act maximizes the well-being of your present and future selves, it is rationally permissible for you to prefer that act. But the future self that comes into being when you have children wouldn't exist if you didn't have children. Moreover, its existence is worth having. So, according to the **Comparative Notion of Harm,** your future self in the scenario where you perform the transformative act isn't worse off, and isn't harmed. Therefore, it cannot be rational for you to prefer not to have children in this case.

There might be a different error theory that we could appeal to.

> **Error Theory II.** What explains our intuitions with respect to **Disability II** and **III** is a kind of *irrational* status quo bias, i.e., a preference for the current state of affairs to continue.

This error theory is slightly more plausible than the previous one; for it can also explain why it might seem rational for the blind person in **Disability I** to prefer not to undergo the cataract surgery and gain sight.[18] The general strategy behind this error theory is to maintain that what makes such preferences seem rational to us is that we often irrationally prefer that the current state of affairs continue as it is, and take ourselves to be rational to do so.

However, this isn't entirely obvious. Following Parfit (2011), we may distinguish two kinds of views about the rationality of preferences. According to *objective* theories, our reasons to prefer one outcome to another ultimately depend on the features of these outcomes. According to *subjective* theories, we have no reasons for our preferences, except in a derivative case where we prefer one outcome to another because the former helps us fulfil some other preference we have. Now, if we hold a subjective view about the rationality of preferences, then in a case like **Disability II,** a status quo bias may indeed be rational. After all, you might attach greater value to a situation where you retain your sight and thus are able to continue the activities that you

[18] Bostrom and Ord (2006) offer an error-theory of this sort to explain why seems wrong to enhance human intelligence by genetic engineering. For a related discussion of the rationality of status quo bias, see Nebel (2015). Some of our discussion is based on things Nebel says.

currently value than to a situation where you lose your eyesight and are unable to continue those activities. If we are subjectivists, we cannot dismiss such preferences as irrational.[19]

Suppose we are objectivists about the rationality of preferences. Even then, it's not obvious why we couldn't vindicate the rationality of your preferences in **Disability II.** According to the description of **Disability II**, your blind self is worse off than your possible sighted future self: the latter has certain capacities that the former lacks. These capacities, on an objective view, may contribute to the well-being of the sighted self to an extent that cannot be compensated by the other benefits that the blind self could receive. If this description of the case is correct, then the blind self, even though its existence is worth having, may indeed be worse off than the sighted self. If the choice is between a situation where your future self is better off and a situation where your future self is worse off, then it does seem rational to prefer that you retain your eyesight. In this sense, this case doesn't seem obviously like a case of irrational status quo bias.

B. Denying the Comparative Notion of Harm for Selves

Another strategy involves denying the **Comparative Notion of Harm for Selves.** Someone who adopts this strategy would have to say that in a case like **Disability I,** the disabled future self of the agent is genuinely harmed by her accident, even though she isn't in a position to recognize the harm itself.

[19] Perhaps, there's room for rejecting this response. For example, the subjectivist about the rationality of preferences could impose a diachronic constraint whereby an agent is required to take into account the values or preferences of her future selves into account while making her decisions. So, in **Disability II,** given that the agent's future blind self (if brought into existence) will prefer its own existence to its non-existence, the agent's current self might have *some* reason to prefer that that future self exists. This response doesn't immediately convince us: we need to know more about what this diachronic constraint is. First, suppose the diachronic constraint is some sort of "I'll be glad I did it" principle: if an agent is rationally certain that if she performs an act, she'll be (rationally) glad she did it, then she is required by rationality to prefer to perform that act. This principle (as Harman (2009) has convincingly argued), is questionable especially in the kinds of cases we are discussing. Setting these cases aside, this principle also leads to inconsistent verdicts.

Second, suppose the relevant diachronic constraint is some kind of a "linear pooling" principle: namely, that the expected value of an act (which partially determines the agent's preferences about it) must be calculated in light of a value function that is formed by aggregating the value functions of the agent's present self *and* the future selves that will come to exist if the agent performs the relevant act. This principle is more plausible: as Pettigrew (2019) shows, if an agent doesn't conform to such a constraint, she will be predictably exploitable. But this account fails to make any concrete prediction about **Disability II**: depending on the weight that each value function gets, it could still be rational for the agent to continue to prefer that she retain her sight. So, the challenge for the opponent would still be to come up with an appropriate method of aggregation that predicts that this preference isn't rational.

First of all, a response of this sort is only available to a person who accepts some form of *welfare objectivism*. Welfare objectivism is the view that there is a certain thing or certain things – a set of freedoms, functions, or capabilities, a list of basic goods, etc. – that constitute the good life, independently of whether these things are desired by the particular person who lives the relevant life or can be said be happy without them. A person who wishes to say that you (as you exist years after your accident) in **Disability I** are in fact harmed by your accident cannot accept a theory of welfare on which welfare is a matter of happiness or desire satisfaction; for all your (actual or suitably idealized) preferences may indeed be satisfied in this case, and you may reasonably be called happy.

However, welfare objectivism isn't sufficient for us to deny the **Comparative Notion of Harm for Selves.** What we need is a notion of harm on which even though your later self couldn't be better off in **Disability I**, that is, in a case where the accident that disables you still harms your later self insofar as it deprives your later self of certain freedoms, capabilities, or functions. Perhaps a defender of this position could come up with a list of goods, such that if an act or event prevents an agent from possessing certain items on that list at a certain time, the self that exists at that time is harmed.

The problem is this. Even if we could come up with a list like that, an agent who finds herself in a situation of transformative choice would not necessarily *know* what that list of goods is, or which items on the list are such that lacking them would constitute a harm. For example, in **Disability II**, if you rationally take the testimony of disabled people seriously, you might be rationally extremely confident that their disability doesn't actually constitute a harm. Thus, by your lights, the expected harm that disability causes to your present or future selves may indeed be negligible (or, at least, need not be greater than the harm that the surgery causes). Therefore, according to the **Self-Affecting Principle**, you cannot rationally prefer to undergo the cataract surgery; for the expected harm that going blind poses isn't more than the expected harm posed by the cataract surgery. This, in turn, will conflict with the intuition that you are rational to prefer to undergo the cataract surgery. Thus, even if the **Comparative Notion of Harm for Selves** is false, the problem that we saw in **Disability II** can be raised here again.

C. Denying the Self-Affecting Principle

This discussion makes salient a different strategy for solving the problem: rejecting the **Self-Affecting Principle**. According to this principle, an agent can rationally disprefer an available act *A* only if, in comparison with other available acts, performing *A* is expected to make things worse for, or harms, her present self or one of her future selves. However, there are alternative, plausible, theories of prudence that are incompatible with this principle.

Start with the following simple theory.

> **The Presentist Theory of Prudence I.** It is rationally permissible for an agent to prefer an available act *A* to an available act *B* iff the expected well-being of the agent's current self conditional on her performing A, is greater than the expected well-being of the agent's current self conditional on her performing B.

It's unclear what this theory entails in cases like **Disability II**. In that scenario, there are two acts that are available to the agent: going blind or not going blind. While it is clear that these acts will affect the agent's future self, it's unclear whether they will affect the well-being of the agent's current self. At least, it seems coherent to say that they won't affect the agent's current well-being at all. So, this version of the present theory won't do at all. We may replace it with:

> **The Presentist Theory of Prudence II.** It is rationally permissible for an agent to prefer an available act *A* to an available act *B* iff the expected value of *A* is greater than the expected value of B, where the value of the different outcomes of *A* and *B* is fixed by the agent's current *actual* preferences.

This proposal is substantively different from the previous presentist theory, since it appeals not to the well-being of the agent's current self but rather to the actual preferences of the agent's current self. It might solve the problem that we are dealing with. In **Disability II**, you might already have preferences about blindness: you might actually prefer to have sight to going blind. Given these preferences, undergoing the cataract surgery may indeed uniquely maximize expected value. But note that this solution won't always work. What if you haven't given the matter of going blind any thought at all? In such a scenario, you may not have a preference either way, so no value can be assigned to the different outcomes of the available acts. As a result, the expected value of the options will be undefined.[20] Moreover, we cannot solve the problem of value gap by appealing to your rational preferences. Since the theory is supposed to predict what your preferences about going blind and not going blind should be, the theory cannot generate this prediction by appealing back to your rational preferences about going blind and not going blind. Doing so would make the theory circular.

---

[20] This theory is terrible in other ways too. Consider a case of future-discounting.

> **Surgery III.** I have been given the option of undergoing a painful surgery under a weak local anaesthetic either tomorrow or in a month. If I undergo the surgery tomorrow, it will last an hour. But if I undergo the surgery in a month, it will last four hours, so the pain will be four times as much.

In this scenario, if it seems irrational for me to prefer to undergo the surgery in a month. But if I am a future discounter and actually want pain to be further away in the future than nearby, then, given my current preferences, the **Presentist Theory of Prudence II** could entail that it's rationally permissible for me to prefer to undergo the surgery in a month. This seems bad, or at least, needs more of a defense than we can muster.

What this shows is that the problem cannot solved by appealing to the agent's current well-being or preferences. We might hope that it can be solved by a theory of prudence that takes into account the agent's future well-being (or preferences). Consider the following theory of prudence.

> **The Totalist Theory of Prudence.** It is rationally permissible for an agent to prefer an available act *A* to an available act *B* iff the expected *net* well-being of the agent's possible current and future selves conditional on her performing A, is greater than the expected *net* well-being of the agent's possible current and future selves conditional on her performing B.[21]

We can see how this theory easily takes care of cases like **Disability II** and **Procreation**. In those situations, the agent has no uncertainty: she knows exactly what will happen if she takes any of her options. In **Disability II**, she is certain that if she goes blind, she will bring into existence a future self which will be worse off in comparison with the future self that will exist if she undergoes the surgery. So, given that the net well-being of the possible current and future selves that will exist if she undergoes cataract surgery is greater than the net well-being of those selves that will exist if she doesn't undergo the surgery, it's rationally permissible for the agent to prefer to undergo the surgery in this case. A similar diagnosis will allow us to address **Procreation**.

However, this view faces the same problem that totalist theories, i.e., versions of utilitarianism, face in population ethics: namely, the *repugnant conclusion*.[22] To see why, consider the following case.

> **Pills.** You are twenty-five now, and in relatively good health. The bad news is that you have been diagnosed with a life-threatening disease. There are two pills available to you. Pill A will allow you to live for about twenty-five years in roughly the same physical conditions that you are now in. But if you take Pill B, your health will deteriorate

[21] This version of the totalist theory only takes into account the well-being of the agent's future selves. But we can offer a preference-based analogue of this theory.

> **The Preference-Based Totalist Theory of Prudence.** It is rationally permissible for an agent to prefer an act A to an B iff the expected value of A is greater than the expected value of B where the value of any outcome of A or B is just the sum of different values assigned by the agent's current and future selves to that outcome.

If an agent's preferences over outcomes can be represented as cardinal utilities, this proposal can work. However, this proposal might raise intra-personal analogues of the problem of interpersonal utility comparisons. Moreover, it will be subject to the same problem that the **Totalist Theory of Prudence** is subject to.

[22] See Parfit (1987, ch. 17).

> radically and you will be bedridden for the remainder of your life. But you will be able to live for sixty more years, and your existence will be worth having.

Suppose you will undergo personal transformations every year, so your self will change every year for the rest of your life. And you know this. You also know that taking Pill B will diminish your annual well-being exactly by half in comparison with what it otherwise would be. So, you can be rationally certain that all else equal, the net well-being of your current and future selves conditional on your taking Pill A will be $k + 25x$, where k is the well-being of your current self, while x is you're the well-being of each of your annual future selves (if things continue as they are now). By contrast, if you take Pill B, your net well-being of your current and future selves will be $k + 60\ (x/2)$. According to the **Totalist Theory of Prudence**, rationality requires you to prefer Pill B, and thus be bedridden for sixty years. This seems bad.

We may try to fix this problem by adopting:

> **The Averagist Theory of Prudence.** It is rationally permissible for an agent to prefer an available act *A* to an available act *B* iff the expected *average* well-being of the agent's possible current and future selves conditional on her performing A, is greater than the expected *average* well-being of the agent's possible current and future selves conditional on her performing B.[23]

This will avoid the problem that **Pill**-style cases raise for the **Totalist Theory of Prudence**. Since the average well-being of your future selves when you take Pill B is exactly half of the average well-being of your future selves when you take Pill A, the average well-being of your current and future selves in the former scenario is greater than the average well-being of your current and future selves in the latter scenario. So, you are required by the **Averagist Theory of Prudence** to take Pill A.

While this might be a solution to the problem raised above, it faces another problem.

> **The Life-Prolonging Drug.** You know that the rest of your life will be exceptionally wonderful. Now, you are offered a drug that will allow you to live for one more day than you are supposed to: on that day, you will be slightly worse off than you were earlier, but things will still be quite nice.

There are two possible situations: in one, you take the drug, and in the other, you don't. Suppose you will undergo personal transformations every year, so your self will change every year for the rest of your life. And you know this. In the first situation, the average well-being of your current

[23] Once again, we can construct a preference-based version of this theory as we did for the totalist theory, and it would be subject to the same problem that we raise below.

and future selves is lower than it is in the second; for your well-being on the last day of your life in the first situation is lower than what it was earlier. So, the **Averagist** is committed to saying that you are required not to take the drug in this case. Once again, this seems too strong: while you might be rationally permitted not to prolong your life, it doesn't seem as if you are required by rationality not to do so.

The challenge here is to come up with a principle of prudence that satisfies two desiderata. First of all, it should allow us to bring into existence a non-existent future self that is better off in comparison with a distinct non-existent self. Second, this principle shouldn't allow us to prolong human lives in cases by creating additional future selves just because those future selves are guaranteed to have an existence that is barely worth having. Yet it should allow us to do when they have a reasonably high degree of well-being. Similar attempts at finding similar theories have been unsuccessful in population ethics, and we expect the same problems to arise for proposed principles of prudence. [24]

§5. Conclusion

Let's take stock. In this essay, we began by showing that scenarios of transformative choice can create a problem that is exactly analogous to the non-identity problem. We then went on consider three possible responses to this problem and showed that none of them obviously succeed. This discussion has two consequences: the first for population ethics, and the second for theories of prudence.

First, since we have shown that there is an intrapersonal analogue of the non-identity problem, we suggest that the non-identity problem has nothing in particular to do with population ethics. It belongs to a more general class of problems that arise whenever an agent faces a choice of creating another agent (whether it's a self or a person) whose existence would be unavoidably flawed but still worth living. Second, the discussion poses a challenge for existing theories of prudence or practical rationality. Standard theories of prudence require the practically rational agent to prefer acts that maximize expected value of some sort. Our intrapersonal analogue reveals a problem with almost all versions of this theory.

References

Adler, M. (2012). *Well-Being and Fair Distribution: Beyond Cost-Benefit Analysis*. Oxford: Oxford University Press.

[24] For a clear survey of these attempted solutions, see Arrhenius, Ryberg, and Tännsjö (2017).

Albrecht, G.L. an G. Devlieger (1999). The disability paradox: high quality of life against the odds. *Social Science and Medicine*, pp. 977–988.

Arrhenius, Gustaf and Rabinowicz, Wlodek (2015). The value of existence. In: Hirose, Iwao and Olson, Jonas, (eds.) *The Oxford Handbook of Value Theory*. Oxford: Oxford University Press, pp. 424-444.

Arrhenius, Gustaf, Ryberg, Jesper and Tännsjö, Torbjörn (2017). The Repugnant Conclusion. *The Stanford Encyclopedia of Philosophy* (Spring 2017 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/spr2017/entries/repugnant-conclusion/>.

Barnes, Elizabeth (2009). Disability and adaptive preference. *Philosophical Perspectives* 23 (1):1-22.

Barnes, Elizabeth (2016). *The Minority Body: A Theory of Disability*. Oxford: Oxford University Press.

Bostrom, Nick & Ord, Toby (2006). The reversal test: Eliminating status quo bias in applied ethics. *Ethics* 116 (4):656-679.

Briggs, R. A. (2015). Transformative Experience and Interpersonal Utility Comparisons. *Res Philosophica* 92 (2):189-216.

Brink, David O. (1997). Rational Egoism and the Separateness of Persons. In J. Dancy (ed.), *Reading Parfit*. Blackwell. pp. 96-134.

Brink, David O. (2011). Prospects for Temporal Neutrality. In Craig Callender (ed.), *The Oxford Handbook of Philosophy of Time*. Oxford: Oxford University Press.

Bykvist, K. (2007). The Benefits of Coming into Existence. *Philosophical Studies*, 135 (3): 335–362.

_____ (2015). Being and Wellbeing. In I. Hirose and A. Reisner (eds.), *Weighing and Reasoning*. Oxford: Oxford University Press, pp. 87–94.

Callard, Agnes (2018). *Aspiration: The Agency of Becoming*. Oxford: Oxford University Press.

Dougherty, Tom, Horowitz, Sophie & Sliwa, Paulina (2015). Expecting the Unexpected. Res Philosophica 92 (2):301-321.

Egan, Andy (2006). Secondary Qualities and Self-Location. *Philosophy and Phenomenological Research* 72 (1):97-119.

Fleurbaey, M. & Voorhoeve, A. (2015). On the Social and Personal Value of Existence. In I. Hirose and A. Reisner (eds.), *Weighing and Reasoning*. Oxford: Oxford University Press, pp. 95–109.

Fraser, Rachel Elizabeth (2018). Stakes sensitivity and transformative experience. *Analysis* 78 (1):34-39.

Gill, Carol J. (2000). Health Professionals, Disability, and Assisted Suicide: An Examination of Empirical Evidence. *Psychology, Public Policy, and Law*, 6(2) 526–45.

Goering, S. (2008). 'You say you're happy, but …': Contested Quality of Life Judgments in Bioethics and Disability Studies", *Journal of Bioethical Inquiry*, 5: 125–135.

Harman, Elizabeth (2009). "I'll be glad I did it" Reasoning and the Significance of Future Desires. *Philosophical Perspectives* 23 (1):177-199.

Isaacs, Yoaav (forthcoming). The problems of transformative experience. *Philosophical Studies*.

Jackson, Frank (1982). Epiphenomenal qualia. *Philosophical Quarterly* 32:127-136.

Johnston, Mark (2016). The Personite Problem: Should Practical Reason Be Tabled? *Noûs* 50 (4):617-644.

Kaiserman, Alex (forthcoming). Stage theory and the personite problem. *Analysis*.

Kavka, Gregory (1981). The Paradox of Future Individuals. *Philosophy & Public Affairs*, 11: 93–112.

Kristjánsson, Kristján (2010). *The Self and its Emotions*. Cambridge University Press.

Lewis, David (1990). What experience teaches. In William G. Lycan (ed.), *Mind and Cognition*. Blackwell. pp. 29--57.

Ladau, Emily (2013). The Complexities of Curing Disabilities. Blogpost on *Words I Wheel By*. URL = https://wordsiwheelby.com/2013/08/complexities-of-cures/

Loar, Brian (1990). Phenomenal states. *Philosophical Perspectives* 4:81-108.

McMahan, J. (1981). Problems of Population Policy. *Ethics*, 92: 96–127.

—— (2009). Asymmetries in the Morality of Causing People to Exist. In M. Roberts & D. Wasserman(eds.) *Harming Future Persons: Ethics, Genetics and the Nonidentity Problem*. Berlin: Springer.

—— (2013). Causing People to Exist and Saving People's Lives. *The Journal of Ethics*, 17: 5–35.

Miller, Kristie (2015). Prudence and Person-Stages. *Inquiry: An Interdisciplinary Journal of Philosophy* 58 (5):460-476.

Moss, Sarah (2016). *Probabilistic Knowledge*. Oxford University Press.

Nebel, Jacob M. (2015). Status Quo Bias, Rationality, and Conservatism about Value. *Ethics* 125 (2):449-476.

Olson, Eric T. (1997). *The Human Animal: Personal Identity Without Psychology*. Oxford: Oxford University Press

Ostrovsky, Y., Meyers, E., Ganesh, S., Mathur, U., & Sinha, P. (2009). Visual parsing after recovery from blindness. *Psychological Science*, *20*(12), 1484-1491.

Parfit, Derek (1987). *Reasons and Persons.* Oxford: Clarendon Press.

Parfit, Derek (2011). *On What Matters*. Volumes I and II. Oxford: Oxford University Press.

Paul, L. A. (2014). *Transformative Experience*. Oxford: Oxford University Press.

Paul, L. A. (2017). "The Subjectively Enduring Self", *Routledge Handbook of the Philosophy of Temporal Experience*, ed. Ian Phillips. Routledge.

Pettigrew, Richard (2015). Transformative Experience and Decision Theory. *Philosophy and Phenomenological Research* 91 (3):766-774.

Pettigrew, Richard (2016). Transformative Experience, by L. A. Paul. *Mind* 125 (499):927-935.

Pettigrew, Richard (ms.) *Choosing for Changing Selves*. Bristol University.

Rabinowicz, Wlodek (2009). Broome and the intuition of neutrality. *Philosophical Issues* 19 (1):389-411.

Saigal, Saroj and P. Rosenbaum (1996). Health Related Quality of Life Considerations in the Outcome of High-Risk Babies. *Seminars in Fetal and Neonatol Medicine*, 1(4): 305–312.

Shiffrin, Seana (1999). Wrongful Life, Procreative Responsibility, and the Significance of Harm. *Legal Theory* 5 (2):117-148.

Strohminger, Nina & Nichols, Shaun (2014). The essential moral self. *Cognition* 131 (1):159-171.

Ullmann-Margalit, Edna (2006). Big decisions: opting, converting, drifting. *Royal Institute of Philosophy Supplements* 58: 157-172.

Woodward, James (1986). The Non-Identity Problem. *Ethics*, 96: 804–31.