What You Can't Expect When You're Expecting*

L. A. Paul

Abstract: It seems natural to choose whether to have a child by reflecting on what it would be like to actually have a child. I argue that this natural approach fails. If you choose to become a parent, and your choice is based on projections about what you think it would be like for you to have a child, your choice is not rational. If you choose to remain childless, and your choice is based upon projections about what you think it would be like for you to have a child, your choice is not rational. This suggests we should reject our ordinary conception of how to make this life-changing decision, and raises general questions about how to rationally approach important life choices.

It seems natural to choose whether to have a child by reflecting on what it would be like to have one. I argue that choosing on this basis is not rational, raising general questions about our ordinary conception of how to make this life-changing decision.[1]

1 Deciding Whether to Start a Family

Scenario: You have no children. However, you have reached a point in your life when you are personally, financially and physically able to have a child.[2] You sit down and think about whether you want to have a child of your very own. You discuss it with your partner and contemplate your options, carefully reflecting on the choice by assessing what you think it would be like for you to have a child of your very own and comparing this to what you think it would be like to remain childless. After careful consideration, you choose one of these options:

* This paper is dedicated, with much love, to my two children.

[1] My point has larger consequences for how we plan our futures and attempt to become the kind of person we think we want to be. I develop the discussion and show how my argument applies to a wide range of decisions and life experiences in Paul (2014).

[2] In this example, I am assuming that you and your partner are physically able to have a child. Below, I will consider an implication of my argument for those who cannot physically produce a child. For simplicity, I am not discussing the decision to adopt a child, although I believe that a version of my argument would apply.

Res Philosophica, *Vol. 92, No. 2, April 2015, pp. 149–170*
http://dx.doi.org/10.11612/resphil.2015.92.2.1

For: You decide to have a child.
Against: You decide to remain childless.

The way you went about making your choice seems perfectly apt. It follows the cultural norms of our society, where couples are encouraged to think carefully and clearly about what they want before deciding that they want to start a family. Many prospective parents decide to have a baby because they have a deep desire to have children based on the (perhaps inarticulate) sense that having a child will help them to live a fuller, happier, and somehow more complete life.[3] While many people recognize that an individual's choice to have a child has important external implications, the decision is thought to necessarily involve an intimate, personal component, and so it is a decision that is best made from the personal standpoints of prospective parents.[4] Guides for prospective parents often suggest that people ask themselves if having a baby will enhance an already happy life, and encourage prospective parents to reflect on, for example, how they see themselves in five and ten years' time, whether they feel ready to care for and nurture the human being they've created, whether they think they'd be a happy and content mother (or father), whether having a baby of their own would make life more meaningful, whether they are ready for the tradeoffs that come with being a parent, whether they desire to continue with their current career plans or other personal projects, and so on.[5]

This assessment of one's prospects and plans for the future is a culturally important part of the procedure that one is supposed to undergo before attempting to become pregnant. Since (in the usual case) the parents assume primary responsibility for the child they create, it seems appropriate to frame the decision in terms of making a personal choice, one that carefully weighs the value of one's future experiences.[6] People often frame the decision this way when they make this choice, and more importantly for my purpose here, we are (culturally speaking) supposed to frame the decision this way. Given the magnitude of the responsibilities we are

[3] This may or may not be the same as increasing one's "life satisfaction" or "meaningfulness." I will return to this at the end of the paper.

[4] I am ignoring external, nonphenomenal factors one might weigh when making a choice about whether to procreate, such as the values of environmental impact or population control. A version of my argument that takes these factors into account holds unless these values are supposed to swamp the personal phenomenal values.

[5] Sixty seconds of googling will turn up plenty of examples. Claims like "You long to nurture and raise a little person who will likely be similar to you but still completely unique. Perhaps, you and your spouse feel like something is still missing, and a baby would complete your vision of family" (http://newlyweds.about.com/od/havingababy/tp/Reasons-to-Have-Kids.htm). Or see Caplan (2011). A different kind of example is provided by initiatives that try to convince young teens that they are not ready to become parents by giving them baby dolls to care for that need constant attention, wake up three times a night, etc.

[6] The importance of this sort of reflective approach is underscored by the general cultural prescription against unplanned pregnancies and in the attention given to family planning by many social and religious organizations.

considering taking on, we are *supposed* to think carefully about the personal implications of the choice. Many choose to have a child. Many prefer to remain childless.

2 Decision Theory: A Normative Model

When we make a choice to do something, we make a decision: we consider various things we might do and then choose to do one of them, and decision theory provides the best account of rational decision-making. Ideal agents in ideal circumstances make choices rationally by conforming to the models of an idealized decision theory. To make a choice rationally, we first determine the possible outcomes of each act we might perform. After we have the space of possible outcomes, we determine the value (or utility) of each outcome, and determine the probability of each outcome's occurring given the performance of the act. We then calculate the expected value of each outcome by multiplying the value of the outcome by its probability, and choose to perform the act with the outcome or outcomes with the highest overall expected value.

Now, decisions made by real agents in real-world circumstances do not conform to this standard model. Ordinary reasoners may be imperfect reasoners; their reasoning may only imperfectly conform to the way an ideal rational being would reason, and their assessments of the values of the outcomes may only imperfectly conform to their actual values. A more realistic version of a decision-theoretic approach, that is, what I'll call a *normative* decision theory, can capture norms for ordinary successful reasoning. If we can glean approximate values for our outcomes and apply the right decision theoretic rules, we can conform to the ordinary standard for rational decision-making. Decisions made by ordinary people can be rational if they conform to the realistic standards set by a normative decision theory, where such standards make allowances for a certain amount of approximation, ignorance, uncertainty, and mistaken beliefs.[7]

For example, when considering an outcome, perhaps we can do no better than glean its approximate expected value. After all, it is probably impossible for a person to calculate the expected value of each outcome with precision. And perhaps we do not know about all the possible outcomes. But we can approximate a rational choice by choosing between approximate expected values of the relevant or the most important outcomes. A normative decision theory describes the range and combination of rules and standards that agents must meet for their decisions to be rational, normatively speaking. It thus provides a *normative model* that real agents can

[7] For simplicity, I am assuming a 'realist' interpretation of decision theory according to which the utility of outcomes corresponds to a real psychological quantity, such as the individual's strength of preference for outcomes or her perception of how good each outcome is. (I am indebted to Lara Buchak here.)

conform to so that their decisions are rational by our lights.[8] In this paper, I will assume that we want to meet the standard for normative rationality when we make the decision of whether or not to have a child.

In any non-ideal case, complicating features may be present. For example, sometimes outcomes have equal expected values. Then no unique act is the rational one to choose. Sometimes expected values are metaphysically indeterminate. Then it is metaphysically indeterminate which act is the rational one to choose. Or perhaps we cannot adequately partition the space of possible outcomes. Etc. For simplicity, I assume that such features are not present in *Scenario*. In particular, I assume that we can partition the space of relevant possibilities into a set of suitably fine-grained, exclusive and exhaustive propositions describing each relevant outcome.

In *Scenario*, the acts in question are either having one's own child or not having one's own child. The decision is the choice between whether to have a child or whether to remain childless. The outcomes of either act are its effects, which have dramatic emotional, mental and physical consequences. The dramatic effects follow the act of not having a child as much the act of having one: for example, not having a child means that you'll have very different experiences from ones you'd have had if you had a child, and has follow-on effects, such as the fact that you'd have significantly fewer financial costs for at least eighteen years following the date from when the omission can be said to "obtain."

The primary concern in *Scenario* is with the value of the outcome "for the agent," where this describes the value of the outcome brought about by the agent, centering on the outcome that involves the agent's perspective or point of view, that is, on the subjective value of what it is like to be the person who made the choice. In particular, the agent in *Scenario* is concerned with phenomenal outcomes that involve what it's like for her to have her own child. Since what it is like to be the agent includes what it is like to have her beliefs, desires, emotions, dispositions, and to perform subsequent acts, in *Scenario* the relevant outcomes include what it is like to have these additional effects and their attendant consequences as part of what it is like for her to have her child.

When choosing between *For* or *Against*, you compare the overall expected values of the outcomes of each act. Since we are concerned here with

[8] Not just anything goes. After all, the madman in the asylum can reason in accordance with his mad beliefs and come to the "right" decision given the beliefs he started with. But his decision to follow the voices in his head and attack his fellow inmates does not conform to what we would ordinarily describe as rational behavior. The madness of his starting point—his mad beliefs—and hence the mad values he assigns to the outcomes of his choices, violate our ordinary standard. As Weirich (2004, 21) points out, "an agent who maximizes utility may fall seriously short of other standards of rational action. For instance, an agent's utility assignment may be mistaken. Then, he may act irrationally even though he maximizes utility." We can allow that an agent may rationally make a merely approximately correct utility assignment and thus act approximately rationally. The point is that the madman's original utility assignments are not rationally acceptable.

ordinary decision-making, we use a normative model to guide our choice, allowing for approximation and estimation in place of perfect precision. To choose rationally, given our normative model, you determine the approximate value of each relevant outcome, you determine the approximate probability of each of these outcomes actually obtaining, and then use this information to estimate the expected value of each act. After estimating the expected value of each act, you choose the act that brings about the outcome with the highest estimated expected value.

In the case where you have a child, the relevant outcomes are phenomenal outcomes concerning what it is like for you to have your child, including what it is like to have the beliefs, desires, emotions and dispositions that result, directly and indirectly, from having your own child. Thus, the relevant values are determined by what it is like for you to have your child, including what it is like to have the beliefs, desires, emotions and dispositions that result, directly and indirectly, from having your own child. (I will sometimes call these values "phenomenal values": they are values of being in mental states with a phenomenal "what it's like" character.) In the case where you remain childless, the relevant outcomes are phenomenal outcomes involving what it is like for you to experience the effects of remaining childless, and thus the relevant values depend on what it is like for you to experience childlessness. In other words, the value of your act in *Scenario*, given the way the choice is made, depends largely on the phenomenal character of the mental states that result from it. This is neither surprising nor unusual from a commonsensical point of view.

Of course, having a child or not having a child will have value with respect to plenty of other things, such as the local demographic and the environment. However, the primary focus here is on an agent who is trying to decide, largely independently of these external or impersonal factors, whether she wants to have a child of her own. In this case, the value of what it is like for the agent plays the central role, if not the only role, in the decision to procreate. That said, the value of the choice is also affected if we assess the wider scope of the value of the act, since even in cases with a wider purview, the value of what it is like for the agent to have her own child must be evaluated in order to determine the overall expected value of her choice. For instance, you might choose to have a child because you desire to have some of your DNA transmitted to future generations. But the value of satisfying this desire must be weighed against other outcomes. If, say, the value of what it was like for you to have your own child was sufficiently positive or sufficiently negative, it could swamp the value of satisfying your desire to leave a genetic imprint.

3 What Experience Teaches

All of this might seem perfectly straightforward and unexceptionable. But there is a problem lurking beneath the surface. To see it, begin by reflecting

on an interesting fact about "what it's like" knowledge, such as knowledge of what it's like to see red. The interesting fact is that this sort of knowledge, that is, knowing what it's like, can (practically speaking) *only* be had via experience.

Frank Jackson developed a famous thought experiment to make this point. His example features black-and-white Mary, a brilliant neuroscientist, who is locked in a colorless cell from birth. Mary has never experienced color. Now, she knows all the facts in a complete physics (and other sciences), including all the causal and relational facts and functional roles consequent on knowing these facts, and including all the scientific facts about light, the human eye's response to light with wavelengths between 600 and 800 nanometers and any relevant neuroscience. Yet, when she has her first experience of red, she learns something new: she learns what it is like to see red.

> Mary is confined to a black-and-white room, is educated through black-and-white books and through lectures relayed on black-and-white television. In this way she learns everything there is to know about the physical nature of the world. . . . It seems, however, that Mary does not know all there is to know. For when she is let out of the black-and-white room or given a color television, she will learn what it is like to see something red. . . ." (Jackson 1986, 291)

As Jackson points out, when Mary leaves her cell for the first time, she has a radically new experience: she experiences redness for the first time, and from this experience, and this experience alone, she knows what it is like to see red.

Because of Mary's lack of experience, before she leaves her black-and-white cell, she lacks a certain kind of knowledge. Perhaps that knowledge is knowledge of a physical fact. Perhaps that knowledge involves a lack of a certain kind of ability or know-how. Perhaps it's knowing an old fact in a new way. Or perhaps, after leaving her room, she knows a new fact of some other sort.[9] None of that matters here.[10] The lesson for us is simply that, before she leaves her cell, black-and-white Mary is in an impoverished epistemic position. Until she actually has the experience of seeing red, she cannot know what it is like to see red.

An important feature of this example relies on the fact that, given Mary's exclusively black and white experiences, the experience of seeing red is unique and distinctive for her. Before she leaves her room, she cannot project forward to get a sense of what it will be like for her to see red, since she cannot project from what she knows about her other experiences to

[9] See Lewis (1990) for relevant discussion.

[10] In other words, we are not concerned here with the debate over physicalism that the example was originally designed for.

know what it is like to see color. As the example is described, then, before she leaves the room, her previous experience is not projectable in a way that will give her information about what it is like to see red. As a result, when she leaves her room and sees red for the first time, her experience is *epistemically transformative*.

Now let's restrict Mary's epistemic situation a little more than it was in Jackson's thought experiment. Before she leaves her room, because she doesn't know what it is like to see red, or indeed what it is like to see any sort of color at all, she also doesn't know what feelings and thoughts she'll experience as the result of seeing red.[11] And so she doesn't know whether it'll be her favorite color, or whether it'll be fun to see red, or whether it'll be joyous to see red, or frightening to see it, or whatever. And even if she could know, say, that she would find seeing red frightening, she wouldn't know how phenomenologically intense this experience would be.

For our purposes, Mary's impoverished epistemic situation means, first, that since Mary doesn't know how it'll phenomenally feel to see red before she sees it, she also doesn't know what emotions, beliefs, desires, and dispositions will be caused by what it's like for her to see red. Maybe she'll feel joy and elation. Or maybe she'll feel fear and despair. And so on. Second, because she doesn't know what emotions, beliefs, desires, and dispositions will be caused by her experience of seeing red, she doesn't know what it'll be like to have the set of emotions, beliefs, desires, and dispositions that are caused by her experience of seeing red, simply because she has no guide to which set she'll actually have. And third: she doesn't know what it'll be like to have any of the phenomenal-redness-involving emotions, beliefs, desires, and dispositions that will be caused by her experience of seeing red. Even if she could somehow know that she'll feel joy upon seeing red, she doesn't know what it will be like to feel-joy-while-seeing-redness until she has the experience of seeing red. And these are all ways of saying that, before she leaves her cell, she cannot know the value of what it'll be like for her to see red.

This means that, when Mary chooses to leave her black-and-white cell, thus choosing to undergo an epistemically transformative experience, she faces a deep subjective unpredictability about the future. She doesn't know, and she cannot know, the values of the relevant phenomenal outcomes of her choice.

[11] In Jackson's thought experiment, because Mary has all the scientific information we'd have at the end of scientific enquiry, Mary might know what brain states will be caused by seeing red, and thus might, at least arguably, know what beliefs and desires, etc. will be caused. This kind of epistemic access is unavailable to ordinary humans reflecting on what they should do, so we can dispense with this possibility.

4 The Transformative Experience of Having a Child

A person who is choosing whether to become a parent, before she has a child, is in an epistemic situation just like that of black-and-white Mary before she leaves her cell. Just like Mary, she is epistemically impoverished, because she does not know what it is like to have a child of her very own.

Why is she epistemically impoverished? At least in the normal case, one has a uniquely new experience when one has one's first child. Before someone becomes a parent, she has never experienced the unique state of seeing and touching her newborn child. She has never experienced the full compendium of the extremely intense series of beliefs, emotions, physical exhaustion and emotional intensity that attends the carrying, birth, presentation, and care of her very own child, and hence she does not know what it is like to have these experiences.

Moreover, since having one's own child is unlike any other human experience, before she has had the experience of seeing and touching her newborn child, not only does she not know what it is like to have her child, she cannot know.[12] Like the experience of seeing color for the first time, the experience of having a child is not projectable. All of this means that having a child is epistemically transformative.

Now, having a child is not just a radically new epistemic experience, it is, for many people, a life-changing experience. That is, the experience may be both epistemically transformative and *personally transformative*: it may change your personal phenomenology in deep and far-reaching ways. A personally transformative experience radically changes what it is like to be you, perhaps by replacing your core preferences with very different ones.[13] For most people, having a child is transformative in both ways: it is an epistemically transformative experience that is also personally transformative.

Why do parents experience such dramatic phenomenological changes? It is a normal reaction to the intense series of new experiences that one has when one has a child of one's own. This is most obvious when the parent in question is the mother. The intensity and uniqueness of the extended act of carrying the child, the physicality of giving birth, the recognition of the new fact of the existence of one's very own child, and the exertion involved in caring for a newborn results in a dramatic change in one's physical, emotional and mental states. The experiences are also very intense for involved fathers. It is common for fathers to date their changed phenomenal state from the moment they saw or held their newborn.

[12] Even having a perfect duplicate of yourself around to undergo it and then tell you about the experience probably wouldn't be enough for you to know what it is like—just like a perfect duplicate couldn't tell you enough for you to know what it was like to see color if you'd never seen color before.

[13] See Ullmann-Margalit (2006).

Perhaps the primary basis for the radical change in phenomenology in both parents is the simple fact that the content of the state of *seeing and touching your own newborn child* can carry with it an epistemically unique and personally transformative phenomenological character.[14] This may be the source of why this experience is both epistemically and personally transformative.

There are probably attendant biological reasons for the phenomenological change in parents: when producing, breastfeeding and caring for a child, mothers experience enormous hormonal and other biological changes, and new fathers also undergo significant hormonal changes. Fans of evolutionary biology will hold that there is a biological mandate for the physiological changes in the parents that underlie the felt attachment to one's offspring. In any case, whether the primary basis for one's new phenomenology is simply the experience of producing, seeing, and touching your newborn child, or whether it is being in some new biological state, or whether it is a more extended and complex series of experiences, the parent has an experience he or she has never had before—an experience with an epistemically unique phenomenal character, and moreover, one which can also be personally transformative.[15]

The combination of the epistemically and personally transformative experience of having one's own child brings with it profound changes in other epistemic states. In particular, because you cannot know what it is like to have your own child before you've had her, you also cannot know what emotions, beliefs, desires, and dispositions will be caused by what it's like to have her. Maybe you'll feel joy and elation when she is born. Or maybe you'll feel anger and despair (many parents experience postnatal depression). And so on. Moreover, you can't know what it'll be like to have the particular emotions, beliefs, desires, and dispositions that are caused by your experience of having your child. As a result, if you have a child, and if your experience is both epistemically and personally transformative, many of your epistemic states will change in subjectively unprojectable ways, and many of these changes will be profound changes.

5 Choosing the Ordinary Way Is Not Rational

Recall the normative model for ordinary decision-making given in §2. You, as a normatively rational agent, are supposed to deliberate between acts: you determine the relevant outcomes of each act, the approximate

[14] The phenomenological character of having a child for a blind or otherwise differently abled person will be different but just as unique.

[15] Even the parent who reacts with numb disbelief or shock upon the presentation of her child has an experience with a uniquely new phenomenal character, despite the fact that the experience does not have the phenomenal character it is "supposed" to have. Indeed, this shocked reaction could have its distinctive character in part *because* it does not have the joyous character the agent was expecting.

probability of these outcomes, the approximate value of these outcomes, and then estimate the overall expected value of each act. After estimating the expected value of each act, you choose the act that has the highest expected value.

The lurking problem I alluded to in §3 comes from the fact that the normative model requires one to determine values of outcomes. And, in fact, *any* standard decision-theoretic model requires one to determine values, at least approximate ones, of outcomes. The problem surfaces when we realize that, first, we want to make the decision based on the phenomenal outcome, that is, based on what we think it will be like to have a child. And second, that if our choice involves an outcome that is epistemically transformative, we cannot know the value of this outcome before we experience it. And if we cannot determine the value of the relevant outcome, we are in the same epistemic position as the agent who, because he doesn't know what the prize will be, cannot rationally determine the utility of winning the lottery (Weirich 2004, 65).

Recall Mary in her black-and-white cell. Imagine that she is trying to decide whether she wants to leave her cell for the first time. As we saw, Mary doesn't know what it will be like to see color. In addition to its being a certain way to see red, maybe it will be terrifying and overwhelming to see color after living in black and white for so long. Maybe the particular fear created by seeing redness will be mind-numbingly awful and paralyzing. Or maybe seeing red for the first time will be blissfully wonderful. She just doesn't know. As I noted above, this means Mary doesn't know what values to assign to the phenomenal states that are the outcomes of her choice to leave her cell. If she cannot rationally determine the values of the relevant outcomes, she cannot use normative decision theory to make a rational choice. (And if she assigns values to these phenomenal states anyway, she is making an unacceptable mistake, for if she cannot know their values, there are no rationally acceptable values she can assign.) Either the decision theoretic model does not apply, because there is no value known for the relevant outcome, or the value she assigns to the outcome is based on an unacceptable belief about what the value should be, and a decision based on an unacceptable belief is not rational.[16]

The very same problem arises in *Scenario*. Here, you are deciding whether to have a child based on the expected value of the act for you and your partner. You think about what it would be like to have a child, how

[16] If the outcome is assigned a value based on an unacceptable mistake, the case is parallel to other cases involving decisions based on mistaken or unacceptable beliefs. "[T]ake a case in which a decision to travel by train rests on an irrational belief that the plane will crash. The decision is irrational even if it follows by utility maximization from the agent's beliefs and desires" (Weirich 2004, 106). Mary might believe she can assign a value to her future phenomenal state of seeing red, but she is necessarily wrong—and so if she assigns it a value, she is making an unacceptable mistake. Her belief is not rational: the value cannot be known and so her belief about it cannot be based on evidence.

it will affect you and your partner, and how it will affect the other parts of your life, and you decide on the outcome with the best overall effects, where "best overall effects" is short for "effects that maximize expected value." Even if the contemplation is not as detailed or precise as the perfect rational agent could make it, an approximation of this approach embodies our ordinary way of trying to take a clear-headed, normatively rational approach to this extremely important decision.

The trouble comes from the fact that, because having one's first child is epistemically transformative, one cannot determine the value of what it's like to have one's own child before actually having her. This means that the subjective unpredictability attending the act of having one's first child makes the story about family planning into little more than pleasant fiction. Because you cannot know the value of the relevant outcome, there is no rationally acceptable value you can assign to it. The problem is not that a prospective parent can only grasp the approximate values of the outcomes of her act, for then, at least, she might have some hope of meeting our norms for ordinary decision-making. The problem is that she cannot determine the values with any degree of accuracy at all.

As a result, no matter which option in *Scenario* you choose, your decision is not even an approximation of a normatively rational act. It is impossible for you to follow the decision procedure in *Scenario* and choose *For* in a way that is consistent with the ordinary standard for rational decision-making. It is also impossible for you to follow the decision procedure in *Scenario* and choose *Against* in a way that is consistent with the ordinary standard for rational decision-making. Arguably, ordinary rationality does not even *permit* making either choice. Generalizing this, you cannot use our ordinary, phenomenal-based, normative decision procedure to rationally make one of the biggest decisions of your life. You cannot use this procedure to rationally choose to have a child, nor to rationally choose to remain childless.

Distinguishing between evidential and causal probability does not help: it is not rational to choose either option whether we understand your decision as one based on evidence or as one based on a judgment about the causal efficacy of the act. Finally, even a distinction between practical rationality and theoretical rationality will not help: your choice in *Scenario* is neither theoretically nor practically rational in the intended sense.[17]

It should be obvious that, in this discussion, I am abstracting from any moral considerations that might affect the choice to have or not to have children, and I am not taking a position on the nature of moral

[17] I have been focusing on our inability to assess states with phenomenal characters that directly involve what it's like to have a child. But there are familiar knock-on effects that are less direct. Once you have a child, will you care less about your career? Will you value your child's welfare over your own? Will you still love your cat just as much? Will you love your partner more? Will you love your partner less?—Who knows? It depends on what it's like for you to have your child.

deliberation—i.e., whether it is a form of rational deliberation, and whether its aim is to maximize value. I am starting from what I take to be our predominant cultural paradigm of how to consider the question of whether to have or not to have a child. According to that paradigm, we are to approach this decision as a personal matter where what is at stake is our own expected happiness and a sort of personal self-realization.[18]

And so we find a conflict between the ordinary way we are supposed to make the decision to have a child and the fact that having one's own child is an epistemically transformative experience. This conflict is interesting precisely because the decision to have a child may also be personally transformative. When a decision involves an outcome that is epistemically transformative for the decision-maker, she cannot rationally assign a value to the outcome until she has experienced the outcome. When that outcome may also be personally transformative for the decision-maker, the conflict matters—for she needs to make a big decision, a possibly self-transformative decision, and she cannot conform to ordinary or "folk" norms for rational decision-making when doing so.

6 Objections

My conclusion is controversial. The remainder of the paper will discuss some objections.

6.1 Subjective Ability

Perhaps you think that you can know what it's like to have a child, even though you've never had one, because you can read or listen to the testimony of what it was like for others. You are wrong.

> If you want to know what some new and different experience is like, you can learn it by going out and really *having* that experience. You can't learn it by being told about the experience, however thorough your lessons might be. . . . You may have tasted Vegemite, that famous Australian substance; and I never have. So you may know what it's like to taste Vegemite. I don't, and unless I taste Vegemite (what, and spoil a good example!) I never will. (Lewis 1990, 292)

The experience of having a child is exactly the sort of epistemically unique, epistemically new experience that Lewis is referring to.[19] Having one's first child and tasting Vegemite for the first time are both epistemically transformative (though tasting Vegemite is rarely personally transformative,

[18] I'm indebted to Tamar Schapiro for this point.

[19] I suppose it is one of the very few ways in which tasting Vegemite is, in fact, similar to having a child.

unless you are an Australian who has been away from home for a long time).

Being around other people's children isn't enough to learn about what it will be like in your own case. The resemblance simply isn't close enough in the relevant respects. Babysitting for other children, having nieces and nephews or much younger siblings—all of these can be wonderful (or horrible) experiences, but they are different in kind from having a child of your very own, perhaps roughly analogous to the way an original artwork has aesthetic value partly because of its origins. (Thus the various memes about "other people's children," including those about how one can dislike other people's children while loving one's own, about how adopting a child "isn't the same" as having one,[20] etc.) Experience with other peoples' children might teach you about what it is like to hold a baby, to change diapers or hold a bottle, but not what it is like to create, carry, give birth to and raise a child *of your very own*. This is obvious even if we discount the conceptual or indexical basis for the uniqueness of the experience, for as I pointed out above, there are purely biological causes that may be sufficient for its uniqueness: the hormonal reactions and other biological responses that stem from physically growing, carrying and giving birth to your own child (mutatis mutandis for fathers). One simply does not get this biological response from babysitting one's niece or changing one's nephew's dirty diaper.

You might think that having a description of what it's like to have a child will tell you what you need to know if it tells you about other experiences that closely resemble the new experience. But it doesn't, at least if you haven't experienced anything that closely resembles the experience, such as already having a child of your own. Lewis (1990, 265–266) points out that even if one can be told that the taste of Vegemite somewhat resembles Marmite, unless one has tasted Marmite, this misses the point. Without the relevant experience, no amount of information about resemblances will help.

The claim that having a child is epistemically transformative does not entail that, if you ascribe a value to what it will be like for you to have a child before you've actually had a child, the value you ascribe will be incorrect. You might get lucky. You might ascribe a value that, once you have the child, turns out to be reasonably close to the actual value. But this doesn't mean that it was rationally acceptable for you to ascribe this value before you could know what it was going to be. It was not rationally acceptable, for you could not know the value before you'd had the experience.[21]

[20] Please do not confuse this first claim with a second, different claim that adopting a child is somehow less valuable than having a child of one's own. I endorse the first claim and categorically reject the second.

[21] Moreover, the claim that having a child is epistemically transformative does not entail that it is also personally transformative: for most people, it is. For some people, it isn't.

Back to Mary in her colorless cell: Mary might guess that the experience of seeing color for the first time will be stressful and frightening. When she leaves her cell, she might indeed find her experiences of redness to be stressful and frightening. Or Mary might guess that the experience of seeing color for the first time will be fulfilling and satisfying. When she leaves her cell, she might indeed find her experiences of redness to be fulfilling and satisfying. But none of this entails that she was able to know what it would be like for her to experience redness before she actually experienced it, and so none of this entails that it was rationally acceptable for Mary to assign these values before she left her cell.

Can there really be anyone who would grant that the relatively mundane experience of tasting Vegemite for the first time is epistemically transformative, while *denying* that growing, carrying, giving birth to, and raising one's first child is epistemically transformative? If you grant that epistemically transformative experiences are possible at all, you should grant that having your first child is one of them.

6.2 Alternative Decision Procedures

The normative model captures the structure of an ordinary decision-making process. Many people, myself included, take the normative model (or close variations thereof) to provide the most natural framework for decision-making in this particular context, even if it gives us unsatisfactory results. However, it is well-known that decision-making under ignorance creates special problems for agents, and models for decision-making under ignorance have been developed for agents to use.[22] How does this fact affect my argument?

In a nutshell: it doesn't. Our option is to replace the simple version of the normative model with a different version, one which would apply under epistemically impoverished circumstances. This might seem like the obvious way to approach the problem. After all, the real world is messy, and as I discussed in §2, the difficulty of fitting the pristine, clear and precise models of decision theory with the murky viewpoints of actual agents is well-known. Can we accommodate decisions involving epistemically transformative experiences by using special models for decision-making under ignorance?

No. The same problem that arose for our simple normative model arises with these special models, for it is a condition of application for all such models that we are able to legitimately determine the values (or

But because it is epistemically transformative, you can't know whether you will find the experience personally transformative until you experience it, and so the problem for rational decision-making remains.

[22] See, for example, Levi (1986) and Weirich (2004). Joyce (1999) and Hansson (Unpublished manuscript) give excellent general discussions.

utilities), at least approximate ones, of the relevant outcomes of the act.[23] In the most common models for decision under ignorance, the models specify the values of the outcomes of the act, but—representing agent ignorance—no probabilities are determined. Just as with our original normative model, your choice to have your own child is based on your phenomenal preferences, so to use these decision theoretic models, you have to be able to determine the approximate values of the phenomenal outcomes, outcomes including *what it is like for you to have your own child*. But because you do not know what it is like to have your own child, you lack the relevant phenomenal knowledge you need in order to rationally determine these values.

For example, a simple model for decision-making under ignorance could use the "maximin" rule for making decisions. When "maximining" the agent decides conservatively, that is, makes a safe bet, with the objective of minimizing bad results. To use this decision procedure, we first determine the desirability and undesirability of each relevant outcome. Then we choose the act whose worst outcome has the highest desirability relative to the worst outcomes of all the acts under consideration, that is we, choose the act with the "least bad" outcome. A different, more optimistic model uses a version of the "maximax" rule: calculate the value of each relevant outcome, and then simply choose the outcome that has the highest value. That is, we "maximax" by choosing the act whose best outcome is the most desirable outcome. Either approach allows for rational decision-making under ignorance.

To apply these models, we determine the values of outcomes and then apply a decision rule. The appropriate decision rule depends on the context, which includes the agent's circumstances and dispositions. If, for example, you are choosing from a range of unfamiliar dishes at a new restaurant somewhere in the Midwest, you might wish to employ the maximin rule, selecting the simply prepared steak instead of the interesting, but unusually flavored, seafood dish. Here, outcomes include having a decent steak, having a delicious seafood dish, or having a disturbingly chewy, unpleasantly fishy evening meal. On the other hand, if the restaurant has enough Michelin stars, you might decide to throw caution to the winds and employ maximax reasoning to go for the Aguachile de Pulpo y Calamar after all.

But what if you are visiting Australia for the first time, and need to choose between having toast with orange marmalade and toast with Vegemite? If you've never had Vegemite, nor anything resembling it (such as Marmite), and you want to choose based on what it will be like for you to taste

[23] Weirich (2004) discusses a range of ways for agents to make normatively rational decisions under ignorance, including models where the standard for rationality is much more tolerant of ignorance. Such models permit cases that lack precise utility assignments. However, in the case of having a child, we are unable to rationally restrict the range of utilities and their probabilities in any reasonable way, preventing us from meeting even this more tolerant standard.

Vegemite, you are out of luck.[24] Neither maximin nor maximax will work for you. In the Midwestern restaurant, you chose between outcomes that resembled what you'd experienced in the past (a decent steak, good seafood, bad seafood), and so you were able to assign values to them. But in a case where you really don't know what it's like to taste the menu item, you can't use maximin, or maximax, or any other decision-under-ignorance rule to rationally make a decision based on what you think it will taste like. You just don't have enough information to deploy the model.[25] You might be able to rationally make your menu choice on another basis, say, where you regard the choice merely as a fun, low-stakes gamble, but a decision on that basis is not analogous to the phenomenally-based decision to have a child.

You might think, hang on, we can just parse the range of outcomes so that they are described as outcomes like "Vegemite tastes delicious," and "Vegemite tastes disgusting."[26] But simply adding terms like "delicious" or "disgusting" to the description of the outcome won't give you the information about values that you need. Intuitively speaking, you need to know more in order to assign them values. You need to know how phenomenally intense the state described by "Vegemite tastes delicious" and how phenomenally intense the state described by "Vegemite tastes disgusting" is, and you need experience in order to know this.[27]

We find ourselves with the very same problem in *Scenario*. No standard model of decision under ignorance is available to the prospective parent who chooses based on what she thinks it will be like to be a parent, for, just as in the Vegemite case, she cannot determine the values of the relevant outcomes. As a result, the models don't apply.

Now, of course, I am assuming various constraints here: it isn't *metaphysically impossible* to determine the values of the outcomes. It is simply epistemically impossible given very reasonable and appropriate real-world constraints. For example, if you had a perfect physical duplicate who underwent the experience of having a child and then told you how to assign values to the outcomes for your version of the experience, you could employ a decision-theoretic model. This sort of pretend scenario, and various other sci-fi alternatives we might be able to dream up, are obviously irrelevant in this context.

[24] Some people find Vegemite absolutely disgusting. Others think it is delicious.

[25] As Weirich points out: "It would be difficult, even for a perfect mind, to sensibly assign intrinsic utilities to states of affairs in the absence of relevant experience. For instance, it would be difficult to assign intrinsic utility to tasting pineapple in ignorance of its taste, or to assign intrinsic utilities to eating items on the menu in an Ethiopian restaurant, even given their full descriptions, in the absence of experience with Ethiopian cuisine" (2004, 65).

[26] I'm indebted to Elizabeth Harman for raising this objection.

[27] One way to put it is to say that you need to be able to grasp the phenomenal content of the proposition described by "Vegemite tastes disgusting," and you can't grasp this content until you've actually tasted Vegemite. Weirich puts the point this way: "the experience may be needed to entertain a proposition in the vivid way required for its intrinsic utility assessment" (2004, 66).

There is another issue here that should be raised: not only is the phenomenal outcome *what it's like to have your own child* a relevant outcome of your choice, it's an outcome whose value might *swamp* the other outcomes. In other words, even if other outcomes are relevant, the value of the phenomenal outcome, when it occurs, might be so positive or so negative that none of the values of the other relevant outcomes matter.[28]

Now, we need not take the fact that normative decision theoretic models don't work well for the case of having children as a criticism of decision theory, for sophisticated decision theorists often think of decision theory as a useful evaluative tool, not as a method one should use to determine, in practical circumstances, what sort of deliberation is rational.[29] The point being made here is that you cannot rationally decide to have a child based on what you think it will be like for you to have a child, and debates about how to make this important life choice should reflect this fact.

6.3 Eliminate the Subjectivity in the Decision Procedure

The source of the problem is the epistemically transformative nature of the experience of having one's child. One way to circumvent this problem is by dispensing with projectability, that is, ignoring your own personal preferences when you choose. You can change the decision procedure and choose to have a child based *solely* on the assumption that anyone who has a child is more likely to end up in a class of individuals who maximize their overall utility, ignoring your own personal beliefs, desires and other phenomenal projections about the future.

Let's consider this possibility. After choosing, you could end up in one of four different classes. The class of individuals for whom, after having a child, the overall value of having a child is higher than it would have been if they had remained childless, is *Lucky Parents*. The class of individuals for whom, after having a child, the overall value of having a child is lower than it would have been if they had remained childless, is *Unlucky Parents*. The class of individuals for whom, having decided to not have a child, the overall value of the choice to be childless is higher than it would have been if they had had a child, is *Lucky Child-frees*. Finally, the class I'll label *Unlucky Child-frees* is the class of individuals for whom, having decided to be childless, the overall value of the choice to not have a child is lower than it would have been if they had had a child.

[28] Of course, swamping can work in the other direction as well. There may be cases where the stakes are relatively low, and values of, say, certain nonphenomenal outcomes will clearly swamp the values, whatever they might be, of the relevant phenomenal outcomes. For example, if in the interest of promoting Australian tourism, foreigners receive a large financial reward for trying Vegemite for the first time, then if you are not Australian, you might rationally choose to try it on this basis. But in high stakes cases like that of having a child, one would have to make the case that such nonphenomenal outcomes exist. What is much more likely is that the value of what it is like to have the child will swamp the other outcomes.

[29] I'm indebted to Kenny Easwaran for this observation.

Now if Lucky Parents is much larger than Unlucky Parents, and Unlucky Child-frees is much larger than Lucky Child-frees, it might seem rational to choose to have a child, simply because you think, given the numbers, if you have a child you are far more likely to be in Lucky Parents than in Unlucky Parents, and you successfully avoid being classed in Unlucky Child-frees. And indeed, many people seem to assume something like the claim that Lucky Parents is much larger than Unlucky Parents. They also seem to assume that Unlucky Child-frees is much larger than Lucky Child-frees: they assume that people increase their happiness and well-being by having children and that childless people decrease their well-being (and as a result are unhappy or unfulfilled) because they do not have children of their own.

However, current empirical evidence suggests that this assumption is false. While the highs seem to be higher for parents, the lows seem to be lower, and many measures suggest that parents with children in the home have, on average, a lower level of overall life satisfaction.[30] Moreover, individuals who have never had children report similar levels of life satisfaction as individuals with grown children who have left home (Simon 2008; Evenson and Simon 2005). A recent analysis of survey data covering a wide range of the empirical results concerning parenthood indicates that *no* group of parents, including those whose children have grown and left home, where those groups are determined by standard sociological classifications such as income, marital status, gender, race, education, and mental health, report higher levels of overall emotional well-being than non-parents (Simon 2008; Evenson and Simon 2005).[31] Psychological results are more mixed: some studies report that parents have lower levels of subjective well-being (Kahneman et al. 2004), while others report that fathers enjoy a higher level of life satisfaction but mothers do not (Nelson et al. 2013).

At best, we have little or no evidence that Lucky Parents is much larger than Unlucky Parents, or that Unlucky Child-frees is much larger than Lucky Child-frees. At worst, the evidence suggests that choosing to have a child is likely to reduce your overall well-being. If you reject the empirical results (which are mixed and admittedly controversial), you find yourself without evidence to guide your decision. If you accept what the balance of evidence seems to show, then the rational choice requires you to act as though your own feelings don't matter. Independently of your own feelings on the issue, you must remain childless, for those who remain childless are

[30] McClanahan and Adams (1989) describe how a number of studies "suggest that parenthood has negative consequences for the psychological well-being of adults." The negative impact of children on happiness and life satisfaction has been widely discussed in sociology, psychology and economics. See, for example, Nomaguchi and Milkie (2003) and see Simon (2008) for a nice overall summary.

[31] The research does show that marital status, education and financial status influence the degree to which parenthood negatively impacts emotional well-being. See Kahneman et al. (2004) and Nelson et al. (2013).

more likely to end up in a class of individuals that have maximized their overall utility.

Thus far, it looks like, if you accept the new decision procedure, you should either hold off on deciding, due to lack of conclusive evidence, or you should ignore your own feelings and decide to remain childless.[32] This is an interesting result. But it is *strange*. First of all, it does not bode well for the future of the species. Second, deciding *solely* on the chance that you'll end up in a class of individuals who maximized their overall utility cuts hard against the way we ordinarily consider the decision.

Imagine Sally, who has always thought that having a child would bring her happiness, deciding not to have a child simply because she knows not having one will maximize her utility. For her to choose this way, ignoring her subjective preferences and relying solely on external reasons, seems bizarre. How could Sally's own phenomenal preferences not matter to her decision? Even Lisa, who, antecedently, does not want a child, and then decides not to have a child based solely on the evidence, is not choosing in an ordinary way. Her choice, if rational, has nothing to do with her phenomenal preferences to not have a child. Lisa does not have special insight into how she has always known that she'd be worse off as a parent: instead, she merely gets lucky. It just so happens that her phenomenal preferences support the same choice as the evidence does. Alternatively, imagine that the sizes of the classes were reversed so that Lucky Parents was much larger than Unlucky Parents, and Unlucky Child-frees was much larger than Lucky Child-frees. Now consider Anne, who has always thought that having a child would bring her misery, deciding to have a child simply because she knows it will maximize her utility. Again, the decision procedure seems bizarre from our ordinary perspective. Choosing rationally requires a very different way of thinking about the decision than we ordinarily think it does—to be rational, *we have to ignore our phenomenal preferences*.[33]

You might think that none of this applies to you. For you are a sophisticated thinker—you know, or at least you have educated, sophisticated beliefs—about which psychological characteristics really matter when you become a parent. You, unlike the unwashed masses, can judge for yourself whether you are more or less likely to end up in Lucky Parents if you have a child. I see no rational basis for a belief in such super-empirical abilities. There just isn't enough evidence available to support this sort of reasoning. Moreover, assessments of subjective well-being using the sorts of sophisticated psychological classifications that individuals would need to use to make an individually tailored, evidence-based decision are

[32] Depending on the context, this may amount to the same thing.

[33] A way of putting the problem is like this: decision-theoretic models are constructed as tools for evaluating decisions from the third-person perspective. But our ordinary way of making personal decisions relies on the first-person perspective. This can result in a fundamental conflict.

in their infancy (Kahneman and Kreuger 2006). Future empirical research might uncover the properties an individual needs to have in order to end up classed in Lucky Parents.[34] But we lack such evidence right now.[35]

As a result, the prospective parent finds herself in an interesting dilemma: ignore what she personally thinks about whether she wants to have a child and decide rationally, or take into account her own beliefs and projections about what it would be like and fail to decide rationally. Neither horn is attractive.

7 Conclusion

Contrary to popular opinion and common sense, contrary to what your parents might tell you, and contrary to the picturesque ideal romanticized by many a chick-lit novel, popular parenting guide, life coach website, and fashion magazine, you cannot rationally choose to have a child based on what you think it will be like to have a child. And, contrary to what those who are committed exclusively to their careers, or who dislike being around the children of other people, or who value their lazy weekends might believe, you cannot rationally choose to remain childless based on what you think it would have been like to have a child.

You can change the method of choosing so as to make it rational by making your choice based on something other than your phenomenal preferences. And indeed, in the past, non-subjective facts and circumstances played a much larger role in the causal process leading up to parenthood. Before contraceptive devices were widely available, you didn't choose to have a child based on what you thought it would be like. Often, you just ended up having a child. And to the extent you actively tried to choose to have children, often it was because you needed an heir, or needed more hands to work the farm, or whatever. But this is not the approach we ordinarily take now.[36] If you dispense with your phenomenal preferences,

[34] Another interesting possibility is that, just by having a child, one's preferences may change in a way that changes her assessment of the value of having a child. This is directly related to the way that the experience of having a child can be both epistemically and personally transformative. If the preferences had by the prospective parent before she has a child were unchanged by the experience, they might entail that the phenomenal outcome of having a child would have a negative value. But perhaps the very fact of having the child changes the prospective parent's preferences such that the phenomenal outcome of having a child turns out to have a positive value. (There is sociological evidence that this actually happens.) This possibility raises interesting questions about how one might employ higher-order decision-theoretic structure. (I'm indebted to Tania Lombrozo here.) Ullmann-Margalit (2006) discusses related issues.

[35] Frankly, I suspect that more evidence will only go so far, because the ability to determine which class one would be located in after the decision still requires a kind of self-knowledge that we can't have with epistemically transformative experiences. But that issue is beyond the scope of this discussion.

[36] See Zelizer (1985) for the classic account of how children have come to be regarded as emotionally priceless.

you reject a central tenet of the ordinary, twenty-first century way of thinking about the choice.

How could common sense have gotten things so wrong? I suspect that the popular conception of how to decide to have a child stems from a contemporary ideal of personal psychological development through choice. That is, a modern conception of self-realization involves the notion that one achieves a kind of maximal self-fulfillment through making reflective, rational choices about the sort of person one wants to be. (The rhetoric of the debate over abortion and medical advances in contraceptive technology have probably also contributed to the framing of the decision to have a child as a personal choice.) While the notions of personal fulfillment and self-realization through reflective choice might be apt for whether one chooses to grow one's own vegetables, what music one listens to or whether one does yoga, it is not apt for the choice to have a child. Some will conclude from my argument that we should base the decision to have a child on the values we assign to nonphenomenal outcomes or that moral considerations need to play a larger role. These conclusions might be warranted.

My view is not that it is right or wrong to have children, nor that you should not be happy with your choice, whatever choice you make. My view is simply that you need to be honest with yourself about the basis for this choice. For example, when surprising results surface about the negative satisfaction that many parents get from having children, telling yourself that you *knew* you would not be among that class of parents, and that's why you chose to have a child, is simply a rationalization—in the wrong sense—of your act. Likewise, telling yourself that you *knew* you wouldn't be happier as a parent, and that's why you chose not to have a child, is simply an act of self-deception. You can be happy that you have a child, or happy that you are childless, without wrapping that happiness in a cloak of false rationalization.

My argument also has consequences for those who want to be able to physically conceive, carry and give birth to a child, but are unable to do so. If you want to have a child because you think having a child will maximize the values of your personal phenomenological preferences, and as a result of your inability to have a child (and thus your inability to satisfy these preferences) you experience deep sadness, depression, or other negative emotions, my argument implies that your response is not rational. This is disturbing and some might find it offensive, but it is true. Such a response is not rational. That does not mean your response is wrong, or blameworthy, or subjectively unreasonable.

All of this raises larger issues, for the sort of subjective information that experience brings is central to many of our most important personal decisions.[37] Any epistemically transformative experience that changes the

[37] I discuss this in more detail in my *Transformative Experience*, where I consider ways in which my argument applies to choices that change our phenomenological capacities, such as getting cochlear implants, and life-course-decisions such as choosing a career.

self enough to generate a deep phenomenological transformation creates significant trouble for the hope that we could use our ordinary subjective perspective to make rational decisions about major life events.

L. A. Paul
E-mail : lapaul@unc.edu

References:

Caplan, Bryan. 2011. *Selfish Reasons to Have More Kids*. New York: Basic Books.
Evenson, Ranae and Robin Simon. 2005. "Clarifying the Relationship Between Parenthood and Depression." *Journal of Health and Social Behavior* 46 (4): 341–358. http://dx.doi.org/10.1177/002214650504600403.
Hansson, S. O. Unpublished manuscript. "Decision Theory: A Brief Introduction."
Jackson, Frank. 1986. "What Mary Didn't Know." *The Journal of Philosophy* 83 (5): 291–295. http://dx.doi.org/10.2307/2026143.
Joyce, James. 1999. *The Foundations of Causal Decision Theory*. Cambridge: Cambridge University Press.
Kahneman, Daniel and Alan Kreuger. 2006. "Developments in the Measurement of Subjective Well-Being." *Journal of Economic Perspectives* 20 (1): 3–24. http://dx.doi.org/10.1257/089533006776526030.
Kahneman, Daniel, Alan Krueger, David Schkade, Norbert Schwarz, and Arthur Stone. 2004. "A Survey Method for Characterizing Daily Life Experience: The Day Reconstruction Method." *Science* 306 (5702): 1776–1780. http://dx.doi.org/10.1126/science.1103572.
Levi, Isaac. 1986. *Hard Choices: Decision Making under Unresolved Conflict*. Cambridge: Cambridge University Press.
Lewis, David. 1990. "What Experience Teaches." In *Mind and Cognition: A Reader*, edited by William Lycan, 499–519. Oxford: Blackwell.
McClanahan, Sara and Julia Adams. 1989. "The Effects of Children on Adults' Psychological Wellbeing." *Social Forces* 68 (1): 124–146. http://dx.doi.org/10.1093/sf/68.1.124.
Nelson, S. Katherine, Kostadin Kushlev, Tammy English, Elizabeth W. Dunn, and Sonja Lyubomirsky. 2013. "In Defense of Parenthood: Children Are Associated With More Joy Than Misery." *Psychological Science* 24 (1): 3–10. http://dx.doi.org/10.1177/0956797612447798.
Nomaguchi, Kei and Melissa Milkie. 2003. "Costs and Rewards of Children: The Effects of Becoming a Parent on Adults' Lives." *Journal of Marriage and the Family* 65 (2): 356–374. http://dx.doi.org/10.1111/j.1741-3737.2003.00356.x.
Paul, L. A. 2014. *Transformative Experience*. Oxford: Oxford University Press.
Simon, Robin. 2008. "Life's Greatest Joy? The Negative Emotional Effects of Children on Adults." *Contexts* 7: 40–45. http://dx.doi.org/10.1525/ctx.2008.7.2.40.
Ullmann-Margalit, Edna. 2006. "Big Decisions: Opting, Converting, Drifting." *Royal Institute of Philosophy Supplement* 58: 157–172. http://dx.doi.org/10.1017/S1358246106058085.
Weirich, Paul. 2004. *Realistic Decision Theory: Rules for Nonideal Agents in Nonideal Circumstances*. Oxford: Oxford University Press.
Zelizer, Viviana. 1985. *Pricing the Priceless Child: The Changing Social Value of Children*. Princeton, NJ: Basic Books.

Acknowledgements I have been helped by discussion with many people. I owe special thanks to Lara Buchak, Kieran Healy and Matt Kotzen. Thanks are also due to Erik Angner, Peter Baumann, Rachael Briggs, Ruth Chang, Tyler Doggett, Kenny Easwaran, Jordan S. Ellenberg, Daniel Gilbert, Elizabeth Harman, Hud Hudson, Jonathan Jacobs, Tania Lombrozo, Ram Neta, Alvin Plantinga, Michael Rea, Geoffrey Sayre-McCord, Tamar Schapiro, Christina van Dyke, Paul Weirich and Mary Beth Willard.

Social Identities and Transformative Experience

Elizabeth Barnes

Abstract: In this paper, I argue that whether, how, and to what extent an experience is transformative is often highly contingent. I then further argue that sometimes social conditions are a major factor in whether a certain type of experience is often or typically transformative. Sometimes social conditions make it easy for a type of experience to be transformative, and sometimes they make it hard for a type of experience to be transformative. This, I claim, can sometimes be a matter of social justice: social conditions can make transformativeness too easy or too hard, in a way that harms people.

Much attention has been paid, in recent discussions, to the epistemic and decision-theoretic implications of transformative experiences. In this paper, I focus on a different and less explored aspect of transformative experiences: their normative significance.

L. A. Paul (2014; 2015), helpfully distinguishes between two ways in which an experience can be transformative. An experience is *epistemically transformative* if it gives you new 'what it's like' information that you didn't previously have access to (2014, 155). And experience is *personally transformative* if it significantly alters your priorities, your preferences, and your self-conception (2014, 156). I begin by making three very simple observations about both types of transformative experience. The first is that it is often contingent whether a particular type of experience is transformative in either sense. The second is that the transformativeness, in either sense, of a given experience is something that can come in degrees. The third is that how or in what way a particular type of experience is transformative can vary. I'm then going to use these three observations to argue that whether, how, and to what extent an experience is transformative can sometimes be a matter of social justice.

1 Transformativeness Is Contingent

When we got our dog, my husband—who had never had a dog, didn't want a dog, and only caved in to getting a dog after years of my pestering—fell instantly, deeply in love with her, and with dogs in general. The experience,

Res Philosophica, *Vol. 92, No. 2, April 2015, pp. 171–187*
http://dx.doi.org/10.11612/resphil.2015.92.2.3

by his own recounting, was both personally and phenomenologically transformative. He became aware of new and surprising information that was previously opaque to him—*what it's like* to share a deep emotional bond with a non-human animal. And his priorities and preferences changed in drastic ways. He rearranged his entire work schedule to make it maximally dog-friendly, he began giving money to dog charities, he no longer wanted to travel in his time off because he hated leaving the dog. Getting a dog had a profound, transformational effect on his life.

Needless to say, however, getting a dog doesn't always have this effect. Some people just aren't dog people. And some people, while they love their dog and really like dogs in general, nevertheless aren't emotionally transformed by the experience. Their dog is wonderful, but not life changing. Only the select few—the genuine *dog people* of world—seem to be convinced that dogs are the single greatest thing on earth, unrivaled in the love and companionship they bring. Whether getting a dog is transformative depends in part on whether you are such a person. And as my husband's experience shows, it can be difficult to predict whether you are such a person.

But whether an experience is transformative doesn't depend merely on what sort of person you are. It can also depend, at least in part, on your social environment and circumstances. In the novel *Great Expectations*, coming into wealth—and learning he has a substantial inheritance—is a both a personally and an epistemically transformative experience for Pip. He learns new information that was previously opaque to him—*what it's like* to have economic and social prospects, and to not be limited by his social status. He also shifts both his priorities and his self-conception. He decides he's going to be a respectable gentleman, and that his chief priority is to maintain his newly found social status. But Pip's coming into money has the potential to be so transformative for him in part because of his social class. Had he been slightly less poor or faced slightly fewer class barriers, coming into the same inheritance might well have altered him less radically. Transitioning to an upper-middle-class education and lifestyle is transformative for Pip at least in part because, due to the social constraints at the time, his poor, working class background had made him believe that such a transition was impossible.

So here is the first general observation I want to make about transformative experiences. Whether a particular token of a general type of experience is transformative is contingent. An experience which is actually transformative might have failed to be so, and vice versa. And, more specifically, whether a particular token experience is transformative can sometimes depend on features external to the experience itself. Whether a particular experience is transformative can depend, in part, both on contingent features of a person's psychological makeup and on contingent facts about their wider social situation.

2 Transformativeness Comes in Degrees

Epistemically transformative experiences are those in which a person gains new phenomenological information which they did not previously have access to. Personally transformative experiences are those in which a person's preferences, desires, and self-conception are altered. Both types of transformation are, arguably, things that admit of greater and lesser degrees.

In N. K. Jemisin's novel *The Hundred Thousand Kingdoms*, the main character Yeine becomes a god. This experience is, unsurprisingly, described in the novel as extremely epistemically transformative. Suddenly Yeine understands the connectedness of things, suddenly she can experience reality both temporally and atemporally, suddenly she has a nearly omnipresent sense of first-person perspective. Her sense of knowing *what it's like* to be a god is profoundly transformative, and was certainly something that was epistemically opaque to her when she was a human.

The first time I tried Irn Bru, I also gained new phenomenological information—I learned *what it's like* to taste Irn Bru. And it's fair to say that this information was previously opaque to me. No amount of previous soft drink tasting could have prepared me for the uniquely bizarre taste of Irn Bru. But there's a very wide phenomenological gulf between my first taste of Irn Bru and Yeine's becoming a god, even if we both gain some new 'what it's like' information.

Between these two cases lie many types of experiences we might think of as epistemically transformative. Holding your newborn child for the first time, experiences a type of synesthesia, having a migraine aura, falling in love, taking peyote—these will all give you new access to specific types of 'what it's like' information that you didn't previously have access to. So there's a sense in which all these experiences might be considered epistemically transformative.

But plausibly these experiences might give you both different amounts of new information and differently significant new information. Holding your newborn child might allow you to understand what it's like to love someone completely unconditionally, to feel fully responsible for another life, etc. A type of synesthesia might allow you to understand what it's like to associate numbers with colors. Both experiences may well give you new access to phenomenological information—it might be impossible to know what it's like to have either experience until you've actually had the experience. But holding your newborn child may well give you both more such information and more epistemically or personally significant such information.

In the Book of Acts, we are told the story of St. Paul's sudden, profound religious experience. The experience is clearly personally transformative for Paul. It completely rearranges his priorities, his desires, and even his

own self-conception—all he wants, after the experience, is to evangelize, and he's willing to put his own life at risk to do so.

My introduction to philosophy also had an effect on my priorities, my beliefs, and even my self-conception. I became very excited about philosophy, I began to apply philosophical methodology to other parts of my life, I began to consider the prospect of further study and career opportunities in philosophy, and I even began to think that maybe, one day, I could be a philosopher. There's certainly a sense in which being introduced to philosophy had a striking effect on my beliefs, my desires, and perhaps even my self-conception. But I very much doubt that my introduction to philosophy was transformative to the extent that Paul's vision on the road to Damascus is described as being transformative.

In between my first experience of philosophy and Paul's transformative religious experience we can find many of the kinds of things we might typically think of as personally transformative experiences. Coming close to death or being diagnosed with a serious illness, falling in love, getting divorced, becoming involved in a social justice movement, caring for an aging parent—these can all be the kind of thing that might rearrange one's priorities, desires, and sense of self. But they plausibly don't all always do so to exactly the same extent, or with exactly the same degree of personal significance.

With all this in mind, I contend that transformativeness—in either sense—isn't an on/off status of experiences. It's not the case, that is, that either an experience is transformative or it isn't. Transformativeness is something that can come in degrees. An experience e_1 can be more transformative than an experience e_2, even though they are both transformative. Whether there is a threshold for how much an experience must change you in order to count as personally transformative or how much 'what it's like' information an experience must give you in order to count as epistemically transformative isn't a question I'm going to address here. All I want is the simple claim that transformativeness comes in degrees.[1]

3 The Character of Transformativeness Is Variable

So far I have argued that there is variation in both whether and to what extent a particular type of experience is transformative. Transformativeness is contingent, and it comes in degrees. I'm now going to claim, somewhat more nebulously, that how or in what way a type of experience is transformative is also something that is contingent, and which can and does vary.

[1] This observation brings up an interesting puzzle—which I will simply mention in passing—for Paul's account of the connection between rationality and transformativeness. Paul argues that we cannot rationally decide to undergo (or fail to undergo) an experience which is transformative. But if transformativeness comes in degrees, the simple question arises: *how much* transformativeness is required to preclude rational decision making?

Perhaps when Anna holds her newborn baby for the first time, she undergoes an epistemically transformative experience—she learns what it's like to hold her newborn baby. And perhaps Bob also undergoes a transformative experience when *he* holds *his* newborn baby. But there isn't much reason to think that the information they now have access to, via their transformation, is the same—even if it shares some commonalities. That is, there isn't much reason to think that what it's like for Anna to hold Anna's baby is the same thing as what it's like for Bob to hold Bob's baby. Indeed, it would pretty implausible if the phenomenal content of these experiences were the same, given all the different experiences that will have led up to them, and all the differences in the two people who are the subjects of the experience. Perhaps holding your new baby is a type of experience that is generally epistemically transformative. That doesn't mean it's always transformative in the same way. It might generally lead to new phenomenological information—but to different new phenomenological information for different people.

Similarly, suppose that near-death experiences are often personally transformative. Even given this commonality, such experiences will likely be transformative in strikingly different ways for different people. Suppose that Ciara and Dani both survive sudden, near-fatal car accidents. Ciara decides, in the wake of this experience, that you only live once, so you have to live to fullest. She quits her city job to pursue her dream of becoming a white water rafting guide. She starts working on her 'bucket list', learns to parachute and bungee jump, and generally begins to pursue high-octane adventure. Dani, in contrast, becomes strikingly more risk averse. She makes a will and begins to carefully invest her savings. She starts to exercise, eat healthily, get plenty of sleep, and generally take better care of herself. She spends more time with family and friends.

Both Ciara and Dani's experiences are personally transformative, but the way in which they were personally transformative is very different. They each re-evaluate their goals, priorities, and preferences—and perhaps even their self-conception—but they do so in very different ways, and to very different results. *That* an experience is personally transformative doesn't tell you *how* it is personally transformative. The same type of experience can be equally transformative for two different people, but transform those people in two very different ways.

4 Hard and Easy Transformative Experience

It's tempting to think of transformativeness as an inherent aspect of experience. Some experiences are just *special*. But as discussed in section 1, this isn't quite right—whether an experience is transformative can depend on factors external to that experience. Whether an experience is transformative can be partly determined by independent facts about the person having the experience, and partly determined by facts about the wider social context

in which the experience is had. It's this latter set of factors I now want to focus on.

It's the wider social context of *Great Expectations*—and Pip's position in it—that make his inheritance transformative. No doubt aspects of Pip's personality play a role as well. But the socio-economic structures of Victorian England facilitate the kind of transformation Pip experiences—they make it *easy* for coming into wealth to be (very) transformative. In a society where there was less socio-economic stratification, or less social emphasis placed on class, it would be less easy for Pip's experience of inheritance to be transformative, or transformative to the same degree.

Similarly, let's follow Paul (2014) and assume that becoming a parent is often a very transformative experience. Conditions and expectations surrounding parenthood for wealthy, educated people in modern, Western societies no doubt facilitate the transformativeness of the experience of parenthood. Parenthood is often the result of careful deliberation, it is highly anticipated (and typically delayed well beyond the beginning of reproductive age), and it is upheld within our society as something that adds special meaning or significance to life. With all these conditions in place, it's not surprising that parenthood might often be experienced as transformative. But this isn't obviously a feature of parenthood simpliciter—parenthood devoid of the complex socio-economic circumstances in which it occurs. Whether a 17-year-old living in a multi-generational agrarian community in the 1800s would, for example, experience parenthood as transformative in the same way, or to the same degree, seems doubtful.

But just as social conditions can make it easy for an experience to be transformative, they can also make it hard. In a society with very little emphasis on class and a high degree of social mobility, it would be hard for an experience of sudden inheritance like Pip's to be transformative, or transformative to the same degree. It wouldn't be impossible—there might still be people who care a very great deal about wealth and social standing, even if that isn't the social norm—but transformativeness of such an experience would be unusual or atypical.

With all this in mind, I want to make the following general claims. A set of social conditions, *S*, make it *easy* for a type of experience, *E*, to be transformative just in case: (i) in nearby worlds in which *S* obtains, *E*-type experiences are often or typically transformative; (ii) in nearby worlds in which *S* does not obtain, *E*-type experiences are not often or typically transformative.[2] Conversely, a set of social conditions, *S*, make it *hard*

[2] This account of will, of course, face the standard types of problems encountered by counterfactual definitions. It will, for example, give the wrong results if the nearby worlds at which social conditions *S* don't obtain are such that social conditions *S** obtain, and *S** also make it easy for *E*-type experiences to be transformative. I'm giving these counterfactuals in order to give a basic gloss on how I'm understanding what it is for transformative experience to be made easy (or hard). I don't want to read too much into this as a counterfactual *analysis*, and it will no doubt be subject to funny counterexamples.

for a type of experience, *E*, to be transformative just in case: (i) in nearby worlds in which *S* obtains, *E*-type experiences are not often transformative or are atypically transformative; (ii) in nearby worlds in which *S* does not obtain, *E*-type experiences are more often or not atypically transformative.

Some experiences might be transformative regardless of the social circumstances in which they occur. Gaining a new sense modality, for example, might be epistemically transformative no matter the social context. And some experiences might depend primarily on personal, rather than social circumstances. Whether a particular type of experience is transformative might be primarily a function of whether the experiencer is a dog person, or has a religious cast of mind, or etc. And plausibly many experiences we tend to think of as transformative depend on a combination of both personal and social factors—whether you're a dog person in a pet-owning society, whether you're a religiously-minded person in a somewhat religious society, and so on.

When I say that a particular set of social conditions make it easy for a type of experience to be transformative, I don't simply mean that those social conditions facilitate the transformativeness of that type of experience together with some quirk or personality or character. Our social norms about pet ownership no doubt facilitate the transformativeness of dog ownership *for dog people*. But dog people are a quixotic bunch, and they certainly aren't the majority. When I say that a particular set of social conditions make it easy for a type of experience to be transformative, the thought is that most people—regardless of quirks of personality—who undergo such an experience given those conditions will find it transformative. It is typical or usual, in those conditions, for that experience to be transformative.

That needn't mean that the experience is itself common or typical. Perhaps the experience of becoming a sovereign ruler in the social context of absolute monarchy is typically transformative. The experience itself is a rare one. But most people, regardless of contingent facts about their personality, would find such an experience transformative. The social conditions of absolute monarchy can make it easy for becoming king or queen to be transformative without that experience being commonplace.

Social conditions making it hard for a type of experience to be transformative is not simply the converse of their making it easy. Easy and hard aren't exhaustive options, though they are exclusive. Modern norms about pet ownership might not make it easy for getting a dog to be transformative, but neither do they make it hard. For particular social conditions to make it hard for a type of experience to be transformative, it needs to be the case both that the experience isn't often transformative or is atypically transformative given those conditions, and that it would be transformative more often, or not atypically, transformative in the absence of those conditions.

Note that this is weaker than the requirement that in the absence of those conditions such experiences would often or typically be transformative.[3] Social conditions in which dogs are commercially reared as food and eaten as part of a standard diet would plausibly make it the case that getting a dog is very rarely a transformative experience. In the absence of those conditions, it still wouldn't be common for getting a dog to be transformative (since it still wouldn't be common to be a dog person). But it would be substantially more common. The presence of dog-eating social conditions can make it hard for getting a dog to be transformative, even though the absence of such conditions isn't sufficient to make it easy for getting a dog to be transformative.

With this basic understanding of hard and easy in place, we can then further complicate them by combining them with both degree and character of experience. We can say, for example, that set of social conditions, *S*, make it easy for a type of experience, *E*, to be transformative *in way W* just in case: (i) in nearby worlds in which *S* obtains, *E*-type experiences are often or typically transformative in way *W*; (ii) in nearby worlds in which *S* does not obtain, *E*-type experiences are not often or typically transformative in way *W*. Similarly, we can say that a set of social conditions, *S*, makes it easy for a type of experience, *E*, to be transformative *to degree n*[4] just in case: (i) in nearby worlds in which *S* obtains, *E*-type experiences are often or typically transformative to degree *n*; (ii) in nearby worlds in which *S* does not obtain, *E*-type experiences are not often or typically transformative to degree *n*.

So, for example, current social conditions for affluent, educated people might make it easy for having a child to be very transformative, or transformative in specific ways (involving a sense of added meaning to your life, perhaps). In different social conditions, having a child might tend to be somewhat less transformative, or might tend to be transformative in different ways. Similarly, in our current social conditions, if someone falls in love with a person of the same gender, this experience can be transformative in the familiar ways in which falling in love can be transformative. But in different social conditions, a person's falling in love with someone of the same gender might be transformative in very different ways—it might convince them they are particularly sinful, for example, or change their life to one of secrecy and isolation. How, and to what extent, an experience is transformative is shaped by social factors.

In what follows, I'm going to argue that whether and how social conditions make it easy or hard for a type of experience to be transformative can sometimes be a matter of social justice.

---

[3] I'm assuming here that 'atypical' is stronger than 'not typical.'

[4] This is a convenient fiction—I don't want to suggest that the degree to which an experience is transformative is (always) precisely quantifiable in this way. Talk of 'transformative to degree *n*' is just to highlight that experiences can vary in how transformative they are.

5 When Transformation Is Too Easy

Sometimes, social conditions make it easy for a type of experience to be transformative—or for a type of experience to be transformative in a particular sort of way. And sometimes, it shouldn't be easy for a type of experience to be transformative, or shouldn't be easy for an experience to be transformative in that particular way. One way in which social conditions can be harmful is by making certain kind of transformations easy.

Consider, for example, transformations that are made easy because of gender stereotypes and entrenched gender roles. In *Middlemarch*, Dorothea Brooke's marriage to Mr. Casaubon is described as a personally transformative experience. Dorothea's wishes, her values, and her priorities are all reshaped—they are completely reordered—in order to comply with Mr. Casaubon's. Upon marrying, she believes that her primary purpose (perhaps even her sole purpose) is to be of assistance to her husband. The transition is not an easy one for Dorothea, by any means. But she undergoes it willingly, believing it to be her calling as Mr. Casaubon's wife:

> By a sad contradiction, Dorothea's ideas and resolves seemed like melting ice floating and lost in the warm flood of which they had been but another form. She was humiliated to find herself a mere victim of feeling, as if she could know nothing except through that medium: all her strength was scattered in fits of agitation, of struggle, of despondency, and then again in visions of more complete renunciation, transforming all hard conditions into duty. (Eliot 2007 [1871], 208)

Eliot describes Mr. Casaubon as receiving, without question, this humbling transformation from Dorothea. She writes of Mr. Causabon that:

> It had occurred to him that he must not any longer defer his attention of matrimony, and he had reflected that in taking a wife, a man of good position should expect and carefully choose a blooming young lady—the younger the better, because more educable and submissive—of a rank equal to his own, of religious principles, virtuous disposition, and good understanding. On such a young lady he would make handsome settlements, and he would neglect no arrangement for her happiness: in return, he should receive family pleasures and leave behind him that copy of himself which seemed so urgently required of a man. . . . And when he had seen Dorothea he believed that he had found even more than he demanded: she might really be such a helpmate to him as would enable him to dispense with a hired secretary. . . . Providence, in its kindness,

> had supplied him with the wife he needed. A wife, a modest young lady, with the purely appreciative, unambitious abilities of her sex, is sure to think her husband's mind powerful. Whether providence had taken equal care of Miss Brooke in presenting her with Mr Casaubon is an idea which could hardly occur to him. (2007 [1871], 293)

It's plausible that becoming *a wife* was often, in the context of such gender norms and stereotypes, a transformative experience. Personally transformative experiences are those which reshape your priorities, your preferences, and your self-conception or sense of identity. And that's exactly what getting married was supposed to do for women (though not for men, of course). Massive shifts in priorities and self-conception were the expectation for women—and women only—upon marriage.[5]

So here is one striking characteristic of the hierarchical gender norms described in *Middlemarch*: they suggest that becoming a wife *ought* to be a transformative experience. When someone becomes a wife, she should rearrange her priorities, her desires, and her projects to cohere with and conform to her husband's. Being *her husband's wife* should be her primary role, and her primary self-conception.

Dorothea is intelligent, brave, thoughtful, and ambitious. In different circumstances, she would've pursued her own career and her own ideas. But within the restrictive gender hierarchy of 1830s England, her best sense of how to pursue her love of learning is by devoting herself—completely—as the wife of a scholarly man. In order to do this, she must undergo a deeply transformational experience. She must learn to prioritize his feelings over her feelings and she must begin to attempt to view things as he does.

The gendered norms of 1830s England make it *easy* for Dorothea's marriage to be a transformative experience. And more specifically, they make it easy for her marriage to be transformative in specific ways—ways which subsume her wishes, her preferences, and her sense of self to that of her husband. Arguably, that they make it so easy is a bad thing—it is part of the structural badness of such norms that they make transformative experiences like Dorothea's easy. The kind of self-abnegation involved in Dorothea's transformative experience is harmful to her. It changes her in a way that leaves her feeling lonely, unfulfilled, and frustrated. And it's not just harmful to Dorothea. Eliot suggests that Dorothea is a better, clearer thinker than Casaubon. If she had been able to pursue her own projects and ideas, she would likely have produced more valuable work than he ever could. But the transformative experience she undergoes, upon her marriage, leaves her with a very poor opinion of her own taste and judgement, and teaches her to value Casaubon's opinion above her own.

Abstracting away from the particular case of Dorothea and 1830s gender norms, the more general point I'd like to make is this. Sometimes the fact

[5] See especially Yalom 2002.

that social conditions make it easy for a particular type of experience to be transformative is harmful. It can be *too easy* for an experience to be transformative, and it can likewise be too easy for an experience to be transformative in specific ways. There can be cases in which an experience's being transformative—or being transformative in a particular way—constitutes a harm, and insofar as social conditions make that kind of transformation easy, they perpetuate that harm.

6 When Transformation Is Too Hard

But just as social conditions can facilitate transformative experience in ways that are harmful, they can also prevent or impede transformative experience in ways that are harmful.[6] Consider the social conditions and norms surrounding disability. There is perhaps a minimal and not very interesting sense in which becoming disabled is always at least an epistemically transformative. You learn what it is like to have a certain kind of physical condition—knowledge you did not previously have access to. But becoming or being disabled can also be personally transformative—and whether, how, and to what extent it is so is a more complex issue.[7]

Simi Linton is a disabled scholar and activist whose experience of becoming disabled as an adult was personally transformative. It changed the way she thought about herself, her priorities, and her relationship to others. Moreover, she views this change as a positive one—her sense of self has been importantly shaped by being disabled, and being a disabled person is a valued part of her identity. In her book *Claiming Disability: Knowledge and Identity* she describes the importance of disability as a type of self-identity, and as a way of building a disability community. Disability identity, she argues, is in part:

> an account of the world negotiated from the vantage point of the atypical. . . . The cultural stuff of the community is the creative response to atypical experience, the adaptive maneuvers through a world configured for nondisabled people. The material that binds us is the art of finding one another, of identifying and naming disability in a world reluctant to discuss it. . . . My experience as a disabled [person] and my alliance with the community are a source of identity, motivation, and information. (Linton 1998, 5)

But *becoming* disabled isn't the only way in which disability can provide transformative experience. Sometimes a transformative experience occurs, not in virtue of a newly acquired disability, but in virtue of a newly acquired way of viewing a disability. For example, disability activist Steven E. Brown,

[6] For further discussion of the intersection of identity, transformative experience, and oppression see McKinnon 2015.

[7] For further discussion of transformative experience and disability see Howard 2015.

in his essay 'I was Born in a Hospital Bed (When I was 31 Years Old)' (2003), recounts his experience of a sudden shift in the way he viewed his disability. Brown was born with a painful degenerative condition, and had spent most of his life up to this point feeling as though this was his own 'cross to bear' or his own personal tragedy. But then, in the wake of having been denied work because of his disability and attending a disability rights event to ascertain whether he might be able to combat this discrimination, something changed. He writes:

> I was born in a hospital bed when I was thirty-one years old. . .
>
> As I lay on that bed I benefitted from the luxury of unhurried contemplation. I focused on my body—which had steamrolled me into this predicament. I was tired of that body. . .
>
> As I began this mental meandering I could only think about the past twenty-five years in a cloud of unbridled agony. But, then, in the time it took to inhale the scent wafting from nearby flowers, I underwent one of those sudden transformations that people often label revelations. . .
>
> I was thirty-one years old and my body had borne more scars than most people feel in a lifetime twice as long. I thought about those heroes of my youth—stars of various sports—and the scores of times commentators bemoaned the aches and pains athletes lived and played through. I realized that my body had taken an athlete's abuse over and over again and rebounded every time. . .
>
> I began to view my body differently. For a long time I had been consumed with bitterness and anger. . . . The hospital inspired rendering of this litany of breaks and bruises awakened me to another truth. My body had weathered a storm of abuse—some of which was inherent in my being and some of which I had heaped upon it in my rebellion against its limitations. Laying in that hospital bed I also saw that the thunder and lightning had alternated with periods of sunshine and calm. I decided right then and there to be nice to my body. In essence I made a life-affirming decision. I recognized myself for who I was, with my disability and its limitations—and with my disability and its affirmations. A funny thing happened when I chose to like my body. I also began to like myself a lot more. And to embrace life itself. (Brown 2003, 61–63)

Brown's experience was personally transformative—so much so, even, that he describes it as the day he was born. From this point on, he became a disability activist and immersed himself in the disability rights community.

But the transformation wasn't due to acquiring a disability, it was due to changing the way in which he viewed his disability.

Similarly, disability rights activist Tammy S. Thompson describes a transformational experience that occurred in virtue of a shift in disability-related perspective, rather than disability status:

> I've spent many years on a mission to cancel out my disability by frantically stacking up achievements, hoping that someday I would find that final, magic accomplishment which would absolve me of the sin of being disabled. . . . No matter what I did, I collided with that hard fact. I couldn't seem to accept it and carry on without shame. Then one day, riding the bus, I met a fellow with a disability who was proud. He was comfortable with himself and his disability. Disability pride—wasn't that an oxymoron? I had to find out, so I got involved in the independent living movement he told me about.
>
> Participating in the Center for Disability Leadership program brought me up to speed and launched me into the disability rights movement. My life and my thinking were liberated. I got connected with powerful, wonderful people who were also disabled. These disability warriors taught me a new way to live that frees me from my past. (Thompson 1997)

Like Brown, Thompson's transformative experience arises via a shift in her perspective about her own disability. And like both Brown and Linton, the key aspect of this transformational experience—a sense of positive self-identity as a disabled person—arises due to interaction with the disability rights community.

For each of Linton, Brown, and Thompson, it seems that whether, how, and to what extent their experiences of disability were transformative is highly contingent. They each attribute their formation of a strong, positive disability identity to their interactions with the disability community and the disability rights movement. Nor do they appear to be alone in this. Research suggests that a strong, positive sense of self-identity as a disabled person is common within the disability rights community.[8] But, of course, whether one has access to the affirming, encouraging, often life-altering (as it was for Linton, Brown, and Thompson) support of the disability community is a highly contingent thing—many, perhaps most, disabled people in contemporary society do not.

The type of personally transformative experiences reported by Linton, Brown, and Thompson are those in which disability positively reshapes their identity and self-conception. They come to think of themselves as

[8] See Hahn and Belt 2004. Hahn and Belt's study further suggests that positive disability self-identity is strongly correlated with negative attitudes toward 'cures' for disability.

disabled people (not just as people who happen to have disabilities), in a way that's personally valuable to them. And this kind of positive sense of disability self-identity isn't just a theoretical curiosity. Whether disability is transformational in this way is something that has the potential to beneficially impact disabled peoples' lives. For example, current research suggests that, for disabled people, non-acceptance of disability is correlated with depression (and predicts future depression),[9] that positive disability identity predicts self-esteem,[10] and that positive disability identity predicts satisfaction with life.[11]

Forming a positive sense of self-identity as a disabled person is one way in which being or becoming disabled can be personally transformative. But, I suggest, it is *hard* for being or becoming disabled to be transformative is this way, given the current social norms and stereotypes surrounding disability. As Linton (1998) points out, many of the positive transformative aspects of disability have to do with experiencing an affirming and accepting sense of disability identity, and the sense of community with other disabled people that this can bring. And yet, she argues, dominant stereotypes about disability suggest precisely the opposite. Disability is not, as standardly understood, something that gives you access to—or something you experience with—a community. Disability is individual tragedy or private burden. Similarly, we tend to think of the potential good effects of disability only in terms of *overcoming* disability—the perseverance, the patience, the fortitude that being disabled can teach. The thought that disability could actually be a positive aspect of someone's self-conception—something they value about themselves, for its own sake—is an idea that's incredibly foreign to most people.

Nowhere is this more telling than in the fact that "I've never really considered you disabled" or "I don't think of you as disabled" are things that non-disabled people say, to disabled people, as *compliments*. When a non-disabled person says "I've never really considered you disabled," they don't typically mean that they don't consider you to have a condition that is generally thought of as a disability. They aren't expressing surprise that you use an accessible parking spot or bathroom stall. What they're saying is that they've never really considered you *less than* or *deficient* in some important way. (Cheer up, disabled person—this normal person thinks of you as normal! You should be flattered.)

It's hardly surprising, in the context of such flagrant stereotypes about disability, that transformative experiences involving a positive sense of disability self-identity stand out as atypical or rare. They are certainly

[9] See Townend et al. 2010

[10] See Nario-Redmond et al. 2013

[11] See Bogart 2014. Bogart interprets her findings as follows: "Results suggest that rather than attempting to 'normalize' individuals with disabilities, health care professionals should foster their disability self-concept. Possible ways to improve disability self-concept are discussed, such as involvement in the disability community and disability pride" (9).

not the norm or the expectation—and seem very often to be mediated by interaction with the disability rights community, an interaction which is itself not the norm or the expectation. We expect disabled people to try to 'overcome' their disabilities and to hope for 'a cure.' Neither of these expectations cohere well with a positive sense of disability as an important, valuable part of disabled peoples' self-identity.

And so, I contend, current norms and stereotypes about disability make the kind of personally transformative experiences described by Linton, Brown, and Thompson *hard*. These experiences are atypical, but I suggest that they are atypical—at least in part—because of the dominant norms and stereotypes about disability. Furthermore, I suggest that it is harmful to disabled people if our current norms and stereotypes about disability make these transformative experiences hard. These experiences are a valuable aspect of being disabled, and they have the potential to have significant positive impact on the wellbeing of disabled people. If they are hard to come by, that's harmful.

7 Transformative Experience and Social Identities

I have argued that social conditions can make it hard for certain kinds of experiences to be transformative (or to be transformative in certain ways or to certain extents), and that social conditions can likewise make it easy for certain kinds of experiences to be transformative. And I've further argued that sometimes whether it is hard or easy for a certain kind of experience to be transformative can be a matter of social justice. Sometimes the fact that social conditions make it hard (or easy) for an experience to be transformative constitutes can constitute a harm (or a benefit).

I want to summarize by making a claim about the relationship between personally transformative experience and identity. Experiences are personally transformative when they re-shape your self-conception or sense of self-identity. But self-conception and sense of self-identity aren't developed in cultural isolation. Social norms and structures make certain ways of interpreting or thinking about ourselves readily available. Faithful husband, loving mother, brilliant genius, tragic overcommer, self-sacrificing caregiver, breadwinner, muse—these are all ways we can think about ourselves and our own experiences. *Which* ways of thinking about ourselves are most salient or readily available will be, at least in part, a function of the social norms and structures in which we find ourselves.

If a personally transformative experience is one that re-shapes our sense of self, then personally transformative experiences can be radically affected by which ways of re-shaping our sense of self are salient to us. 'Submissive and dutiful wife' was, in 1830s England, an easy way for Dorothea Brooke to understand herself and her own experience. 'Free-thinking scholar' was not. 'Brave inspiration' is an easy way for disabled people to understand

their own experiences now. 'Thriving person in an unconventional body' is not.

What ways of understanding yourself and your own sense of identity your social situation makes salient needn't always be a normatively weighty matter. Plausibly, sometimes a type of identity might be readily available—and a corresponding transformational experience might be made easy—for reasons of (not very interesting) cultural accident. Perhaps, for example, being a Mod or a Rocker in 1960s England really was an important part of some peoples' sense of identity, and perhaps some people really did undergo personally transformative experiences when they found their scene. Nevertheless, whether one can easily identify as a Mod or a Rocker doesn't seem to be a particularly pressing matter of social justice. Indeed, it seems large a matter of accident—to be a Mod or a Rocker you just have to be in the right place at the right time.

In other cases though, the availability of specific identities is more plausibly something that matters. The fact that it was so easy for women to re-shape their self conception to cohere with the image of a dutiful, submissive wife was something that was bad for women. Part of achieving justice for women is making identities like this less readily available, and making other identities more readily available.

The relevance of transformative experiences to epistemology and decision theory is something that's received a lot of attention recently. But if I'm right, transformative experiences aren't of interest only for their epistemological or decision-theoretic import. Whether, how, and to what extent a type of experience is transformative is something that can sometimes matter morally as well.

Elizabeth Barnes
E-mail : e.j.barnes@virginia.edu

References:

Bogart, Kathleen. 2014. "The role of disability self-concept in adaptation to congenital or acquired disability." *Rehabilitation Psychology* 59 (1): 107–115. http://dx.doi.org/10.1037/a0035800.

Brown, Steven E. 2003. "I was born in a hospital bed (when I was 31 years old)." In *Movie Stars and Sensuous Scars: Essays on the Journey from Disability Shame to Disability Pride*, edited by Steven Brown, 61–69. New York, NY: iUniverse.

Eliot, George. 2007 [1871]. *Middlemarch*. London: Vintage.

Hahn, Harlan D. and Todd L. Belt. 2004. "Disability Identity and Attitudes Toward a Cure in a Sample of Disabled Activists." *Journal of Health and Social Behavior* 45 (4): 453–464. http://dx.doi.org/10.1177/002214650404500407.

Howard, Dana Sarah. 2015. "Transforming Others: On the Limits of "You'll Be Glad I Did It" Reasoning." *Res Philosophica* 92 (2): 341–370. http://dx.doi.org/10.11612/resphil.2015.92.2.9.

Acknowledgements Thanks very much, for helpful feedback and discussion, to Ross Cameron, Jon Jacobs, Laurie Paul, and two anonymous referees.

Linton, Simi. 1998. *Claiming Disability: Knowledge and Identity*. New York, NY: New York University Press.

McKinnon, Rachel. 2015. "Trans*formative Experiences." *Res Philosophica* 92 (2): 419–440. http://dx.doi.org/10.11612/resphil.2015.92.2.12.

Nario-Redmond, Michelle, Jeffrey Noel, and Emily Fern. 2013. "Redefining Disability, Reimagining the Self: Disability Identification Predicts Self-Esteem and Strategic Responses to Stigma." *Self and Identity* 12 (5): 468–488. http://dx.doi.org/10.1080/15298868.2012.681118.

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

Paul, L. A. 2015. "What You Can't Expect When You're Expecting." *Res Philosophica* 92 (2): 149–170. http://dx.doi.org/10.11612/resphil.2015.92.2.1.

Thompson, Tammy S. 1997. "Escape From Shame." *Mouth Magazine* 43.

Townend, Ellen, Deborah Tinson, Joseph Kwan, and Michael Sharpe. 2010. "Feeling Sad and Useless: An Investigation into Personal Acceptance of Disability and its Association with Depression Following Stroke." *Clinical Rehabilitation* 24 (6): 555–564. http://dx.doi.org/10.1177/0269215509358934.

Yalom, Marilyn. 2002. *A History of the Wife*. New York, NY: Harper.

Transformative Experience and Interpersonal Utility Comparisons

Rachael Briggs

Abstract: I consider an old problem for preference satisfaction theories of wellbeing: that they have trouble answering questions about interpersonal comparisons, such as whether I am better off than you are, or whether a particular policy benefits me more than it benefits you. I argue that a similar problem arises for intrapersonal comparisons in cases of transformative experience. I survey possible solutions to the problem, and point out some subtle disanalogies between the problem involving interpersonal comparisons and the problem involving transformative experience.

1 Preference Satisfaction

According to preference satisfaction theories of wellbeing, a person is better off in one scenario than in an alternative scenario if and only if the person's preferences are better satisfied in the first scenario than in the second. So for instance, I am better off in a scenario where I have a pet goldfish rather than a pet dog if and only if my preferences overall would be better satisfied in the goldfish scenario than in the dog scenario.

A few qualifications are in order. First, while I don't have space to give a detailed theory of preference, I will need to say a bit about what preferences are. Hausman (2011) distinguishes four concepts of preference: preference as level of enjoyment (so that I prefer chocolate ice cream to vanilla ice cream just in case I enjoy chocolate more than vanilla); preference as comparative evaluation (so that I prefer one political policy over another just in case I judge the first policy to be better than the second); preference as favoring (so that I prefer local contractors just in case I give them a better chance at getting a contract than non-local contractors, whether or not I believe that local contractors are better); and preference as choice ranking (so that I prefer soup over salad just in case I am disposed to choose soup and not salad when offered both as options).

While the first concept of preference might plausibly ground wellbeing, I will not consider it, since it does not generate the most obvious version of the problem I am interested in. And the third concept can be ruled out

Res Philosophica, *Vol. 92, No. 2, April 2015, pp. 189–216*
http://dx.doi.org/10.11612/resphil.2015.92.2.7

immediately as irrelevant to wellbeing. Preference as favoring has little to do with the ultimate value of things, or even anybody's idea of the ultimate value of things; it's often done to satisfy purely procedural requirements.

This leaves the first two concepts of preference as candidates: either preferences should be interpreted as dispositions toward choice, or preferences should be interpreted as comparative evaluations. For the purpose of explaining and predicting choice behavior, it is best to interpret preferences as choice dispositions. But this interpretation looks less appealing if preferences are supposed to constitute what is valuable to people. Choice dispositions are influenced by many states that seem irrelevant to wellbeing, including (possibly unjustified) beliefs about the options on offer. I will assume that the most plausible preference satisfaction theories take the fourth interpretation: preferences are comparative evaluations.

This choice might look worrisome: if to evaluate x more highly than y is to believe a mind-independent proposition 'x is better (for me) than y,' then the degree to which something satisfies my preferences is the degree to which I believe that it is good, or attribute goodness to it. But I claim that we can make sense of comparative evaluations even if 'better than' does not pick out any mind-independent relation, just as we can make sense of the concept 'looks red' even if 'red' is a response-dependent concept that does not pick out a mind-independent property. To rate x as better than y is to be in a state that plays a complex functional role: one that results in choosing x when one believes x and y are the only two available alternatives, stating verbally that x is better than y, and perhaps certain emotional responses toward x and y (when one recognizes them for what they are).

Second, I'll assume that a person's wellbeing depends on global, not just local, preference satisfaction. Even if I would rather have a dog than a goldfish the goldfish might make me better off overall—since it might be that having the dog thwarts my other preferences (such as the preference to have the freedom to travel and have someone look after my pet, or the preference not to have an animal destroy my couch cushions).

So I take preference satisfaction theories to claim that what is good for a person is for the person to get what she judges (in a global way) to be best. The problems I consider will be problems for preference satisfaction theories, so understood.

1.1 Utility

Philosophers and economists often assume that wellbeing can be measured by a utility function u, which maps each scenario w to a real number $u(w)$ representing how well off I am in w. Preference satisfaction theories can then be glossed in terms of utilities: the utility that a person assigns to world w is determined by features of their preferences, together with features of w that satisfy or fail to satisfy those preferences.

Given a utility function, we can consider two types of comparisons.

Ordinal Comparisons: of the form "person A is better off in scenario w than in scenario x."

Cardinal Comparisons: of the form "the benefit to A of inhabiting w rather than x is greater than the benefit to A of inhabiting y rather than z."

To assume that a person's preferences can be represented by utility functions is to assume that the greater-than ordering among real numbers accurately reflects the person's ordinal comparisons. This is fairly strong idealization. It requires that all scenarios be *comparable*—so that for any two scenarios, the person either prefers one to the other, or is indifferent between them. It also requires that the person's preferences be *transitive*, so that if x is preferred to y and y is preferred to z, then x is preferred to z, and *irreflexive*, so that no scenario is preferred to itself.

In realistic cases, people may fail to have transitive, complete preferences. Imperceptible differences between objects can create intransitive preferences. For instance, consider a person who rates their chocolate chip cookies on three criteria: size, sweetness, and darkness of chips. Suppose the person is contemplating a choice between three cookies, A, B, and C. The table shows the ranking of the cookies along each of the three dimensions, where 1 is the best rating and 3 is the worst.

	A	B	C
Size	1	2	3
Sweetness	2	3	1
Chips	3	1	2

Suppose the differences are very small, so that the person cannot tell the difference between a 1 and a 2, or a 2 and a 3, along any dimension: a cookie with sweetness 2 tastes indistinguishable from a cookie with sweetness 1. But the person can distinguish, along each dimension, between items marked 1 and items marked 3. Where a is the scenario in which the person eats A (and similarly for b and c), it seems reasonable to suppose that $b \succ a$, because B is better than A on the dimension of sweetness—the only dimension on which the cookies detectably differ. Similar reasoning suggests that $c \succ b$ and $a \succ c$, in violation of transitivity.

A case of incomplete preferences can be adapted from Raz (1988, 342). A student is contemplating two scenarios: one where she pursues a career as a lawyer, and another where she pursues a career as a clarinetist. The two scenarios are valuable in different ways: the law career provides security and financial stability, while the clarinetist career provides excitement and creative fulfillment. The student does not prefer either scenario to the other. However, the student is not indifferent between the two scenarios: though she prefers a clarinetist career with a bonus of $100 per year to the original clarinetist career, she does not prefer the clarinetist career with a bonus to

the law career (which she would, if she were indifferent between the law career and the original clarinetist career).

Despite these difficulties, I hold that measuring preference satisfaction with utility functions makes sense. Cases of preference incompleteness can be accommodated requiring that each person's preferences be extendable to a complete, transitive, and irreflexive ordering—in other words, that there be some way of filling in preferences or indifferences where the person lacks an opinion so that the resulting set of preferences is complete, transitive, and irreflexive. Someone with incomplete preferences that are extendable in this way can be represented by a set of candidate utility functions—one corresponding to each way of extending the person's preferences. We can then say that a claim about the person's level of wellbeing is true if and only if it comes out true according to each candidate utility function.

But some intransitive preference orderings, including the preference ordering in the cookie example, cannot be extended to complete, transitive, and irreflexive orderings. For a person whose preference ordering is not extendable in this way, we might consider the set of transitive, complete preference orderings that are most similar to the original preference ordering. We can then say that a claim about the person's level of wellbeing is true if and only if it comes out true according to each preference orderings that is among the most similar to the original preference ordering.

The use of utility functions does not commit the theorist to any claims about cardinal comparisons. A utility function may either be cardinal, or merely ordinal. This is often understood in terms of which transformations to the utility function are allowable. A merely ordinal utility function, which provides only information about a person's preferences between scenarios, is equivalent to any other ordinal utility function that orders the scenarios in the same way—so squaring the utilities, or doubling them, has no effect on how the utility function represents the world. For a cardinal utility function, the only allowable transformations are positive linear transformations, which consist of uniformly adding a number to all utilities, or multiplying all utilities by some positive number. When any of the claims or arguments below requires appeal to a cardinal utility function, I will note this explicitly.

In addition to *intra*personal ordinal and cardinal comparisons, theorists often wish to make *inter*personal utility comparisons. But interpersonal comparisons pose a prima facie problem.

2 The Problem of Interpersonal Utility Comparisons

If wellbeing (as represented by the utility function) is ultimately grounded in preference, then we can easily say what it is for me to be better off in world w than in x: it is for for me to prefer the features of w, all things considered, to those of x.

But when it comes to comparing my wellbeing to yours, things are trickier. There is no single preference ordering that compares how much I prefer the scenario I'm in with how much you prefer the scenario you're in. So how can the preference satisfaction theorist compare wellbeing across individuals? This is commonly known as the *problem of interpersonal utility comparisons*.

The problem is not about intuitive understanding; people often make informal judgments about who is better off than whom, or about whether a particular policy benefits one person more than it benefits another. Rather, the problem for the preference satisfaction theorist is to demonstrate that such interpersonal comparisons are compatible with her views on the nature of wellbeing.

There are really three types of comparisons we might want here (List 2003):

> **Level Comparisons:** of the form "A is better off in scenario w than B is in scenario x."
>
> **Unit Comparisons:** of the form "the benefit to A of inhabiting w rather than x is greater than the benefit to B of inhabiting y rather than z."
>
> **Zero Comparisons:** of the form "A's wellbeing in w is better than an absolute level of 0."

The three types of interpersonal utility comparisons are logically independent: for any set of them, there exist utility measures which admit of exactly those comparisons, and no others. (Unit comparisons presuppose that the utility functions in question are cardinal; level and zero comparisons do not.) In addition to being logically independent, the three types of comparisons also have different uses in ethics.

Level comparisons are crucial to egalitarian principles, which demand that people be made equally well off, and for prioritarian principles, which demand that social policies benefit the least well-off members of society. One might try to sidestep these issue by advocating equal (or prioritarian) distribution of something other than wellbeing—such as money, or primary goods. But the rationale for the equal distribution of resources is that resources are a good and easy-to-measure proxy for wellbeing. In cases where the same resources have disparate effects on different people—for instance, where one of the goods distributed is a staple food to which part of the population is allergic—equal distribution of primary goods does not suffice for fairness.

Unit comparisons are crucial to the utilitarian principle that one should act so as to maximize total utility. To assess how much whether option x produces more total utility than option y, a utilitarian must compare the sizes of harms and benefits that the move from x to y will produce for different people. (And as long as the population is fixed, this is all that the utilitarian needs.)

Zero comparisons are important for measuring total utility in worlds where the population is variable. Is introducing more people into the world a good or bad thing? That depends on whether the new people's lives are worth living. We may also need zero comparisons to make judgments about entitlement: if there is a moral difference between benefits that bring wellbeing up to the level that the recipient is entitled to, and benefits above that level, then we will need zero comparisons to draw the line.

It might appear, at first glance, that utility comparisons should be restricted to political and moral disputes, which involve trading off the interests of different individuals against each other. But cases of transformative experience give rise to analogous practical problems, which involve tradeoffs between the interests of one individual in at different times or in different scenarios.

3 The Problem of Intrapersonal Utility Comparisons

In what Paul (2014, 15-16) calls *personally transformative experience*, an individual undergoes a radical change to her point of view. Paul gives a list of examples: "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 your parent, making a major scientific discovery, or experiencing the death of your child." These experiences are so radically new that they change the preferences and values of the person who undergoes them. Paul argues that personally transformative experiences pose problems for decision-makers.

One of the problems, which I raise primarily to set aside, is epistemic. Paul's central interest is in experiences that are not just personally transformative, but also *epistemically transformative*, giving the person who undergoes them new knowledge of what the outcomes of actions are like. When deciding whether to undergo an epistemically transformative experience, Paul argues, the decision-maker typically does not have enough information to choose rationally—she lacks the requisite knowledge of what the possible outcomes are like, and so cannot rationally assign utilities to them. This problem is hard to present in the preference satisfaction framework, which seems best motivated by the view that there are no *objectively* correct preferences.

This second problem is normative, and easy to capture within the preference satisfaction framework: how does one negotiate between one's earlier values and one's later transformed values? Paul illustrates this problem using a science-fictional example: suppose you have the opportunity to be implanted with a new microchip that replaces the sense of taste with a completely new sense, never before possessed by humans. Paul writes,

> Before having the chip, perhaps I assign a low value to revelation, since I love good food and fine wine. But if I were to get chipped, I'd assign revelation a high value, embracing the new, tasteless me. When assessing what to do, which preferences trump? Which ones am I supposed to use to make my decision? (41)

Paul is concerned with experiences that are both epistemically and personally transformative, where both the epistemic and the metaphysical problems are at play. I will focus on personally transformative experiences, and the metaphysical question they raise. What does wellbeing amount to in the face of a personally transformative experience? A complete answer to this question requires intrapersonal level comparisons, unit comparisons, and zero comparisons.

3.1 Level Comparisons

Level comparisons are important in cases of transformative choice, where an individual decides whether or not to undergo a transformative experience. Many cases of personal transformation are also cases of transformative choice: having a child, moving cities, changing careers, devoting one's life to a cause, and joining a religion are often chosen, rather than forced. But transformative experience and transformative choice are not quite the same: there are cases of unchosen transformation, and cases where a person makes a transformative choice *not* to undergo a transformative experience.

As an illustration of the importance of level comparisons, consider Paul's microchip example example. When deciding whether to have the chip implanted, I should ask: will I be better off in a future where I enjoy the new sense experiences brought on by the chip, or in a future where I continue to indulge my love of gourmet food? I must compare levels of preference satisfaction not across people, but across two scenarios in which I have different preference orderings.

Level comparisons matter not only for transformative choices made on one's own behalf, but also for choices that influence the transformative experiences of others. Decisions about how to educate children shape their preferences. So do decisions about whether to authorize medical procedures aimed at helping children "fit in" to mainstream expectations about their bodies, such as cochlear implants for deaf children or cosmetic surgeries performed on intersex children. Guardians of children are meant to choose with the child's wellbeing in mind—but when the consequences of the choice include the child's ultimate preferences and sense of self, it is not at all clear what wellbeing even amounts to.

3.2 Unit Comparisons

Often, it is uncertain whether a given choice will lead to a transformation. Maybe walking into a Catholic Church will result in my conversion, but maybe I'll just be bored by the liturgy. Maybe I'll be transformed by love for my first child, but maybe we'll fail to bond, and I will remain my boring, selfish self. Maybe ingesting *salvia divinorum* will open up new vistas of perception for me, or maybe it will only make me feel ill. When is it a good idea to risk a transformation? To answer this question based on considerations of wellbeing, we need level comparisons as well as unit comparisons.

To see why, consider a variant on Paul's microchip experiment. The inventors of the chip have decided to run a controlled study on its effects, which involves implanting the chip in 50% of subjects, and giving the remainder a placebo chip that leaves their underlying sensory capacities, preferences, and values intact. Given my current sensory capacities, preferences, and values, getting the placebo chip is a mildly bad thing: I'll have to miss work, the implant will hurt for a few days, and my incision might become infected. Should I enroll in the study (that is—take a lottery between receiving the chip and receiving a placebo), or should I stay as I am?

To answer this question, it is not enough to know that the (genuine) chip is beneficial to my wellbeing. (If the chip is harmful, that suffices to show I shouldn't enroll in this particular experiment—but we can cook up a similarly puzzling experiment by making the control condition better than my current condition.) I must also assess whether the amount of wellbeing that I stand to gain from the chip is greater or less than the amount I stand to lose from the placebo. According to traditional decision theory, it is reasonable to enroll in the experiment if and only if enrolling has positive expected value—and assuming that the only relevant value is my wellbeing, this happens if and only if the chip benefits me more than the placebo harms me.

Unit comparisons are also important in situations where someone expects to undergo a personal transformation (chosen or not) and hopes to choose wisely in light of that expectation. Run-of-the-mill decision problems involve trading off present costs against future benefits: should you sleep in, or get started on the odious tasks that you hope to complete by midday; skip your daily run, or continue your long-term project of getting into shape; buy a new suit, or set aside the money for a down payment on a house? To make these decisions, you must compare the present cost of an action to its anticipated future benefit. For instance, in the sleeping-in example, you must consider the magnitude of the difference that sleeping in rather than waking up makes to your current wellbeing, and compare it with the magnitude of the difference that struggling with your tasks rather than being finished with your tasks will make to your future wellbeing.

When your current preference ordering is different from your later preference ordering, these kinds of unit comparisons are not straightforward. Suppose you have just enlisted in the military, and you expect that you will come to value discipline. On your last night at home, you are deliberating about whether to practice folding your socks with military precision. Right now, you find the task dull and valueless, but you expect that once you've finished with your training, you will be grateful for the extra practice. Should you spend your evening folding socks? Or suppose you are pregnant with your first child, and are deciding between an evening out with friends (favored by your current childfree preferences) and an evening at home reading books on parenting (favored by your future devoted-parent preferences). Which should you choose?

To answer these questions, you must compare the cost of folding socks or staying in and reading (according to your earlier preferences) with the benefits of greater military precision or parenting skills (according to your later preferences). You must compare benefits and harms not across people, but for yourself in two scenarios where your preferences differ.

3.3 Zero Comparisons

Zero comparisons, as I've said, are relevant to decisions about when to bring people into existence. These comparisons might already be considered intrapersonal: they are, after all, questions about whether a given person is better off existing than not. There is a related question about ending existence: at what point is a person better off ending his or her own life? Ceasing to have a point of view at all might be conceived as a dramatic type of personal transformation. And to assess when a person is better off existing than not existing, we must assess the welfare of the person (given that person's current preferences) against a basic threshold (which corresponds to the level of wellbeing of someone with no preferences at all).

3.4 How Big Is the Problem?

There are a great many candidate cases of personal transformation—some genuinely problematic, some only apparently puzzling. Here is an incomplete catalogue of examples.

Updating Preferences: In the morning, I typically prefer to get up and work rather than sleep. At the end of the day, I prefer to sleep.

Changing Tastes: I begin my life as a patriotic American who prefers peanut butter over Vegemite. After a long stint in Australia, I end up preferring Vegemite over peanut butter. Or: in my youth, I am fond of free-verse poetry, but as I age, I find myself more and more drawn toward formalism.

Changing Personal Values: I start out career-focused, but as I age and mellow, I become more invested in my friendships. Or: as I become close friends with a particular person, I come to see greater value in friendship with that person than with equally virtuous strangers.

Changing Moral Outlooks: After years as a staunch meat-eater, I learn about the horrors of factory farming. I come to prefer eating veggie burgers over eating hamburgers. Or: After years of atheism, I have a religious experience and convert to Orthodox Judaism. After my conversion, I stick loyally to Orthodox interpretations of Jewish laws—when in the earlier part of my life, I would have found such obedience inconvenient and silly.

All of these examples involve *some* change to my point of view. The change in the updating case seems routine, while the change that causes me to keep to Orthodox laws seems radical. But where should we draw the line between the routine and the radical?

I propose the following (somewhat fuzzy) distinction. If a change in my behaviors and choice dispositions can be explained by my having a fixed underlying set of preferences over properties of situations, together with a change of beliefs about which situations have those properties, then it is routine rather than radical. For example, when I switch from preferring wakefulness to preferring sleep, the best explanation is that I have a stable preference to be awake during the day, and asleep at night. My beliefs about whether it is day or night change with position of the sun, but my underlying preferences remain the same. My change from meat-eater to vegetarian can likewise be explained by stable underlying preferences, coupled with a change of belief. I prefer the harmless enjoyment of delicious meat over an acceptable but dull veggie burger, but I value avoiding serious harm to other sentient beings over my own culinary pleasure. When I learn about factory farming for the first time, I learn that the delicious meat on offer is not so harmless after all.

Some of the examples are tricky: there are multiple ways of filling them in, and it's not always clear which is best. When I change from being a peanut-butter-loving American to being a Vegemite-loving Australian, is the best explanation a change in which flavors I value, or a change in my beliefs about what's tasty (coupled with a stable desire for whatever is tasty)?

While there are tricky borderline cases, not every example can be explained by appeal to stable underlying preferences. In the religious conversion case, there is nothing that plausibly plays the functional role of a preference and (in conjunction with my beliefs) and explains both my earlier atheist choices and my later religious ones. A theorist could claim that I have a stable preference to abide by whatever religious rules are true. But unless I am disposed to avow such a preference, or my emotions about possible courses of action depend in some way on whether I associate

them with religious truth, I do not have the sort of state that counts as a preference to abide by whatever religious rules are true.

Likewise, the case of changing personal values is hard to explain by appealing to changes in my beliefs about what will ultimately satisfy me, combined with a preference to end up satisfied. If I have no category of 'satisfying,' or if I am not independently motivated by recognizing that something will lead to my being satisfied (once I have realized that it is an instance of friendship, or success), then it is wrong to describe me as having a stable desire or preference for whatever will satisfy me.

Bradley (2009) shows that if we constrain ourselves to a fixed, countable algebra of propositions that are the objects of belief and desire, then some changes in preference cannot be explained by appealing to changes in belief. On the other hand, if we vary the domains of believed and desired propositions arbitrarily, then every change in preference can be modeled by a change in belief. I claim there are right and wrong ways to fix the domains of belief and desire, so that some logically possible models of a believer are nonetheless inaccurate.

A more thoroughgoing skeptic could push back, and insist that in some (or all) cases, there is no fact of the matter about which changes to my choice dispositions are best explained by my changing preferences, and which are best explained by stable preferences combined with changing beliefs. Such a thoroughgoing skeptic should be even more worried about the problem of intrapersonal utility comparisons (if that skeptic is tempted by preference satisfaction theories of wellbeing). On such a skeptical view there is unlikely to be a *unique* right way of assigning stable underlying preferences to an agent, and different ways of assigning stable preferences may lead to different and incompatible interpersonal utility comparisons.

Paradigm cases of transformative experience result in a profound change to a person's outlook, and make it hard to find underlying, stable evaluations that explain the change. They therefore represent a particularly acute version of the problem of intrapersonal utility comparisons of time—one that is unlikely to admit of easy solutions.

4 Solutions

In cases of transformative experience, the problem of interpersonal utility comparisons creates intrapersonal analogues. In this section, I survey proposed solutions to the problem of interpersonal utility comparisons, and assess how well they handle the intrapersonal version of the problem.

4.1 Doing Without

One option is to deny the possibility of utility comparisons across preference orderings. The preference satisfaction theorist who takes this route will need

to explain why wellbeing is a useful concept, despite being incomparable across—and sometimes within—individuals.

In the interpersonal case, this response looks unappealing, since much of ethics relies on interpersonal utility comparisons. Not only do many popular principles—such as Rawls's difference principle and Bentham's rule of utilitarianism—rely on interpersonal utility comparisons,[1] but such comparisons seem built into the nature of the task. Social policies must ultimately be justified in terms of their benefits to individual citizens. Since people's interests are not always aligned, costs to one person have to be traded against benefits to another.

Still, we might give up on the view that justice can be explicated in terms of wellbeing. Sen (1970b) argues that there are situations where a strict increase in everyone's wellbeing would involve a rights-violation, and would therefore be unjust. There is then a substantive question about whether the concept of wellbeing is doing any work, and what else we need to give an account of justice. (Perhaps rights would be a crucial ingredient.)

Another option is to formulate social justice principles that require only intrapersonal comparisons. For instance, one could argue in favor of a particular *social choice rule* which took as input sets of individual ordinal preferences, and returned as output a single ordering of options according to how good or bad they were for the group. However, it is well known that without interpersonal utility comparisons, no social choice rule can satisfy all of the following principles (Arrow 1963). In fact, it is a necessary and sufficient condition for the existence of a rule satisfying all four principles that at least one type of interpersonal utility comparison be possible.

> **Universal Domain:** If each individual in the group has a complete, transitive, irreflexive preference ordering, then the social choice rule outputs a complete, transitive, irreflexive preference ordering for the group.
>
> **Weak Pareto:** For any two scenarios x and y, if each individual in the group prefers x to y, then the group prefers x to y.
>
> **Non-Dictatorship:** There is no individual i such that no matter what the preferences of other group members, the group's preference ordering coincides with i's preference ordering.
>
> **Independence:** For any two scenarios x and y, the group's preference between x and y depends only on the overall pattern of individual preferences x and y, and not on individuals' opinions about any other pair of alternatives.

So a preference satisfaction theorist who wants a social choice rule, if she rejects interpersonal utility comparisons, will have to decide which intuitively compelling principle to give up.

[1] Neither Bentham nor Rawls is a preference satisfaction theorist, but both the advice to seek the greatest good for the greatest number, and the advice to choose the social policy that benefits the least well off, can be recast in preference satisfaction terms.

If we are prepared to get rid of interpersonal comparisons of utility, doing away with intrapersonal comparisons might not seem like much of an additional loss. Justice may require trading off the interests of different individuals, but it does not require trading off among the interests of one individual. But contrary to appearances, giving up on intrapersonal utility comparisons does have additional costs.

One cost is that, in cases of transformative choice and personally transformative experience, we seem to lose the concept of prudence. Prudent choices are those that are likely (by the chooser's lights) to maximize the chooser's overall level of wellbeing. Even setting aside Paul's epistemic concerns about whether the chooser is entitled to count any outcome as likely or unlikely by her own lights, conditional on her undergoing a transformative experience, there is a normative question about what counts as the goal of prudence in cases of transformative experience and transformative choice. Those who deny that justice can be defined in terms of wellbeing can attempt to define justice in other terms. But defining prudence in terms of something other than wellbeing looks doomed to failure; what else could prudence be?

The preference satisfaction theorist might claim that the concept of prudence does not apply in tricky cases involving personal transformation—just as Paul argues that there is no possibility of a rational response to a choice that is both personally and epistemically transformative. But this response does not do enough to address the problem. If ordinary changes in tastes and values count as personally transformative, we may find that the concept of prudence applies to hardly any real-life cases.

Another possible response is to try to get by with a purely formal concept of rationality, rather than appealing to a substantive concept of prudence.[2] Suppose I am deciding whether to undergo Paul's microchip experiment. Rationality bids me to do whatever maximizes the expected satisfaction of my current preferences—perhaps with the rider that those current preferences be reasonable. If I'm now horrified by the prospect of losing my sense of taste, then I should refuse the microchip, regardless of how much I may come to enjoy my novel sensory experiences after the operation. On the other hand, if I now wish to ensure a pleasant future for myself, and I think the microchip is likely to give satisfaction, then I should be willing to put up with some degree of pain and revulsion now in order to ensure that my later self will get to enjoy the effects of the chip.

The second response, however, faces a dilemma. Either it yields an unsatisfyingly thin concept of rationality, or it smuggles in considerations of prudence. Suppose I don't care at all about my preferences after a transformation—I think my post-religious-conversion self is too corrupt to be worth caring about, or my future sophisticated tastes in wine are hopelessly stuffy, or the microchip will present me with a hopelessly distorting

[2] I thank Miriam Schoenfeld for the suggestion.

picture of the world—and so I fail to take my transformed preferences into account. Have I made an error? If the answer is no, then the concept of rationality in question is disappointingly thin. Many authors think there is something wrong with me if I fail to take my future desires into account, including Parfit (1984, 123–126) and Nagel (1986), Without intrapersonal utility comparisons, preference satisfaction theories are ill-placed to explain why and how I should defer to my future preferences.

But if the answer is yes, then there are at least some wrong answers to the question "how much weight do my future preferences deserve?" And to sort the right from the wrong answers, it is crucial to compare the strengths of my current preferences with the strengths of my future preferences. (Even if it is permissible for me to accept a slightly smaller benefit now for a slightly larger cost later, it is permissible because the cost and the benefit do not differ *too much* in magnitude—a unit comparison.)

In addition to these problems for the concept of prudence, giving up on intrapersonal utility comparisons creates new problems for the concept of justice. In cases where a person faces a transformative choice, or is likely to undergo an unchosen transformative experience, we often need interpersonal utility comparisons to determine which of two alternatives will make them better off. So any theory of justice that relies on interpersonal utility comparisons—including social choice rules that violate one or more of Arrow's principles—will have trouble when members of society encounter (or choose whether to encounter) personal transformations.

So while giving up on interpersonal utility comparisons is already costly, giving up on intrapersonal comparisons adds new costs.

4.2 The Zero-One Rule

The Zero-One Rule is based on Neumann and Morgenstern's (1953) theory of utility, which treats utilities as cardinally (and not just ordinally) meaningful. Von Neumann and Morgenstern assume that utility is bounded, and stipulate that for each individual, the best outcome has utility 1 and the worst outcome has utility 0. It is then tempting to think that the numbers 0 and 1 mean the same for everyone, so that my utility 0 is the same as your utility 0, and my utility 1 is the same as your utility 1. Call this assumption the Zero-One Rule. The Zero-One Rule would enable us to make both unit comparisons (since the distance between my utility 1 and my utility 0 must equal the distance between your utility 1 and your utility 0) and level comparisons (since we can use intrapersonal cardinal comparisons to ensure that for each n between 0 and 1, my utility n is the same as your utility n).

Critics of the Zero-One Rule complain that it is unfair. Hausman (1995) cites several versions of this objection. (Hausman believes that the Zero-One Rule is a consequence of the preference satisfaction theory, but leads to ethically unacceptable results. I will address his views in the next section.

The critics Hausman cites think that the Zero-One Rule is false, because it leads to ethically unacceptable results. This section is an attempt to respond to their arguments.)

> Consider some undemanding person who achieves his upper bound at a low level of consumption. Do we normalize that person's utility scale so that it has the same upper and lower bounds as that as a greedy person? If so, and if we distribute goods to each individual so that each achieves, say 90% of maximum utility, then the greedy person is likely to be given much more than one feels he deserves. (Hammond, 1993, 216)

> Thus for example the Zero-One Rule implies that, all other things equal, greater social utility results from educating people to have simple desires and to be easily satisfied; and that such persons will generally have stronger claims. They are pleased with less and so presumably can be brought closer to their highest utility. (Rawls, 1971, 323)

> It may be argued that some systems, e.g., assigning in each person's scale the value 0 to the worst alternative and the value 1 to the best alternative, are interpersonally "fair," but such an argument is dubious. First, there are other systems with comparable symmetry, e.g., the system . . . of assigning 0 to the worst alternative and the value 1 of the *sum* of utilities of the other alternatives. Neither system is noticeably less fair than the other (one assumes equal maximal utility for all and the other assumes equal average utility for all), but they will yield different bases of social choice. Second, in comparing the utility measures of different persons, one may wish to take account of interpersonal variability of capacity for satisfaction, e.g., one may wish to give special consideration to handicapped people whose enjoyment measure may be thought to be universally lower. (Sen, 1970a, 98)

Here is a way to spell out the problem. To take the Zero-One Rule seriously, we would have to say that people who could conceive of wonderful unactualized scenarios were thereby worse off, and people who could conceive of horrible unactualized scenarios were thereby worse off. But it is absurd to think that the ability to conceive or wonderful or horrible scenarios changes people's welfare in the way the Zero-One Rule requires. Even if being able to imagine better alternatives turns out to cause unhappy moods or social maladjustment, it does not constitute a harm in itself.

There are similar objections to the Zero-One Rule in the intrapersonal case. If being able to imagine good possibilities makes you genuinely worse

off, then it should be possible for you to improve your own wellbeing by maiming your imagination. Likewise, if being able to imagine bad possibilities makes you genuinely better off, then it should be possible for you to improve your wellbeing by thinking up new and increasingly terrifying Lovecraftian horrors. But neither way of manipulating your possible circumstances, while leaving your actual circumstances intact, seems like an acceptable way of improving your wellbeing.

While these objections are typically cast in moral terms, they are ultimately driven by a nonmoral thought: we have some pre-theoretic grasp on when it is that one person is better off than another, and merely adding or removing options should not make a difference to anyone's wellbeing, according to this pre-theoretic concept. It is not just wrong for me to maim your imagination in the name of making you better off; it is imprudent for you to maim your own imagination to improve your wellbeing. So Griffin (1993, 120) puts the objection correctly when he writes, "the Zero-One Rule is just false. It is not the case that we all reach the same peaks and valleys."

This diagnosis also suggests a possible line of response. The objector to the Zero-One Rule has the space of possibilities wrong. If we consider a person's feasible options, or the options that the person has actively considered, then we are leaving out possibilities. We want a broader range—all the person's *de re* metaphysical possibilities.

A possible challenge to this line of response runs as follows. What does it mean for a person's preferences to be satisfied by a scenario they have not even considered? If I haven't thought about the option of living in an ashram, it seems otiose to ask whether I prefer that option to my current job. Living in the ashram may well be better or worse for me than my current job, but this is not because of my preferences—I have no relevant preferences.

While a complete answer to this challenge would go beyond the scope of this paper, a promising line of response is that unconceived possibilities may satisfy my preferences in virtue of sharing salient features with conceived possibilities. Even if I have not thought about living in an ashram, I have opinions about how much I value community (moderately), hard work (I detest it), and religious experience (not at all). These facts about me determine that living in an ashram satisfies my preferences less than my current job.

4.3 Primitive Comparisons

Another possible solution is best explained by considering an objection to preference satisfaction theories developed by Hausman (1995). Hausman claims that preference satisfaction theorists are committed to the Zero-One Rule. Any alternative to the Zero-One Rule, he claims, must tacitly appeal to a non-preference-based theory of wellbeing, such as hedonism. In fact,

the claim that the Zero-One Rule gets the wrong answers (as pressed by the objectors in the previous section) already presupposes a non-preference-based standard of rightness.

Hausman claims that if preference satisfaction theorists are right, then any rule for making interpersonal utility comparisons must satisfy two desiderata. Say that a person's *non-comparative wellbeing* is the position at which she ranks the actual scenario, relative to the top and bottom of her preference ordering (so that the higher up I move in my preference ordering, the greater my non-comparative wellbeing is). A person's *comparative wellbeing* is the extent to which she is better or worse off than others. Where Ira and Jill are any two individuals:

> (1) If anything diminishes Jill's comparative well-being without affecting the preference satisfaction of anyone else, then it diminishes her non-comparative well-being.
> (2) If Ira and Jill have identical preferences and are at the same position in this preference ranking, then they are equally (comparatively) well-off. (481)

The Zero-One Rule is the only rule that satisfies both of these requirements.[3]

Weintraub (1998) raises persuasive objections to Hausman's first desideratum. Ira may be at the top of *his* preference ranking without being at the top of *the universal scale* of preference rankings, just as Ira may be experiencing as much pain as he is capable of experiencing without experiencing as much pain as anyone is capable of experiencing. Or, to coin a new analogy, if Ira is running at his maximum possible speed, and Jill is running at her maximum possible speed, it does not follow that they are running at the maximum speed possible for anyone.

On Hausman's interpretation, Weintraub's response does not make sense. Hausman holds that all preference satisfaction theorists are committed to the view that whether an option makes a person better off depends only on its relative position in that person's preference ranking—i.e., how distant it is from the top and bottom of the ranking.

I think the preference satisfaction theorist should reject Hausman's narrow interpretation of the theory. In addition to facts about whether an option is at the top or the bottom of my preference ranking, there may be further facts about whether my preference between x and y is twice as intense as your preference between z and w. Such facts are properly interpreted as facts about preference satisfaction, and they could ground

[3] Suppose Ira and Jill both have the same option @, which ranks at the top of both their preference orderings. If Jill had the same preferences as Ira, then by 2, they would be equally well off. But transforming Jill's preference ordering into one that matched Ira's preference ordering, while leaving her with @, would have no effect on her non-comparative wellbeing; this is a consequence of the definition of "non-comparative wellbeing." By 1, such a transformation cannot change Jill's comparative wellbeing. Therefore, Jill and Ira are equally well off. The same argument can be run to show that whenever Jill and Ira are at the bottom of their respective preference orderings, they are equally well off.

level comparisons. Likewise, if the preference satisfaction theorist posits desires with absolute strengths (as Griffin [1987] does), then facts about the relative strengths of these desires might ground level comparisons. If my desire for food is twice as strong as yours, then there is a good sense in which feeding me satisfies my preferences twice as well as feeding you satisfies yours. Appealing to such primitive facts is consistent with the motivations of the preference satisfaction theory, even if it is not consistent with the letter of Hausman's interpretation of the theory.

It is true that these primitive facts about interpersonal comparisons would be harder to observe than facts about the relative rankings of options within an individual. To get information about my ordinal preferences, such as whether I would rather eat beef or eat eggplant, you can simply offer me a choice between the beef and the eggplant. (This isn't an infallible test—you'll get the wrong result if I mistake the beef for tofu, and you'll have trouble distinguishing weak preference from indifference based on my choice alone—but it works in good circumstances.) If you know that I rank eggplant above chicken and chicken above beef, you can test the cardinal strengths of my preferences by offering me various choices between the chicken on the one hand, and weighted lotteries between the eggplant and the beef on the other. (This isn't infallible either—you get the wrong result if I'm bad at reasoning about probabilities, or if the lotteries are presented in a misleading way—but again, it works in good circumstances.) There is no similarly straightforward way to test whether my preference for eggplant over chicken is stronger than your preference for chicken over beef—no pair of choices you can offer us where our choice behavior will entail anything useful about the relative strengths of our preferences.

However, it's at least logically possible for the comparative facts to exist without being directly observable. The indirect effects of different levels of preference satisfaction may include different expressed emotions: in cases where people know the facts about the world that are relevant to their preference satisfaction, then, all other things being equal, people whose preferences are better satisfied are likely to be happier. Furthermore, if there is some resource that different people tend to value equally (perhaps money is such a resource), then we can make interpersonal comparisons of preference strength by seeing how much of the resource people are willing to spend to have the preference satisfied when they start out with similar amounts of the resource. (There is no resource for which it is necessary or *a priori* that people tend to prefer it with equal strength, but there may be some resource that, as a contingent matter of fact, people tend to prefer with equal strength. And this is all that is required for interpersonal comparisons of preference strength to be indirectly observable.)

A complete defense of the primitive comparisons solution would include a theory about how to measure the primitive comparisons. Whatever this theory looks like, it seems likely that it will treat intrapersonal comparisons and interpersonal comparisons symmetrically. Comparing two people's

facial expressions, reported feelings, or spending behavior presents much the same challenges as comparing a single person's facial expressions, reported feelings, or spending behavior across two different times.[4] So on the primitive comparisons account, there is no special reason to think that the intrapersonal comparisons differ in kind from the interpersonal ones.

4.4 Value Judgments

Some authors suggest that interpersonal utility comparisons are value judgments. If true, this conjecture would explain why observable facts are not enough to settle the facts about interpersonal utility comparisons: empirical observations are generally not sufficient to settle disputes about what is good or valuable. Perhaps interpersonal utility comparisons are grounded in the way that other value judgments are grounded: by reason, instead of observation. But are interpersonal utility comparisons value judgments? What would it mean for them to be value judgments? I'll consider a few ways of spelling out the view.

A famous proponent of the view is Robbins (1938), who contrasts the assumption that all people have equal capacity for satisfaction—a claim about interpersonal utility comparisons that rests on "an ethical principle"—with "The analysis of the effects of a small tax on particular prices and quantities of particular products," which "would rest upon scientific demonstration." Robbins's idea is that interpersonal utility comparisons are not observable by economists, nor do they play a role in explaining and predicting the behavior of economic actors. Instead, they play a normative role in shaping policy. Therefore, they must be value judgments.

Scanlon (1993, 20) distinguishes three ways that interpersonal utility comparisons might count as value judgments. They might be:

(1) "moral judgments about the kind of consideration we owe to each other" (such as the judgment that each person has a right to a certain amount of clean drinking water, or a certain degree of autonomy)
(2) "judgments about what makes a life better for the person who lives it that figure in the process of defending a general criterion of well-being" (such as the judgment that pleasure is good for everybody), or
(3) "judgments about what makes a life better for the person who lives it that figure in the process of arriving at particular judgments of relative well-being" (such as the judgment that Anne benefits more overall from her good health than Bob does from his achievements at work).

[4] An anonymous referee suggests that the problem is more tractable in its intrapersonal versions, since different people can be expected to vary in their levels of expressiveness. But couldn't one person's level of expressiveness vary over time?

Scanlon argues that preference satisfaction theorists are committed to the claim that interpersonal utility comparisons are value judgments in the first sense. Preference satisfaction theories "are fundamentally different in character" from hedonistic or experiential theories, Scanlon claims, "the preference satisfaction view being at base a moral doctrine, whereas experientialism is an account of the nature of value" (24).[5]

Scanlon uses Harsanyi's (1955; 1977) preference satisfaction theory to illustrate why preference satisfaction theorists are irrevocably pushed toward treating interpersonal utility comparisons as value judgments in the first sense. Harsanyi claims that some preferences are irrelevant to a person's wellbeing, including anti-social preferences based on envy or sadism, moral preferences based on political commitments that may conflict with liberalism, and preferences based on factual mistakes or faulty reasoning. Scanlon argues that the only justification for excluding these preferences is moral: a person is arguably entitled to the satisfaction of well-reasoned, non-moral, and non-anti-social preferences, but is not entitled to the satisfaction of mistaken, moral, or anti-social ones. "The choice of preference satisfaction as a standard of well-being, and the definition of a particular version of this standard, are decisions shaped largely by moral considerations, and not merely by ideas about individual good" (18).

While Scanlon makes a good case that the motivation for Harsanyi's preference satisfaction theory appeals to moral considerations, it does not follow that for Harsanyi, the content of interpersonal utility comparisons must be moral. The preference satisfaction theorist could make the following reply. We have to make a variety of value judgments (of the first type) to determine what property in the vicinity of preference satisfaction counts as the property of wellbeing. But this does not mean that the property that constitutes wellbeing is a moral property: degree of preference satisfaction (of the valued sort) may be a grounded in nonmoral things. That we value wellbeing (i.e., *that* property) is settled by moral considerations. But once we have picked out the property, Harsanyi might claim, interpersonal comparisons of it depend on its nature. Its features are difficult to observe, but real nonetheless.

This reply rests on the idea that judgments about wellbeing are value judgments in the third sense. Scanlon has a separate argument that for the preference satisfaction theorist, judgments about wellbeing cannot be value judgments in the third sense. So, if the reply is to be viable, the preference satisfaction theorist will need to rebut Scanlon's argument.

Typically, says Scanlon, deliberating agents do not take their preferences to be reasons for valuing the objects of preference. Rather, they take their preferences to be tracking independent reasons for valuing. Not only is this

[5] Scanlon's ultimate aim in the article is to refute preference satisfaction theories and replace them with a kind of objective list theory, but my aim here is to find the best solution to the problem of interpersonal utility comparisons from within the standpoint of a preference satisfaction theory.

true of derivative preferences (which might be disputed on the grounds that they fail to reflect the agent's more basic preferences), but it is true of the most basic and fundamental preferences.

I think that preference satisfaction theorists should be willing to adopt an error theory here: preferences (or perhaps the most basic preferences) do provide a reason for valuing the objects preferred, and deliberating agents are systematically wrong about this. This is a costly response, because it involves rejecting the phenomenology of decision makers as misleading. A fully developed version of it would include a debunking explanation of the phenomenology, perhaps along the lines offered by expressivists.

So I do not think that preference satisfaction theorists are forced to deny that judgments of wellbeing are value judgments in the third sense. They can push back against Scanlon, at the cost of having to explain why the phenomenology of decision making is partly in error. But suppose a preference satisfaction theorist does adopt Scanlon's conclusion, and accepts that interpersonal utility comparisons are value judgments in Scanlon's first sense instead. How might this view be extended to intrapersonal utility comparisons across time in the case of transformative experience?

Unlike interpersonal utility comparisons, intrapersonal comparisons do not seem to be moral in character. To decide whether to get Paul's science-fictional microchip, you may need to assess how well off you would be with the microchip and without it—but this is not a matter of assessing what anybody owes to anybody else. Likewise, when assessing how to trade off present effort against future comfort in the face of changing preferences, you don't seem to be making a moral judgment about what your current self owes to your future self. If you are making a value judgment, it is not a value judgment of the first kind.

Scanlon himself explicitly accepts that "not all judgments of relative well-being are made with morality in mind," and that "We can ask, quite apart from any question of right or justice, how well a person's life is going and whether that person is better off than another, or better off than he or she was a year ago" (18). Scanlon has in mind value judgments of the second kind—he thinks the interest of these questions lies in establishing a general criterion of wellbeing—but in cases of transformative experience, the preference satisfaction theorist needs to provide value judgments of the third kind. Claiming that interpersonal comparisons are value judgments of the first kind provides no traction in the case of transformative experience.

4.5 Extended Sympathy

Some authors, such as MacKay (1986) and Goldman (1995), propose that interpersonal utility comparisons are actually intrapersonal judgments in disguise. To ask yourself whether Socrates dissatisfied is better off than a pig satisfied is to ask whether you would be better off with Socrates tastes in

the scenario that dissatisfies Socrates, or with the pig's tastes in the scenario that satisfies the pig.

One defense of this view is grounded in a hedonistic or experiential conception of the good, on which a scenario is good for a person to the extent that it involves good or bad experiences. Since accurately imagining a scenario is a good way to acquaint oneself with the experiences it involves, it is clear why imagination would provide useful information about wellbeing.

However, a preference satisfaction theorist can also motivate the extended sympathy view. Harsanyi (1977) does so by appeal to

> the *similarity postulate*, to be defined as the assumption that, once proper allowances have been made for the empirically given differences in taste, education, etc., between me and another person, then it is reasonable for me to assume that our basic psychological reactions to any given alternative will be otherwise much the same.

If everybody would have the same response to being Socrates dissatisfied (given the same information and tastes), then to vividly imagine being Socrates dissatisfied gives you a good idea of how well Socrates's preferences are fulfilled, overall, in the scenario where he is dissatisfied. Likewise, to vividly imagine being the pig satisfied gives you a good idea of how well the pig's preferences are fulfilled, overall, in the scenario where it is satisfied. Once both levels of satisfaction are present in your mind, you can compare them.

Among philosophers, the extended sympathy view is typically given an epistemic interpretation, on which imagining yourself in Socrates's shoes is a good source of information about how well-satisfied his preferences are relative to other people's. We can contrast this epistemic interpretation with a metaphysical one, according to which what happens when you imagine yourself in Socrates's shoes grounds the facts about how well-satisfied his preferences are relative to other people's.

Let us first consider the epistemic interpretation of the extended sympathy view. Is it plausible? And does it solve the problem of interpersonal utility comparisons?

Paul's puzzles about epistemically transformative experience might suggest that the epistemic version of the extended sympathy view is not much use. Suppose that becoming like a dissatisfied Socrates would be epistemically transformative for you, involving experiences that are deeply alien to you. Then you can't know what it's like to be Socrates dissatisfied. If you can't know what it's like to be Socrates dissatisfied, then you can't accurately imagine his situation, and so imagination leaves you in no position to judge how good or bad things are for him.

The defender of the extended sympathy view can push back here. Goldman (1995) argues that you can know what it's like to be Socrates dissatisfied, and suggests a type of mechanism by which you can know. Normal

agents, he suggests, have mental mechanism that take in some mental states as inputs, and returns others as outputs; for instance, a decision-making system takes in desires and beliefs, and returns choices. In addition to feeding real inputs into these mechanisms, you can feed in simulated mental states. For example, if you have a system that takes in visual perceptions and returns verbal descriptions of them, you can see what it would return when fed in different perceptions by feeding it visual imaginings. Similarly, you know what it is like to be Socrates dissatisfied by feeding the simulated beliefs and desires of of Socrates into one of your mental mechanisms, and seeing what imagined experiences and feelings that mechanism returns.

Assuming that you have introspective access to your own preferences, and that the similarity postulate holds, Goldman's theory explains the following conditional: If you are able to feed simulated mental states of Socrates into mental mechanisms that will return simulated preferences, then you can make reliable judgments about whether Socrates dissatisfied is better or worse off than you in your current condition. Paul's concerns are about the antecedent of the conditional: maybe you are not able to feed enough simulated mental states of Socrates into your mental mechanisms to get a good output from them.

Unfortunately, while the epistemic version of the extended sympathy view addresses the epistemic problem of interpersonal utility comparisons—how do we know Socrates dissatisfied is better off than a pig satisfied?—it does not really address the normative problem—what makes it the case that Socrates dissatisfied is better off than the pig satisfied? It might be used to supplement another answer to the normative question; for instance, extended sympathy might be used to bolster the 'primitive comparisons' view by providing a mechanism by which we learn about preference strengths and absolute levels of desire in other people. But if the extended sympathy view is to work as an independent solution to the normative problem, it must be given a metaphysical interpretation. Philosophers who defend the metaphysical interpretation of the extended sympathy view are thin on the ground, but it is defended by some economists, including Hammond (1993) and Harsanyi (1977).

On the metaphysical interpretation of the extended sympathy view, the facts about relative preference satisfaction for different individuals are grounded in facts about the satisfaction of preferences within an individual. In particular, what makes true my judgment that Socrates dissatisfied is better off than a pig satisfied is that when I correctly imagine the situations of both these individuals, *I* prefer the situation of Socrates dissatisfied over the situation of the pig satisfied. Likewise, what makes true my judgment that satisfaction (versus dissatisfaction) makes a bigger difference to Socrates than to the pig, is that my preference between the situations of Socrates satisfied and Socrates dissatisfied is stronger than my preference between the situation of Socrates dissatisfied and the situation of a pig satisfied?

The metaphysical interpretation as I have construed it raises an awkward question: what happens when two people have different preferences between the situation of Socrates dissatisfied and a pig satisfied? Such disagreements seem possible, even among rational and well-informed observers. Is there any neutral standpoint from which to adjudicate them? If not, then the metaphysical version of the extended sympathy view gives only observer-relative answers to interpersonal comparisons of wellbeing. It may be that relative to my preferences, Socrates dissatisfied is better off than a pig satisfied, while relative to your preferences, the pig satisfied is better off than Socrates dissatisfied. Worse, since preferences can change, interpersonal comparisons of wellbeing would have to be relativized not just to individuals, but to individuals at worlds and times.

4.6 Rigidifying

Preference satisfaction theorists might try to avoid the problem of interpersonal utility comparisons by picking out a single preference ordering, and claiming that wellbeing for everyone consists of the satisfaction of the preferences in the privileged ordering. In the interpersonal case, we have seen that different people have different preferences. An objective list theorist might claim that one particular set of preferences (not necessarily the preferences of any actual individual) determined the good for everyone.

The resulting rigidified theory would typically be classified as a rival to preference satisfaction theories, rather than a type of preference satisfaction theory. Nonetheless, it is instructive to consider this rigidified theory, because its adaptation to the intrapersonal case will turn out to be an interesting type of preference satisfaction theory.

There are two ways to understand the objects of preferences on my rigidified objective list proposal. Suppose I would rather win than lose road races. This preference might be given a personal interpretation—I prefer scenarios in which *I* win road races—or it might be given an impersonal interpretation—I prefer scenarios in which *Rachael Briggs* wins road races. For me, the two interpretations go hand in hand: I know that I am Rachael Briggs. But how good is it for my rivals to get what I want? Intuitively, it is good for my rivals if they satisfy my preferences read personally—if *they* win road races. But it is bad for them if they satisfy my preferences read impersonally—if *Rachael Briggs* wins road races.

The personal interpretation of preferences is the most appealing way to supplement the rigidifying proposal. People's interests sometimes conflict: it is in each runner's interest that *she* win the race, and that others lose. Conflict would be impossible if the same impersonal state of affairs were best for everybody.

Although a preference satisfaction theorist cannot adopt the rigidifying strategy for all interpersonal comparisons and remain a preference satisfaction theorist, might she still adopt it for intrapersonal comparisons? How would such a view look?

In the case of transformative choice, there is a natural set of preferences to privilege: the set of preferences held in the actual world. Why not say that what is good for an individual in counterfactual scenarios is whatever satisfies her actual preferences? (This won't address the question of how to make tradeoffs in cases where a person has different preferences before and after a transformative experience. But it will address the question of whether a given transformation is desirable or not.)

This rigidifying strategy can accommodate a class of examples involving adaptive preferences—usually thought to favor objective list theories over preference satisfaction theories. Often, a possible change seems alien or corrupting to someone's identity, even though, were the person to experience the change, they would come to endorse it. The atheist who passes up the opportunity to convert to a religion whose principles she abominates; the proud Deaf man contemplating an alternative history where he receives a cochlear implant; the writer who contemplates an alternative, less tempestuous life as an office worker—all these people can say, "I would be worse off in that alternative life, even though I wouldn't know it."

The rigidifying strategy can't do everything the objective list theorist demands. There are cases where a merely possible preference adaptation would have made the transformed individual better off by the objective list theorist's lights: the anxious socializer who refuses to stop caring what her peers think of her; the jealous person who never learns to take joy in the accomplishments of friends; the picky eater who never comes to like new foods. But it's worth noting that the rigidifying strategy deals neatly with some cases that are commonly thought to favor objective list theories over preference satisfaction theories.

How might we apply the rigidifying strategy in cases where the chooser expects to undergo a transformative experience, and wants to make prudent choices in light of that expectation? Just like the interpersonal case, the case of transformative experience over time presents us with a choice about how to interpret the contents of preferences. Suppose that at some time t, I prefer eating Vegemite to eating peanut butter. My preference could be interpreted temporally, as a preference for a scenario where I now eat Vegemite over one where I now eat peanut butter. Or it could be interpreted eternally, as a preference for a scenario where I eat Vegemite at t over one where I eat peanut butter at t.

The rigidifying strategy looks unappealing if we interpret my preferences as preferences about what happens now (as opposed to preferences about what happens at time t). My earlier preference for peanut butter cannot possibly make the peanut butter good for me after my transformation, when my tastes have changed. The rigidifying strategy looks more appealing if

we interpret my preferences as preferences about what happens at *t*, so that what is good for me is to get peanut butter when I have peanut butter cravings, and Vegemite when I have Vegemite cravings.

But even on the more plausible way of thinking about objects of preference, there is still a problem. It is not at all clear which preferences to privilege. Perhaps what is good for a person is whatever satisfies the person's last held preferences—so that when after a long and varied life, I settle on a value system that favors artistic achievement, it turns out to be artistic achievement that was good for me all along (even when I was a successful businessperson). But this seems arbitrary. Besides, should it really be possible to improve one's life drastically by a Pollyannaish deathbed conversion? Or perhaps what is good for a person is whatever satisfies the person's first preferences—but this seems to yield wrong results about people who, as adults, come to love things that their childhood selves would have despised, such as Brussels sprouts or kissing.

When it comes to solving the problem of interpersonal utility comparisons, rigidifying is incompatible with preference satisfaction theories. However, the rigidifying strategy helps the preference satisfaction theorist address the problem of intrapersonal utility comparisons. Rigidifying allows the preference satisfaction theorist to give intuitively correct answers in a variety of cases involving transformative choice. In cases involving transformative experience, the rigidifying strategy is less promising, since it requires us to arbitrarily favor an agent's preferences at one time over her preferences at all other times.

5 Conclusion

The problem of interpersonal utility comparisons has intrapersonal analogues in cases of transformative experience.

In cases where I expect to undergo a transformative experience, unit comparisons are particularly important: in order to assess whether an option is good or bad for me, I need to compare the difference it makes to my life before the transformation with the difference it makes to my life afterward. And in cases where I face a transformative choice, level comparisons are particularly important. In order to assess whether undergoing a transformation will make me better off, I must compare how well my unchanged preferences are satisfied, in the closest world where they are unchanged, with how well my transformed preferences are satisfied, in the closest world where they are transformed. If I cannot directly choose to be transformed, but can only raise my probability of transformation, then I must also make unit comparisons to do decision theory.

The problem of interpersonal utility comparisons admits of various solutions: I might do without the comparisons; make the comparisons using the Zero-One Rule; posit primitive facts about interpersonal comparisons; or choose a single privileged set of preferences and rigidly index wellbeing

to it. Attempts to extend these solutions to the problem of intrapersonal utility comparisons sheds new light on them. It is even less attractive to do without intrapersonal utility comparisons in cases of transformative experience, than it is to do without interpersonal utility comparisons. Likewise, trying to sidestep the comparisons by rigidifying looks different—and has different drawbacks—in the interpersonal and intrapersonal cases. The view that interpersonal utility comparisons are value judgments looks much less appealing when applied to intrapersonal utility comparisons, as do objections to the Zero-One Rule based on its unfairness, rather than on its descriptive failures. On the other hand, solutions to the problem that make interpersonal utility comparisons a matter of descriptive fact—appeal to primitive comparisons, or the Zero-One Rule suitably interpreted—seem to extend naturally to intrapersonal comparisons.

Though the solutions play out differently for the intrapersonal and interpersonal puzzles, preference satisfaction theorists have a range of promising solutions to the puzzle of how to assign degrees of wellbeing to people who undergo transformative experiences. Solving my problems about wellbeing, however, is not sufficient to solve Paul's problem about rationality. I have been concerned with evaluative questions: even given full information about the outcome of a transformative experience, it is hard to say whether the outcome is good or bad. Many of Paul's worries are epistemic: it is very difficult (perhaps impossible) to get full information about the outcome of a transformative experience. Decision theorists have their work cut out for them, and a theory of intrapersonal utility comparisons is only part of the task.

Rachael Briggs
E-mail : rachael.briggs@anu.edu.au

References:

Arrow, Kenneth. 1963. *Social Choice and Individual Values*. New York, NY: John Wiley & Sons, Inc.

Bradley, Richard. 2009. "Becker's thesis and three models of preference change." *Politics, Philosophy & Economics* 8 (2): 223–242. http://dx.doi.org/10.1177/1470594X09102238.

Goldman, Alvin I. 1995. "Simulation and Interpersonal Utility." *Ethics* 105 (4): 709–726. http://dx.doi.org/10.1086/293749.

Griffin, James. 1987. *Well-Being: Its Meaning, Measurement, and Moral Importance*. Oxford: Oxford University Press.

Griffin, James. 1993. "Against the Taste Model." In *Interpersonal Comparisons of Well-Being*, 45–69. Cambridge: Cambridge University Press.

Hammond, Peter. 1993. "Interpersonal Comparisons of Utility: Why and How They Are and Should Be Made." In *Interpersonal Comparisons of Well-Being*, edited by Jon Elster and John E. Roemer, 200–254. Cambridge: Cambridge University Press.

Harsanyi, John C. 1955. "Cardinal Welfare, Individualistic Ethics, and Interpersonal Comparisons of Utility." *Journal of Political Economy* 63 (4): 309–321. http://dx.doi.org/10.1086/257678.

Harsanyi, John C. 1977. "Morality and the Theory of Rational Behavior." *Social Research* 44: 625–656.

Hausman, Daniel M. 1995. "The Impossibility of Interpersonal Utility Comparisons." *Mind* 104 (415): 473–490. http://dx.doi.org/10.1093/mind/104.415.473.
Hausman, Daniel M. 2011. *Preference, Value, Choice, and Welfare*. New York, NY: Cambridge University Press.
List, Christian. 2003. "Are Interpersonal Comparisons of Utility Indeterminate?" *Erkenntnis* 58 (2): 229–260. http://dx.doi.org/10.1023/A:1022094826922.
MacKay, Alfred F. 1986. "Extended Sympathy and Interpersonal Utility Comparisons." *The Journal of Philosophy* 83 (6): 305–322. http://dx.doi.org/10.2307/2026093.
Nagel, Thomas. 1986. *The View from Nowhere*. Oxford: Oxford University Press.
Neumann, J. Von and Otta Morgenstern. 1953. *Theory of Games and Economic Behavior*. 3rd edn. Princeton, NJ: Princeton University Press.
Parfit, Derek. 1984. *Reasons and Persons*. Oxford: Oxford University Press.
Paul, L. A. 2014. *Transformative Experience*. Oxford: Oxford University Press.
Rawls, John. 1971. *A Theory of Justice*. Cambridge, MA: Harvard University Press.
Raz, Joseph. 1988. *The Morality of Freedom*. Oxford: Oxford University Press.
Robbins, Lionel. 1938. "Interpersonal Comparisons of Utility: A Comment." *The Economic Journal* 48 (192): 635–641. http://dx.doi.org/10.2307/2225051.
Scanlon, T. M. 1993. "The Moral Basis of Interpersonal Comparisons." In *Interpersonal Comparisons of Well-Being*, edited by Jon Elster and John E. Roemer, 17–44. Cambridge University Press.
Sen, Amartya. 1970a. *Collective Choice and Social Welfare*. San Francisco, CA: Holden-Day.
Sen, Amartya. 1970b. "The Impossibility of a Paretian Liberal." *The Journal of Political Economy* 78 (1): 152–157. http://dx.doi.org/10.1086/259614.
Weintraub, Ruth. 1998. "Do Utility Comparisons Pose a Problem?" *Philosophical Studies* 92 (3): 307–319. http://dx.doi.org/10.1023/A:1004203702557.

Epistemic Expansions

Jennifer Carr

Abstract: Epistemic transformations—changes in one's space of entertainable possibilities—are sometimes rational, sometimes irrational. Epistemology should take seriously the possibility of rationally evaluable epistemic transformations. Epistemic decision theory compares belief states in terms of epistemic value. But it's standardly restricted to belief states that don't differ in their conceptual resources. I argue that epistemic decision theory should be expanded to make belief states with differing conceptual resources comparable. I characterize some possible constraints on epistemic utility functions. Traditionally, it's been assumed that the epistemic utility of a total belief state determines the epistemic utility of individual (partial) beliefs in a simple, intuitive way. Naive generalizations of extant accounts generate a kind of repugnant conclusion. I characterize some possible alternatives, reflecting different epistemic norms.

I've never had a child. I've never tasted an oyster. I've never experienced war. I don't have the slightest idea what any of these experiences is like. I can't even entertain the possibilities for what they are like. How can I make rational decisions about whether to have a child, taste an oyster, or go to war, when I have so little idea what kind of outcome I'd generate? Normally I could decide on the basis of thinking about the possible outcomes of my actions: how valuable each would be, and how likely it is. But I can't do that.

This is L. A. Paul's (2014; 2015) challenge: what could rationalize a decision about whether to perform any of these actions? How can there be a decision theory for partial credence functions, when decisions hinge on possibilities the agent can't entertain? The problem is not uncertainty: it's not simply that the agent is unsure of the outcomes of her actions. Rather, the problem is limited conceptual resources: there are some possibilities that the agent can't "see," propositions she isn't in a position to entertain.

We can formulate the puzzle, then, in overtly decision theoretic terms. Suppose an agent has only a partial credence function, one that doesn't range over some of the possible outcomes of an action available to her. Then

Res Philosophica, *Vol. 92, No. 2, April 2015, pp. 217–236*
http://dx.doi.org/10.11612/resphil.2015.92.2.4

the expected utility of that action is undefined. So what could rationalize the choice of whether to perform it?

An epistemic analogue to Paul's challenge: how can there be an epistemic decision theory for an agent with a partial credence function, when her epistemic "decisions" hinge on possibilities the agent can't entertain?[1]

Suppose I'm deciding whether to ϕ, where ϕing involves changing my credences. But I don't know the epistemic utility of ϕing. Its epistemic utility depends on what state of the world I'm in: in particular, on which credence function ϕing will lead me to adopt. What's worse: I can't even entertain these credence functions, because they involve concepts that I don't possess. They distinguish among possibilities I can't distinguish. So how am I supposed to choose whether to ϕ?

The epistemologized version of Paul's challenge, then, is: can we extend epistemic decision theory for agents with partial credence functions? For agents whose credences can change domain, so that the agent can come to see different possibilities? I'll argue that we can. Moreover, we should: it's sometimes irrational to change the domain of one's credence function.

The plan of the paper is as follows: I argue that epistemic decision theory should be expanded such that credence functions with different domains are sometimes comparable. The argument is based on a weak conservative principle: that it's sometimes irrational to lose conceptual resources. Then I characterize some possible constraints on epistemic utility functions that compare credences with different domains. Traditionally, the epistemic utility of a total credence function is understood as a function of the epistemic utility of credences in individual propositions. The most natural ways of generalizing to partial credences generates a kind of repugnant conclusion.[2] I argue for a general constraint on the space of possible algebra-neutral epistemic utility functions. I characterize some possible versions, reflecting different epistemic theories.

1 Paul's Challenge Epistemologized

1.1 Partial Credence Functions

Jackson (1982) argued that before ever seeing the color red, there was something Mary failed to know: what it's like to see red. Jackson's thought experiment arguably supports a stronger conclusion: Mary can't even conceive of what it's like to see red. Of course, she can entertain the possibility that what it's like to see red is the same as what it's like to see dark gray, or the same as what it's like to taste an oyster. But there are many possibilities for what it will be like to see red that Mary can't entertain, including what it's actually like.

[1] To highlight the analogy with practical decision theory, I use voluntaristic phrasing ("decisions," "options," etc.). But epistemic decision theory doesn't presuppose epistemic voluntarism.

[2] For the familiar repugnant conclusion argument against utilitarianism, see Parfit 1984.

Like Paul, we begin with the assumption: people don't always have attitudes toward all propositions. Within the partial belief framework: sometimes a person's credence function isn't defined over all propositions.

This is already controversial. On the traditional view of the psychology of credences, both credence and utility functions are understood as abstractions from dispositions to choice behavior. There are representation theorems that show that, if an agent's dispositions satisfy some constraints, the agent is describable by a unique credence function and a utility function unique up to positive affine transformation. The traditionalist may insist that credence functions aren't partial. Or, if she allows that they may be partial, the traditionalist may insist that at least credence functions are not partial in any way that could change over time. Credences are automatically defined for all possibilities the agent could possibly encounter, since for each encounterable possibility, the agent will have choice dispositions.[3]

We might respond by simply accepting that the dispositionalist psychology of credences is wrong. But there are also possible replies to this objection that maintain the spirit of the dispositionalist view. First, we might question whether an agent really has the relevant dispositions to choice behavior even for choices that hinge on propositions she can't entertain. It might be that, instead, agents have dispositions to acquire dispositions to choose once the relevant conceptual resources are active.[4]

Second, we might accept that credences are dispositions to choice behavior that meet certain necessary conditions. Plausible theories of intentionality impose some conditions on propositional attitudes. These conditions may be externalist: in order to have *de re* thoughts about individuals, it might be that you need to have a particular sort of causal connection with those individuals. (Shakespeare couldn't have had *de re* thoughts about Cher.) They may also be internalist: in order to have certain kinds of thoughts about phenomenal redness, it might be that you need to have already experienced phenomenal redness.

A different kind of objection: decision theory characterizes agents who are idealized in all kinds of ways. For example, their credences are infinitely non-vague. Traditional decision theory presupposes credences that are infinitely precise. Fans of imprecise credences functions assume credences take sets of reals as values, and so imprecise credences have infinitely precise boundaries. And so on. Ideally rational agents aren't computationally limited. So why allow that they can have merely partial credence functions?

Reply: if there are external or experiential conditions on having propositional attitudes, then plausibly ideal rationality is compatible with having partial credence functions. Ideal rationality doesn't require us to be in a certain sort of environment, or to have had certain kinds of perceptual experiences. Mary is conceptually limited not because of some irrationality

[3] The traditionalist assumes that agents' dispositions are determinate, though not fixed.

[4] Thanks to Robbie Williams for this suggestion.

in her belief state, but because a certain kind of experiential state is a necessary precondition for entertaining what it's like to see red. Rationality doesn't require one to have already experienced red phenomenology.

In short, the presupposition that rational agents have total credence functions is not an intuitive assumption about ideal rationality. It's an idealization only in the sense of "simplification." When this simplification is removed, it opens up substantive questions about rational decision making and rational update.[5]

I'll assume that even partial credence functions are defined over boolean algebras of propositions, the strongest elements of which form a partition over the set of worlds. In other words, I assume there's an exhaustive set of mutually exclusive 'basic possibilities' that are "visible" to an agent; the agent's credences are defined over all unions of visible basic possibilities plus the empty set. What justifies this assumption? Plausibly, while ideal rationality doesn't require being able to see all propositions, it does require being able to negate and conjoin the propositions you do see.

The partition of basic possibilities an agent can see characterizes the distinctions the agent is able to make. As an intuitive shorthand, I'll talk about an agent's "concepts" or "conceptual resources." Note: I don't mean to make any substantive commitments about the psychology of concepts. Indeed, nothing in the way we're modeling things will play the role of "a concept," qua subpropositional mental representation. The distinctions in the space of possibilities that an agent can make are presumably related to her conceptual resources, but I acknowledge that that relation might be very messy. If the shorthand seems misleading or distracting, talk of concepts may be translated into talk of distinctions an agent can make in logical space.

Once we notice that rational agents can't always see all propositions, Paul's challenge arises. How can we make rational choices when we aren't in a position to entertain the possible outcomes of our actions?[6]

1.2 Epistemic Decision Theory

Paul's challenge is a problem for practical decision theory. The epistemologized variant we're considering is a problem for epistemic decision theory.

[5] Other objections to Paul's challenge relate to whether partial credences generate a distinctive problem for traditional decision theory. See, for example, Collins 2015, Dougherty et al. 2015, Harman 2015, and Sharadin 2015.

[6] Here, it might be that I'm formulating the challenge differently from how Paul (2015) sees it. On her view, the specification of outcomes won't determine facts about the agent's phenomenology. So an agent whose credences aren't defined over relevant phenomenological propositions will nevertheless be able to entertain the possible outcomes of her acts. But the agent's inability to conceive of the phenomenology of, e.g., having a child will prevent her from assigning utilities to outcomes in which she has a child. This difference may be substantive. Indeed, it may be that by shifting the focus away from phenomenology, what I call "Paul's challenge" would be better called "one of Paul's challenges."

Epistemic decision theory is an application of a Savage-style decision theory, restricting itself to epistemic "acts" and epistemic utilities. Epistemic "acts," typically not construed voluntaristically, involve possessing or coming to possess a credence function. Epistemic utility functions represent comparative epistemic goodness. Commonly defended epistemic decision rules:

Dominance: if credence function c has higher epistemic utility than credence function c' at every world, don't adopt c'.

Maximize Expected Epistemic Utility: adopt the credence function with the highest expected epistemic utility (i.e. the highest weighted average of epistemic utilities at all possible worlds, weighted by the probability of those worlds).

As with ordinary decision theory, the relevant sort of goodness is ultimate epistemic goodness. Epistemic goodness of this sort is objective, in the sense that it's non-information-dependent. A common view is that the relevant sort of ultimate epistemic goodness should be interpreted in terms of gradational accuracy (Rosenkrantz 1981; Joyce 1998, 2009; Leitgeb and Pettigrew 2010a,b). Gradational accuracy is the closeness of a credence function to the truth, by some measure satisfying a handful of intuitive constraints. Credence 1 in p is maximally close to the truth iff p is true; credence 0 is maximally close to the truth if p is false.

We can distinguish the epistemic utility of a particular credence an agent has in an individual proposition from the epistemic utility of the agent's total belief state. We'll call the former 'local epistemic utility' and the latter 'global epistemic utility.' It's usually assumed that global epistemic utility is determined straightforwardly as a function of local utilities.

Turning back to Paul's challenge: can epistemic decision theory be extended to credence functions defined over different propositions?

2 Epistemic Decision Theory for Partial Credences

2.1 Motivating Comparability

Epistemic decision theory usually presupposes that the credence functions it compares are defined over the same algebra of propositions. Once we abandon this presupposition, new difficulties arise.

For example, to narrow the space of epistemic utility functions, constraints such as 'truth-directedness' and 'immodesty' are placed on epistemic utility functions. Roughly, truth-directedness says credence functions are epistemically better the closer they are to the truth. Immodesty says that probabilistic credence functions assign themselves higher expected epistemic utility than all other credence functions. It's no longer clear what these constraints amount to, or why they are intuitive, once we compare credence functions that are defined over different domains. Should a probabilistic

partial credence function assign itself higher expected utility than probabilistic extensions of itself that see more propositions? Is the extension automatically closer to the truth than the original?[7]

It's tempting to prescind from these questions. Epistemic decision theory wasn't designed to make comparative evaluations of credence functions that see different propositions. It falls silent about these sorts of comparison. So perhaps credence functions with different domains are incomparable in epistemic value.

Here are two arguments for why credence functions that can see different propositions should be comparable.

First: decision theory aims to be neutral with respect to substantive normative questions. It provides only structural constraints on rational choices. Epistemic decision theory involves placing some substantive constraints on epistemic utility functions, but these are only meant to delineate the epistemic subject matter and need to be individually justified. There's no intuitive basis for assuming that credence functions with different domains can't be compared. So epistemic utility theory should accommodate the possibility that conceptual change has epistemic (dis)value.

Second: there are intuitive grounds for making at least some credence functions with different domains comparable. Here's an argument from evidentialism: it's irrational to change your credence in a proposition without new evidence. One way of changing your credence in a proposition is to abandon it, so that your credence function no longer sees the proposition. So when evidence is held fixed, it's irrational to abandon your credence in a proposition. More briefly: it's irrational to undergo an "epistemic contraction." By contrast, it's plausible that it's at least sometimes rational to undergo an "epistemic expansion," whereby you retain your previous credences but come to see new propositions. (Intuitions are murkier about cases where you come to see new propositions and lose sight of old ones,

[7] Pérez Carballo (Unpublished) provides the only other sustained discussion of epistemic utility functions for comparing partial credence functions defined over different algebras. His focus is on assessing avenues of inquiry for epistemic utility: which questions are most fruitful to ask? Pérez Carballo defends a series of constraints on algebra-neutral epistemic utility functions: first, he defends weak generalizations of traditional constraints on epistemic utility functions: partition-wise truth-directedness and partition-wise propriety. Pérez Carballo offers compelling arguments against stronger generalizations of these principles (appendix 2). Second, algebra-neutral epistemic utility functions must be "nice", i.e., they must assign partial credence functions uniform utility at worlds the credence functions don't distinguish. Third, he argues for a third constraint, "resilience," according to which, if two credence functions are on a par with respect to truth-directedness and propriety, then epistemic utility functions should be such that whichever has greater expected explanatory potential (understood as counterfactual resilience) will also have greater expected epistemic utility. I take no stance on whether the resilience constraint is correct. This paper concerns separate constraints on epistemic utility functions, motivated by general questions about the value of conceptual resources. Unlike Pérez Carballo, my discussion abstracts away from computational limitations that might make rich conceptual resources costly, in order to focus on ideal epistemic theory.

and so I focus narrowly on easy cases: pure epistemic expansions and contractions.)

Let's consider stronger and weaker versions of these claims:

(1) It's always rationally impermissible to undergo an epistemic contraction.
(2) It's sometimes rationally impermissible to undergo an epistemic contraction.
(3) It's always rationally permissible to undergo an epistemic expansion.
(4) It's sometimes rationally permissible to undergo an epistemic expansion.

Of these four claims, all that is needed to establish the possibility of cross-algebra comparison is for one of the weaker claims, claim 2 and claim 4, to be true. So here I rely on intuition. There are at least some circumstances where losing a credence in a proposition is irrational. In order for it to be possible to model the comparison between epistemic states before and after an epistemic expansion or contraction, credence functions with different domains must at least sometimes be comparable.

2.2 Strong and Weak Conceptual Conservatism

For the purposes of this paper, I'll commit to claim 2 and claim 4. But in fact, all four have some intuitive plausibility, at least when the space of options is unconstrained. (When the space of options is constrained, there can be cases of forced choices between irrational credences and epistemic contractions, or cases where all the only optional expansions are irrational. In such cases, contractions may be permissible and expansions impermissible, contra claim 1 and claim 3. Note also that my discussion is confined to ideal rationality, where clutter avoidance and other resource constraints are non-issues.)

Consider claim 1, which we can call 'strong conceptual conservatism.' A brief defense of strong conceptual conservatism:

Strong evidentialism: It is irrational to change your credences without acquiring new evidence.

No evidence against concepts: There can be no evidence that justifies losing conceptual resources.

Together these entail strong conceptual conservatism.

One might object: aren't there good reasons to abandon some concepts? For example, concepts that have false presuppositions? For example, once I realize that the concept *slut* has false misogynist presuppositions, isn't it best to abandon that concept altogether? In fact, though, it's better to retain the concept.

One might argue that the real objection to having the concept *slut* is a practical, ethical objection, not an epistemic one. Practical norms have no

bearing on epistemic norms. (We are not epistemically required to have nonprobabilistic credences even when we know that doing so will magically save the life of a child.)

A better argument for retaining even problematic concepts, on my view, is that doing so is preferable both epistemically and practically. What's really wrong with objectionable concepts is not possessing them, but rather (in some sense) applying them. Suppose you know that Joe believes that Mary is a slut. If you abandon the concept of a *slut*, then you will no longer know exactly what Joe believes about Mary. If you want to talk him out of this misogynist belief, better to keep the concept *slut* and explain to him what's wrong with applying it.

Nothing in the present framework models a fully general distinction between possessing and applying a concept, but the idea should be roughly clear. Refraining from applying a concept might mean rejecting (having credence 0) in all propositions that apply the concept. (One might wonder: how is that consistent with Bayesianism? Wouldn't you have to have credence 0 in both a proposition and its negation, thereby violating probabilism? Response: to make sense of concepts with false presuppositions in this sense, we'd have to move to a nonclassical setting.)[8]

So: problematic concepts don't generate a problem for strong conceptual conservatism. But there are other possible different angles of attack.

First, it might be that in some cases, if you face a choice between having credences that employ a problematic concept in an irrational or otherwise problematic way, or else losing the concept altogether, then perhaps you should choose the latter. For example, suppose it's psychologically impossible for you to retain the concept *slut* without lending positive credence to propositions that apply it. Then perhaps it's better just to lose the concept.

Here I think, the fan of strong conceptual conservatism will reply: if your credences are already irrational, then rationality doesn't recommend retaining them. Contracting your conceptual resources is still irrational, but it might be the lesser of two epistemic evils. What this objection is really an objection to is the stronger claim that in all cases, it's better to retain one's own credence function than to adopt any contraction of it. That is

[8] Robbie Williams has pointed out to me that, in the case of concepts with false presuppositions, the problem may be more complex than I make out. On some views, each concept with a false presupposition determines a concept with the same application conditions that doesn't trigger the false presupposition; and so one can abandon the false presupposition while retaining the same distinctions in logical space by switching to the equivalent, presuppositionally innocuous concept. This suffices for conceptual conservation in the sense of "conceptual" I use throughout. But in the case of thick concepts like *slut*, this form of disentanglement may not be possible, and so other resources (e.g. rational but nonprobabilistic credences) may be necessary. With still other forms of problematic concept, we may have to tell a different story. Still, for the purposes of representing others' beliefs, or merely possible beliefs, or counterfactual or counterpossible scenarios, I suggest, strong conceptual conservatism is compelling.

plausibly false, but not something that the strong conceptual conservative is committed to.[9]

A different type of objection to strong conceptual conservatism suggests that sometimes evidential considerations favor loss of concepts. For example, it might be that as our knowledge of the world develops, our concepts change; a naive concept is replaced by a more sophisticated one, or two separate concepts merge when they're revealed to be extensionally equivalent or analytically identical.

But these sorts of objections rely on a more inflated theory of conceptual resources than capacities to draw distinctions in logical space. Extensionally equivalent and analytically identical concepts don't draw different distinctions in logical space. Concept change over time may involve getting rid of concepts in some psychological sense, but it's not clear that it rationally requires abandoning distinctions in logical space.

So these observations are in fact compatible with conceptual conservatism, in my (loose, perhaps unfortunate) sense of the word "conceptual." There are, of course, interesting questions about the epistemic value of concepts in a more inflationary sense; but that's not under discussion here.

A final objection to strong conceptual conservatism: it prohibits ever losing conceptual resources. This is a diachronic epistemic constraint. It's controversial whether there are diachronic constraints on rationality.[10]

Opponents of diachronic constraints on rationality reject strong evidentialism, in favor of a weaker variant:

> **Weak evidentialism:** It's irrational to change your credences without some change in evidence.

Weak evidentialism allows changes in your credences if you either gain new evidence or lose old evidence, e.g. by forgetting information. On its own, weak evidentialism entails claim 2, which we can call 'weak conceptual conservatism.' Indeed, it entails something stronger:

> 2′ 'Medium-strength conceptual conservatism': it's rationally impermissible to undergo an epistemic contraction without some change in evidence.

We can equally provide arguments from claim 4 to comparability. For example, suppose c is a partial credence function, and c^+ is an extension of c such that for all propositions A visible to c^+ and not c, $c^+(A)$ is maximally accurate (or otherwise has maximal epistemic value). Intuitively, c^+ must be at least as accurate as (or as valuable as) c. And so c and c^+ must be comparable.[11]

The upshot: evidentialism supports the conclusion that at least some credence functions with different domains can be compared.

---

[9] Thanks to L. A. Paul, Julia Staffel, and Robbie Williams for discussion.

[10] Against diachronic rationality, see Talbott 1991, Christensen 2000, Williamson 2000, Meacham 2010, Moss Unpublished, Hedden 2013.

[11] Thanks to an anonymous referee for suggesting this compelling intuition pump.

2.3 Impact on Epistemic Decision Theory

What this means for epistemic decision theory: we need epistemic utilities, pr epistemic decision rules, that make at least some credence functions defined over different domains comparable.

One natural temptation would be to be as neutral as possible with respect to the epistemic utilities of partial credence functions: for example, to assign partial credence functions only imprecise epistemic utilities. A neutral imprecise utility assignment for a partial credence function c would equal the set of global utilities of all total extensions of c.

How do we compare imprecise utilities? Perhaps the imprecise utility of c will be greater than the imprecise utility of c' iff the precise utilities of all total extensions of c were greater than the precise utilities of all total extensions of c'. The problem is that this generates widespread incomparability. In particular, partial credences will always be incomparable with all of their extensions. The imprecise utility of a credence function will be a superset of the imprecise utility of its extensions.[12]

We may be happy not to compare the utilities of credence functions defined over disjoint sets of propositions, or overlapping sets of propositions where neither includes the other. But comparing credence functions and their extensions was supposed to be the easy case. An algebra-neutral epistemic utility function should be able to compare at least some partial credence functions and their extensions. Otherwise we can't predict conceptual conservatism.

It would be hasty to rule out imprecise utilities altogether. But I've argued that epistemic decision theory should allow at least some credence functions over different domains to be comparable, including at least some credence functions and their extensions. So either way, we can't avoid substantive epistemological questions about what constraints there are on epistemic utility functions that range over credence functions with different domains.

3 Epistemic Utility Functions for Partial Credences

Instead of retaining a utility function that ranges only over total credence functions, and assigning partial credence functions imprecise utilities, I suggest we consider epistemic utility functions that range over (at least some) credence functions with different domains.

[12] Obviously, we might use a different rule for comparing imprecise credences: for example, perhaps c is strictly preferable to c' iff the maximal utility in c is greater than the maximal utility in c' and the minimal utility in c is greater than the minimal utility in c'. Then we might generalize: c is weakly preferable to c iff the maximum utility of c is greater than or equal to the maximum utility of c' and the minimum utility of c is greater than or equal to the minimum utility of c'. Even then, we can never predict that an extension of c is strictly preferable to c, and so we can't predict weak conceptual conservatism.

Which utility functions? Instead of positing unique epistemic utility functions, epistemic utility theory generally proceeds by imposing constraints on the space of candidate epistemic utility functions, constraints which suffice represent epistemic norms of various sorts. We can think of principles like strong and weak conceptual conservatism as constraints on the space of epistemic credence functions, constraints that encode the epistemic value of conceptual resources. There may be other intuitive principles linking the utilities of partial credence functions and their extensions.

Other constraints on algebra-neutral epistemic utility functions may be natural generalizations of accepted constraints on algebra-specific functions. For example, in place of strict propriety (the constraint that algebra-specific utility functions should make probabilistic credence functions assign themselves higher expected utility than all alternatives), an algebra-neutral epistemic utility function should perhaps satisfy a generalization of strict propriety that makes probabilistic partial credence functions assign themselves higher expected utility than all alternatives defined over the same partition.

3.1 Local and Global Utilities

In order to extend epistemic decision theory to partial credence functions, we need to look at how adding or subtracting individual credences affects the epistemic utility of an agent's total epistemic state. In other words, we need to look at the relationship between the local utilities of credences in individual propositions and the global utilities of overall credence functions.

Global epistemic utility is usually interpreted (noncommittally) as a sum or average of local epistemic utilities. Where g and l are global and local utility functions, respectively, and $\mathscr{A}$ is the algebra of propositions c is defined over,

Summing proposal:

$$g(c,w) = \sum_{A \in \mathscr{A}} l(c,A,w)$$

Averaging proposal:

$$g(c,w) = \frac{1}{|\mathscr{A}|} \sum_{A \in \mathscr{A}} l(c,A,w)$$

Now, epistemic utility is standardly understood as gradational accuracy. Accuracy-first epistemic decision theory assigns credences value in terms of their distance from the truth, by some measure satisfying a handful of intuitively plausible constraints.

So we can ask: is an accuracy-first epistemology compatible with treating conceptual resources as epistemically valuable? The answer can be *yes* only if the epistemic value of conceptual resources is reducible to the epistemic value of accuracy. As it turns out, both the summing hypothesis and the

averaging hypothesis for global (in)accuracy introduce commitments about the value of conceptual resources. These commitments potentially are unattractive.

We could, of course, switch to epistemic utility functions that aren't accuracy-directed. Indeed, we might be forced to do so in order to avoid implausible conclusions about the value of conceptual resources. But it's worth exploring whether we can represent the value of conceptual resources without departing to far from accuracy-first epistemology, especially because the research project has otherwise proven fruitful.

3.2 Summing Positive Disutilities

In accuracy-first epistemic decision theory, local inaccuracy is interpreted as a positive penalty for distance from the truth. There is a maximum degree of accuracy, which is distance 0 from the truth. Any credence other than 1 in truths or 0 in falsehoods incurs a positive disutility.

And so summing local disutilities generates an immediate consequence: if c^+ is an extension of c and any of the new credences it brings are uncertain, c^+ will incur whatever disutilities c has plus disutilities for its new credences. So c^+ will automatically have greater global disutility c. In other words, c dominates c^+, merely because c^+ can make new distinctions but isn't omniscient about them.

So on this proposal, we find ourselves with the result that the fewer imperfect credences you have, the better you are epistemically. This amounts to treating nonattitudes toward propositions as epistemically *perfect*: maximally accurate. They are matched only by the epistemic utility of credence 1 in truths and credence 0 in falsehoods.

This proposal, paired with either weak dominance avoidance or expected utility maximization, yields the following verdicts: rational agents only have credences in propositions such that their credence matches the truth value at every world (i.e. $\top$ and $\bot$ and any other known necessities). Specifically, rational agents will have credence 1 in all tautologies, credence 0 in all contradictions, and no other defined credences. And of course, similar problems afflict any other, non-accuracy-based interpretations of global epistemic disutility that treat global disutilities as sums of positive disutilities.

I take it as a datum that epistemic utility functions shouldn't universally prohibit attitudes toward contingent propositions, or propositions the agent can't be certain of.

3.3 Averaging

What if instead of treating the global (dis)utility of a credence function as the sum of its local (dis)utilities, we treated it as the average?

Then there is no automatic dominance relation between credence functions and all of their extensions. Sometimes seeing new propositions will

increase global utility, sometimes decrease it. At some worlds, new uncertain credences may increase your average accuracy; at others decrease it.

The averaging proposal seems like an obvious move. But it brings with it new problems. Suppose you have credence function c, which sees propositions A and $\neg A$ and assigns .8 in A, which is true. Then suppose you have the option of extending your credence function to c^+, which also sees new propositions B and $\neg B$, and has credence .6 in the true one.[13]

Two consequences. First, adopting c^+ entails a reduction in your global accuracy at the actual world. Your old credences in A and $\neg A$ were pretty accurate. Your new credences in B and $\neg B$ are on the right track, but they're still not as close to the truth as your old credences were. So they drag the average down.

Second, adopting c^+ will typically entail a reduction in your expected global accuracy. After all, you are more confident than not that you're at a world where your average accuracy is dragged down.[14]

Two objections. First, it's not clear that this is an intuitive way of characterizing the overall accuracy of a credence function. Second, if we accept this as a characterization of the accuracy of a credence function, then accuracy is a bad measure of epistemic value.

On the first point: it's clear that your average local accuracy is reduced when you move from c to c^+. But it's not obvious that your global accuracy should be reduced. There is some intuitive sense in which c^+ is doing better at the actual world, accuracywise, than c. For one thing, c couldn't distinguish B from $\neg B$. c^+ not only distinguishes them, but is closer to truth than to falsehood. So this example may motivate rejecting the equivalence of global accuracy with average local accuracy.

On the second, less controversial point: if global accuracy is average local accuracy, then it's implausible that global accuracy is a good characterization of epistemic value.

In this example, suppose our agent is an expected utility maximizer and epistemic utility is simply average accuracy, measured by the most familiar accuracy measure, the Brier score.[15] Then if she has c^+ as her credence, she'll prefer to abandon her attitudes toward B and its negation. c has higher expected accuracy from c^+'s perspective than c^+ itself.

The problem doesn't just afflict expected utility maximizers. Even restricting ourselves to strong accuracy dominance avoidance, we end up at

[13] To keep the example as simple as possible, I don't stick to my assumption that credences must defined be over the boolean closure of a partition.

[14] This depends on the choice of local utility function or accuracy measure, since the loss of average accuracy at some worlds may be offset by a more dramatic gain in average accuracy at other worlds. But for any continuous, truth-directed accuracy measure, we can generate an analogous case where the expected average accuracy will decrease with the addition of new credences.

[15] The Brier score of a probability is the squared Euclidean distance between the probability assigned to a proposition and its truth value (1 = true and 0 = false).

the bottom of the same slippery slope. If the agent has attitudes towards propositions that she's uncertain of, then her average accuracy is imperfect. But if she only has attitudes toward propositions that she's certain of—again, tautologies and contradictions—then her attitudes will have perfect average accuracy, and hence will be more accurate at every possible world. So in order to avoid having credences that are strongly dominated, she must restrict her credences to propositions that she's certain of. In order to avoid having credences that are weakly dominated, she must restrict her credences to propositions that are necessary across all possibilities that she can entertain.

And so, like the summing disutilities proposal, I think we have good reason to reject a algebra-neutral epistemic utility function that derives global utilities from averages of local utilities.

3.4 Summing Positive Utilities

Let's return to the summing proposal. What if instead of penalizing distance from truth, we reward distance from falsehood? That is, what if we derive global utility by summing positive local utilities? Then, of course, the situation is reversed with respect to summing positive disutilities. Each new proposition added to the domain of a credence function increases the credence function's epistemic value, as long as the credence it assigns isn't maximally inaccurate. This means treating nonattitudes toward propositions as maximally inaccurate—just as bad, from an epistemic point of view, as certainty of falsehood.

This is not counterintuitive in the way that summing disutilities is. There's a perfectly reasonable position in logical space according to which any increase in conceptual resources contributes positively to epistemic utility. It says that conceptual resources are a pure epistemic good: they trump any epistemic badness that might be required to achieve them. So it automatically gains us strong conceptual conservatism.

Whereas summing disutilities generated a dominance argument against having credences in any contingent propositions, summing positive utilities generates a dominance argument for having credences in every proposition. One might object: doesn't this contradict the premise of this paper: that there can be rational agents with partial credence functions? Doesn't this entail that before she sees red, Mary is irrational?

Answer: having dominated credences is only irrational if there are any non-dominated credences that are among one's epistemic options. In Mary's case, though, adopting credence functions that can conceive of what it's like to see red is not an option for her. She would need to experience phenomenal redness before she'd be in a position to adopt that credence function. But she can't: she's locked in a black room. So she's not irrational for having a credence function that's dominated only by a non-option, any more than you are irrational for not spontaneously acquiring a billion dollars.

It does mean, though, that if your credence function is needlessly partial, then you're irrational. If you simply fail to have credences in some possibilities, even though you're perfectly well in a position to do so, then this view judges you irrational. Two reasons to think this is not a problem: first, it's presumably controversial whether this is even psychologically possible. Second, if it is, it's not obvious that it should be rational.

A different kind of objection: suppose there were a way to make up new concepts. For example, you might give your friend Rachel a separate name, "Srachel," just for Tuesdays. Then you might conceptually distinguish Srachel (who went to karaoke on Tuesday) from Rachel (who gave a talk on Wednesday). Even though you know they're the same person, perhaps you can now conceive of the possibility that they're two different people.

The details of the example don't really matter; we simply need some way for a person to be able to generate new concepts. Then, the proposal seems to suggest that doing so is epistemically mandatory. That's implausible.

First, it's not clear that the act of voluntarily inventing a new concept is an epistemic act, as opposed to a practical one. Second, it's not clear that in introducing new concepts of this sort, you're really exposing any distinctions in logical space. (After all, you plausibly already have the concept of *rachel on tuesdays*; possessing that concept was instrumental in your invention of the *scrachel* concept. Third, if somehow you are exposing real distinctions in logical space, then it's perhaps doing so is epistemically preferable after all. The only obvious reason against it is clutter avoidance; but for ideally rational agents I take it that epistemic clutter is unproblematic.

Summing positive utilities does have some questionable consequences, though. For example, suppose a partial credence function c has an extension, c^+. Then c^+ dominates c, regardless of what credences it assigns to new propositions. c^+ could be irrational in lots of ways: it might assign credence 1 to both a proposition and its negation. Still, c^+ would dominate c, even if c were probabilistically unimpeachable.

This may not be a deep problem. If c^+ isn't probabilistic, then it will itself be dominated by some probabilistic credence function defined over the same algebra.[16] And so as long as an agent's epistemic options include all those credence functions defined over subsets of the propositions potentially visible to the agent, the summing utilities proposal will never require one to adopt nonprobabilistic credences.

Still: it's not obvious that seeing new propositions in an irrational way is necessarily epistemically better than not seeing them at all, even when both are suboptimal.

There are other reasons why one might not be satisfied with this proposal. Because it treats nonattitudes toward propositions as having the lowest

[16] Assuming our algebra-neutral utility functions preserve the probabilism-entailing properties of algebra-specific epistemic utility functions when comparing credence functions that share an algebra.

possible epistemic utility, expanding one's epistemic vision is an absolute epistemic good. One might doubt this: for example, one might think that having a very inaccurate credence in a proposition is worse, at a world, than not having any credence in that proposition.

I don't want to rule out the summing positive utilities proposal. It has many intuitively appealing features, including its preservation of conceptual conservatism. But these considerations suggest that it's worth exploring other options. There's room in first-order epistemology for controversy about whether new conceptual resources are always an absolute epistemic good, or whether their utility can be outweighed.

3.5 "Better than Chance"?

So far, a pattern has emerged. When we sum positive disutilities, in effect we treat nonattitudes toward propositions as if they had maximal epistemic utility. When we sum positive utilities, in effect we treat nonattitudes toward propositions as if they had maximal epistemic disutility. When we average local (dis)utilities, in effect we treat nonattitudes as if they had the same utility as the average utility of attitudes.

Instead of treating nonattitudes as though they all having maximal utility or minimal utility, it seems plausible that we should treat nonattitudes as having middling utility. The averaging proposal accomplishes that, but not in the right way.

One might be tempted to say: nonattitudes toward propositions are no closer to truth than to falsehood. And so they should have the same utility as credences that are no closer to truth than falsehood, i.e. credence .5.

Something like this might be on the right track. But suppose Mary has no attitude toward the propositions that seeing red is like experiencing phenomenal redness, or like experiencing phenomenal blueness, or like experiencing phenomenal azureness, or . . . . Suppose Mary's partial credence function has the same utility as an extension of itself that assigns credence .5 in *all* of these propositions. Then it has the same utility as a wildly non-probabilistic credence function. This will generate unpredictable consequences for when a very inaccurate credence function has higher utility than locally very accurate but partial credence function. In any case, this certainly makes partial credence functions dominated (since all non-probabilistic credence functions are dominated, assuming familiar constraints on epistemic utility functions). This is a substantive, non-obvious epistemological hypothesis.[17]

To avoid this problem, the local utility of nonattitudes need not be the same for both truth and falsehood. That does satisfy a plausible desideratum: that the global utility of a partial credence function is uniform across worlds it doesn't distinguish. But the proposal is stronger than necessary: one might satisfy the same desideratum without assuming that a

[17] Thanks to Kenny Easwaran for prompting me to consider this proposal and for discussion.

nonattitude's local utility is the same at worlds where it's true and worlds where it's false. Local utilities may differ as long as it all evens out at the global level. Let's consider another sort of proposal that generates uniformity in global utilities across indistinguishable worlds.

Consider again the example in the last subsection. A fan of accuracy-first epistemology might reason: even though c^+ brings down average local accuracy, it seems to be doing pretty well, accuracywise. After all, it assigns .6 credence in B, a truth; so it's closer to truth than falsehood. One might say, intuitively, c^+ is doing "better than chance." Maybe that's the sense in which it seems to be an improvement in accuracy over having a nonattitude.

And so, one might suppose nonattitudes toward propositions are worse than credences that are more accurate than chance, but better than credences that are less accurate than chance. The credences that are neither worse nor better than chance are the maximally unopinionated attitudes. So, on this view, nonattitudes have the same utility as maximally unopinionated attitudes.

Spelling this out: suppose each partial credence function c has a unique maximally unopinionated total extension: call this credence function $c^{:|}$. On the hypothesis we're considering, the utility of c's nonattitude toward A is equal to the utility of $c^{:|}$'s attitude toward A.

This hypothesis has a number of attractive features. Unlike summing disutilities or averaging, there is no automatic epistemic gain in seeing fewer propositions. There is also no automatic epistemic gain in seeing more propositions: this view allows that whatever epistemic good there is in seeing new propositions, it can be outweighed.

Another feature of this view is that it ensures weak conceptual conservatism, and comes close to ensuring strong conceptual conservatism. If c is probabilistic, then so is $c^{:|}$. And so if c's utility matches $c^{:|}$'s, then with standard (strictly proper) scoring rules, c will be non-dominated and will maximize expected utility from its own perspective. So it will never be rationally required to give up having any credence at all in some propositions. And, if the agent is an expected utility maximizer, she'll prefer her own partial credence function to almost any alternative the domain of which is a subset of hers.[18]

But this hypothesis does face standard symmetry worries. It depends on the existence of a unique, maximally unopinionated extension of c. But problems for the principle of indifference suggest that there isn't any objective basis for determining a unique maximally unopinionated $c^{:|}$.[19] In other words, there is no objective way of isolating a uniform distribution of probabilities c_0 over $\mathscr{W}$ such that $c^{:|}$ is (roughly) c_0 updated on c's evidence.

[18] It will, however, be epistemically permissible for an agent to give up credence in a proposition if her credence is already maximally unopinionated (and if giving up that credence doesn't violate the requirement that credences be defined over a boolean algebra).

[19] See van Fraassen 1989.

This is obviously a legitimate worry. And it isn't easy to avoid: for this style of proposal, we need maximal unopinionatedness to ensure that the global utility of c is uniform over worlds that c can't distinguish.[20] It is plausibly a desideratum of an algebra-neutral epistemic utility functions that if a partial credence function doesn't distinguish two worlds, then the credence function should have the same epistemic utility at both worlds.

If we like this proposal, we might have to accept that there's some arbitrariness or assessment-sensitivity in the assignment of epistemic utilities. The motivating thought behind this proposal: having a credence in a proposition is better (at a world) than having no attitude toward it if the credence assigned is "better than chance": if one's credence is closer to the truth than if one were to withhold judgment as much as possible. What amounts to doing "better than chance" depends here on a conception of chance as a uniform distribution over indistinguishable worlds, and so presupposes a specific space of worlds. The relevant space of worlds depends on the perspective of the theorist.

4 Conclusion

We've seen a variety of accuracy-based proposals for algebra-neutral epistemic utility functions. The most natural extensions of traditional epistemic utility functions—summing local disutilities and averaging local (dis)utilities—both generate terrible consequences. They require agents to give up any credence in propositions that the agent can't be certain of. And so they violate both strong and weak conceptual conservatism with respect to epistemically contingent propositions.

The other two proposals we've seen don't face this problem. Summing positive local utilities entails strong conceptual conservatism. Any loss of concepts will generate a loss of epistemic utility and so will be dominated. And the final proposal—where partial credence functions' utilities match that of their maximally unopinionated extensions—entails at least weak conceptual conservatism.

It's obvious that my discussion has been far from exhaustive. There may be other plausible ways of generating algebra-neutral epistemic utility functions that still make epistemic utility a function of accuracy. There are certainly other plausible epistemic utility functions that don't rely on

[20] This rules out a natural generalization. Objective Bayesianism—the view that there's a unique rational prior credence function—was originally envisioned as recommending absolute unopinionatedness, i.e. the uniform distribution. In the face of symmetry worries, contemporary objective Bayesians typically think that the rationally privileged prior need not be the uniform distribution over worlds. Suppose we said that the utility of a nonattitude toward a proposition at a world is equal to the utility of the credence assigned by the rationally privileged prior, updated on the partial credence holder's evidence. If the prior is not uniform, though, then it might assign different credences to worlds that the partial credence function under evaluation can't distinguish. And so its utility at those worlds might differ.

accuracy alone. It is not the ambition of this paper to narrow down the space of epistemic utility functions to one or the other of these proposals.

What I hope to have made clear, however, is the need for certain constraints on algebra-neutral epistemic utility functions. First, they must permit rational agents to have some uncertainty, and to have attitudes toward contingent propositions. Second, they must permit rational agents to have attitudes toward contingent propositions. Neither of the traditional ways of aggregating local inaccuracies have satisfied these constraints. Furthermore, there is some intuitive support for stronger constraints: perhaps weak conceptual conservatism; perhaps strong conceptual conservatism.

Jennifer Carr
E-mail : jenniferrosecarr@gmail.com

References:

Christensen, David. 2000. "Diachronic Coherence Versus Epistemic Impartiality." *Philosophical Review* 109 (3): 349–371. http://dx.doi.org/10.1215/00318108-109-3-349.

Collins, John. 2015. "Neophobia." *Res Philosophica* 92 (2): 283–300. http://dx.doi.org/10.11612/resphil.2015.92.2.6.

Dougherty, Tom, Sophie Horwitz, and Paulina Sliwa. 2015. "Expecting the Unexpected." *Res Philosophica* 92 (2): 301–321. http://dx.doi.org/10.11612/resphil.2015.92.2.5.

Harman, Elizabeth. 2015. "Transformative Experience and Reliance on Moral Testimony." *Res Philosophica* 92 (2): 323–339. http://dx.doi.org/10.11612/resphil.2015.92.2.8.

Hedden, Brian. 2013. "Options and Diachronic Tragedy." *Philosophy and Phenomenological Research* 87 (1): 1–29. http://dx.doi.org/10.1111/j.1933-1592.2011.00563.x.

Jackson, Frank. 1982. "Epiphenomenal Qualia." *The Philosophical Quarterly* 32: 127–136. http://dx.doi.org/10.2307/2960077.

Joyce, James M. 1998. "A Nonpragmatic Vindication of Probabilism." *Philosophy of Science* 65 (4): 575–603. http://dx.doi.org/10.1086/392661.

Joyce, James M. 2009. "Accuracy and Coherence: Prospects for an Alethic Epistemology of Partial Belief." In *Degrees of Belief*, edited by Franz Huber and Christoph Schmidt-Petri, Vol. 342, 263–297. Dordrecht: Springer.

Leitgeb, Hannes and Richard Pettigrew. 2010a. "An Objective Justification of Bayesianism I: Measuring Inaccuracy." *Philosophy of Science* 77 (2): 201–235. http://dx.doi.org/10.1086/651317.

Leitgeb, Hannes and Richard Pettigrew. 2010b. "An Objective Justification of Bayesianism II: The Consequences of Minimizing Inaccuracy." *Philosophy of Science* 77 (2): 236–272. http://dx.doi.org/10.1086/651318.

Meacham, Christopher J. G. 2010. "Unravelling the Tangled Web: Continuity, Internalism, Non-Uniqueness and Self-Locating Beliefs." *Oxford Studies in Epistemology* 3: 86–125.

Moss, Sarah. Unpublished. "Credal Dilemmas."

Parfit, Derek. 1984. *Reasons and Persons*. Oxford: Oxford University Press.

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

Paul, L. A. 2015. "What You Can't Expect When You're Expecting." *Res Philosophica* 92 (2): 149–170. http://dx.doi.org/10.11612/resphil.2015.92.2.1.

Pérez Carballo, Alejandro. Unpublished. "Good Questions."

Acknowledgements For comments on earlier versions of this paper, many thanks to Kenny Easwaran, Laurie Paul, Paolo Santorio, Robbie Williams, an anonymous referee, and audience members at the Res Philosophica Conference on Transformative Experience at Saint Louis University, as well as audiences at the University of Leeds, the University of Birmingham, and the Eastern APA (2014).

Rosenkrantz, Roger. 1981. *Foundations and Applications of Inductive Probability*. Atascadero, CA: Ridgeview Press.
Sharadin, Nathaniel. 2015. "How You Can Reasonably Form Expectations When You're Expecting." *Res Philosophica* 92 (2): 441–452. http://dx.doi.org/10.11612/resphil.2015.92.2.2.
Talbott, W. J. 1991. "Two Principles of Bayesian Epistemology." *Philosophical Studies* 62 (2): 135–150. http://dx.doi.org/10.1007/BF00419049.
van Fraassen, Bas. 1989. *Laws and Symmetry*. Oxford: Oxford University Press.
Williamson, Timothy. 2000. *Knowledge and its Limits*. Oxford: Oxford University Press.

Transformative Choices

Ruth Chang

Abstract: This paper proposes a way to understand transformative choices, choices that change 'who you are.' First, it distinguishes two broad models of transformative choice: 1) 'event-based' transformative choices in which some event—perhaps an experience—downstream from a choice transforms you, and 2) 'choice-based' transformative choices in which the choice itself—and not something downstream from the choice—transforms you. Transformative choices are of interest primarily because they purport to pose a challenge to standard approaches to rational choice. An examination of the event-based transformative choices of L. A. Paul and Edna Ullman-Margalit, however, suggests that event-based transformative choices don't raise any difficulties for standard approaches to rational choice. An account of choice-based transformative choices—and what it is to be transformed—is then proposed. Transformative choices so understood not only capture paradigmatic cases of transformative choice but also point the way to a different way of thinking about rational choice and agency.

> *An angel walks into a fractious philosophy department meeting and says to the Chair: "I'll give you one of three gifts you choose: Wisdom, Truth, or Ten Million Dollars." The Chair chooses Wisdom. She is transformed! But all she does is sit there, staring down at the table. One of her colleagues whispers to her, "Say something!" The Chair replies, "I should have taken the money."*[1]

Some choices are transformative; they *change who we are*. But what is transformative choice? What is transformation and what gets transformed in a transformative choice? How does transformation take place? And if transformative choices are rational, how can they be rational?

[1] Adapted from Carthcart and Klein 2007, 79.

Res Philosophica, *Vol. 92, No. 2, April 2015, pp. 237–282*
http://dx.doi.org/10.11612/resphil.2015.92.2.14

In this paper, I moot a view of transformative choices that answers these questions. While there are different phenomena that go under the label, I will assume that the main interest of transformative choices is in how they push the boundaries of what might be called 'standard approaches' to rational choice. Do transformative choices require us to abandon these approaches? I explore what I take to be the most natural way of thinking about transformative choices and argue that thinking of them in this way poses no threat to standard approaches. I then propose an alternative view that, I argue, requires us to reject a fundamental assumption of such approaches. Its central idea is that transformation is not only something that can happen to us but is something we can *do*. Transformative choices so understood, I suggest, not only capture paradigmatic cases of transformative choice but also point the way to a different way of thinking about rational choice and agency.

1 Two Models of Transformative Choice

We start by proposing two general models of transformative choice, and in particular, two ways in which you might change 'who you are.' We'll have more to say about what it might be to be 'transformed,' that is, to change 'who you are' later, but for now, I want to work with the broad, but I think intuitive, understanding of 'transformation' and 'who you are.'

First, you can be transformed by an event or process. One especially salient kind of event is an *experience*. In 1999, Mike May received an operation that partially restored his sight after 43 years of being blind.[2] We might say that the experience of seeing transformed him—it changed him from being an uncannily talented unsighted person who had broken several downhill skiing records, worked for the CIA, and invented a GPS system for the blind, to a partially-sighted man who has difficulty identifying coke cans at the market and holding a conversation with someone while looking at him or her. May was transformed from a high-functioning unsighted person to a partially-sighted person who struggles—valiantly and, for the most part successfully—to do what sighted people do as a matter of course (Kurson 2007).

Experiences can be, roughly speaking, 'extended,' as in May's case of experiencing life as a sighted person, or 'discrete,' as in the experience of seeing or hearing for the first time. Discrete experiences, such as giving birth, climbing Mount Kilimanjaro, or undergoing violent trauma might not only be themselves transformative, but also be the root of an extended experience. A discrete experience, such as being a victim of a violent crime, can cause you to have the extended experience of seeing strangers as threats.

An experience is one kind of event or process, but there are other kinds of events that can transform you. What transforms you needn't be something

[2] See, e.g., http://www.theguardian.com/science/2003/aug/26/genetics.g2, last accessed May 1, 2015.

subjective in you but something objective in the world. You might be transformed by the event of taking a pill that scrambles your brain and, consequently changes who you are. There need be no appeal to a subjective experience to explain your transformation. More generally, events in the world—including in your own life—can be transformative. When Steve Jobs was fired from Apple, it was arguably the event of being fired that transformed him, not his subjective experience of being fired. Being fired was something that happened in his life, but what 'changed who he was' wasn't the experience of being fired but the event of his being fired. Similarly, the process of military training can transform a cadet; it needn't be how the cadet subjectively experiences the training that changes her; rather it may be the process of being trained to kill that does the transformative work. And when an assistant professor gets tenure, it can be the achievement of having won tenure that transforms her, not the subjective feel of knowing that she can now be fired only for moral turpitude, however wonderful that feeling might be.

Note that an event—whether experiential—can be transformative independently of whether you *choose* it. The nature and features of certain events—including whether they are transformative—may, of course, *depend* on whether you chose them. Experiencing a violent fistfight may transform you if you didn't choose it, but might be all in a day's work if you cage fight for a living. And the arduous discipline of military training might not transform you—or at least not in the same way—if you chose to have it rather than having had it foisted upon you. There are many other factors that will determine whether an event changes who you are—like the social conditions that can determine the meaning of an event (Barnes 2015). For our purposes, we just need to emphasize that in these cases, it is not the choice but rather an event downstream from choice that does the transforming. The first model of transformative choice, then, understands transformative choice as a choice about whether to undergo or bring about an event downstream from choice that transforms you. Call these 'event-based' transformative choices.

There is a second way in which you can change who you are. You can be transformed by a *choice itself*. According to a second model of transformative choice, you change who you are by the very making of a choice, not by some experience or event downstream from your choice. Call these 'choice-based' transformative choices. Of course, a choice is also an event, but for the purposes of this paper, I will understand 'events' as always downstream from a choice in order to mark a distinction between the choice itself transforming an agent and some event or process downstream from choice doing so.

Here are some examples in which the choice itself arguably transforms an agent. Suppose a Hollywood plastic surgeon chooses to give up her 1%-er lifestyle to become a volunteer doctor in a war torn region. Her very choice to give up her material comfort for one of hardship and penury

may change who she is. Or consider Gauguin, whose choice to leave his family to pursue his art in Tahiti transformed him into someone committed to his art at the cost of abandoning prior commitments to his wife and children. And when William Styron's Sophie chose to save her son, Jan, over her daughter, Eva, from certain death in the Nazi death camp, she is transformed by her choice—under duress though it was—into a tragic figure haunted by her role in the death of her daughter. Sometimes the very making of a choice itself can be transformative.

So we might distinguish two models of transformative choice: 1) *event-based transformative choice*, a choice in which an event or process downstream from a choice—perhaps experiential—transforms you, and 2) *choice-based transformative choice*, a choice in which the making of the choice itself transforms you. And there are 'mixed' transformative choices that involve some or other of these two routes to transformation.

My interest in this paper is in understanding whether transformative choices on each model raise any serious challenge to standard approaches rational choice. But I need to be clear about what I mean by 'standard approaches.' Those who worry about transformative choices have tended to assume a rather narrow target—classical expected utility theory, already abandoned by many contemporary philosophers of practical reason—in arguing that transformative choices raise problems for rational choice. By 'standard approaches' to rational choice, I mean something much broader that goes well beyond standard forms of normative expected utility theory and classical rational and social choice theory. My concern is not to show that economists and classical decision theorists have a problematic view of rational choice, though that will be an implication of my argument. I think the challenge that transformative choices pose is much broader. By 'standard approaches' to rational choice, I mean to include *any* reasons- or values- or preference-based normative approach to rational choice that makes the following two fundamental assumptions:

(1) The rationality of a choice is determined by the value (or utility) of the alternatives or the reasons for and against them,[3] and
(2) That in virtue of which we have reasons or values (or utility) is not a choice itself.

The first assumption is clear; the rationality of a choice is given by the value of, or normative reasons for and against, the alternatives.[4] So, for example, the rational choice might be the alternative that is at least as good

[3] By 'determined' I mean to include both the subjective and objective case. That is, standard approaches hold that the values you assign to an alternative determine what it's subjectively rational to choose and that the values the alternatives in fact have determine what it's objectively rational to choose. As I am understanding 'standard approaches,' the first assumption can thus be undermined in either case, but I will for the most part be concerned with the objective case. So when I talk of 'assigning a value,' I assume that the assignment is correct.
[4] I'll be using 'values' and 'reasons' interchangeably. Nothing I say here turns on which, if either, is explanatorily more fundamental.

as the others or most strongly supported by reasons. Or it might be what simply 'ought to be done' or is 'good enough.' So long as the view holds that the rationality of a choice is based on either the value of or the reasons (which might themselves be given by preferences constrained by certain axioms) for the alternatives, it counts as a potential target view to which transformative choices might pose a challenge.

The second assumption is often only implicitly held or can fairly be imputed; choice itself can't be that in virtue of which you have a reason or you ought to do something. Simply by choosing to jump off a cliff, you can't make it the case that you have a reason to do so or that it's a valuable thing to do. To think otherwise would lead to the familiar problem of bootstrapping reasons, viz., that we can simply create reasons for ourselves to x simply by choosing x.[5] Strictly speaking, most standard views don't broach the question of that in virtue of which we have reasons. But if they didn't implicitly assume that choice couldn't be that in virtue of which we have reasons, the substance of their accounts would be very different. They would have to allow that the mere fact that you choose something could create an additional reason to choose it and thereby be self-justifying.[6] This consequence is not consistent with views I include among 'standard approaches.' And some such views have explicitly denied that choice itself can be that in virtue of which we have reasons (Scanlon 2004). According to standard views, our reasons to x are given by facts that count in favor of x—such as the fact that it's delicious or that I promised or that I want or prefer it—and those facts are reasons in virtue of facts *other than* the fact that we have chosen a certain way. 'Standard approaches,' then, include nearly every view about rational choice that has been a going concern in the last few centuries.[7]

[5] An especially vivid illustration of the bootstrapping problem is given by Jerry Cohen's Mafioso objection against Korsgaard's Kant according to which willing can be a source of reasons. If choosing can be a source of reasons, then the Mafioso can bootstrap his way into having all-things-considered reasons to shoot the kneecaps off his enemy simply by choosing to do so. See Korsgaard 1996.

[6] Which is not to say, of course, that norms of structural rationality, such as 'follow through on your choice unless you have a reason not to,' might not be involved. The interest here is on the 'rationality of reasons,' not of the norms governing consistency and coherence among our mental states.

[7] Revealed preference theory isn't normative in the sense of interest, and so I exclude it from 'standard approaches.' Other theories excluded are a particular kind of neo-Kantianism according to which willing is, strictly speaking, the ground of one's reasons, such as that I believe championed by Christine Korsgaard (1996). Other neo-Kantians, such as Elizabeth Anderson (1993) and Barbara Herman (1996), explain the ground of practical reasons in terms of some fundamental value, such as the value of humanity, while still other neo-Kantians, such as Thomas Hill (2001), as I read him, take a roughly Humean approach to the ground of practical reasons. I have argued elsewhere that views that may appear to hold that the rationality of a choice is determined by something other than the value of the alternatives or the reasons for them—such as views according to which the rational choice is just 'the thing to do' or what meets some test or standard, are in fact views that hold that the rationality of

As we will see, event-based transformative choices purport to challenge the first assumption and choice-based transformative choices the second. But, I'll suggest, event-based transformative choices *don't* raise any difficulties for the first assumption and so, to that extent, can be handled by standard approaches to rational choice. However, as I'll argue, choice-based transformative choices *do* require rejection of the second assumption and thus raise a genuine challenge to standard approaches. So one upshot of the arguments here is that only choice-based, and not event-based, transformative choices, raise difficulties for standard approaches to rational choice.

We start with two accounts of event-based transformative choices, one offered by L. A. Paul, who focuses on a particular kind of experiential event, 'epistemically transformative experiences,' and the other by Edna Ullman-Margalit, who considers transformative events more broadly. These are the only two analytic accounts of transformative choices of which I am aware. Both have as their target normative rational decision theory, and they purport to raise a problem for such theories because they claim that the rationality of a transformative choice cannot be based on the value of the alternatives. If successful, these accounts would also pose a problem for standard approaches, as they are understood, broadly, here. But I have doubts about whether these views, while interesting in their own right, succeed in raising any difficulties for the idea that the rationality of choice is based on the value of the alternatives. They don't seem to raise a genuine challenge to standard ways of thinking about rational choice.

We then turn to the choice-based model of transformative choices. We start by proposing a way to understand transformation—who and what is transformed and how transformation takes place. We then describe how choices might themselves be that in virtue of which something is a reason, that is, 'grounds' for something's being a reason. When we choose in a *thick* sense, that is, by *committing* to an alternative, we create reasons for ourselves to choose it—our commitment is that in virtue of which we have a reason to do something. So by choosing, we can create new reasons for ourselves, thereby transforming 'who we are.' This choice-based view of transformation does not run afoul of the bootstrapping problem that led standard approaches to assume that choice cannot ground reasons because it is part of a more general metanormative view about practical normativity, what I have elsewhere called 'hybrid voluntarism,' according to which our normative power to create reasons through our commitments is suitably constrained. Thus, hybrid voluntarism, while I believe an independently attractive view about the source of normativity, also underwrites an attractive and plausible account of transformative choices. Crucial to choice-based transformative choices is the idea that transformation is something we *do*,

choice is determined by the *comparative* value of or strength of reasons for and against the alternatives. See Chang Forthcoming.

not something that happens to us. I suggest that this account not only captures paradigmatic cases of transformative choice but points to a richer understanding of what it is to be a rational agent.

It's worth saying at the outset that since philosophical investigation of transformative choices is relatively new territory, especially for analytic philosophers, I'll be relying on large, abstruse—but ordinary—notions on which we may have little more than an intuitive fix. If you don't share the notions to which I appeal, remember that my aim is to make a case for *one* way of thinking about transformative choice, and in particular, a way that poses a genuine challenge for standard ways of thinking about rational choice.

2 Event-Based Transformative Choices

In an event-based transformative choice, an event downstream from choice transforms you. One particular kind of an event—an experience—might be thought to be an especially good candidate event that can transform an agent.

2.1 Epistemically Transformative Experiences

L. A. Paul has recently proposed that a very particular kind of experience—an "epistemically transformative" experience—can be personally transformative, that is, "change who you are, in the sense of radically changing your point of view" (2014, 10–11). As Paul goes on to say, changing your point of view is a matter of changing your "personal or subjective preferences" (16). And when you radically change your point of view, you "change your post-experience preferences, or change how your post-experience self values outcomes" (48).

An epistemically transformative experience is an experience you can't know what it's like to have without actually having the experience (10). 'What it's like' to have an experience goes beyond its raw phenomenal feel and includes attitudes and emotions you might have in response to that feel (27). As Paul explains, "When a person has a new and different kind of experience, a kind of experience that teaches her something she could not have learned without having that kind of experience, she has an *epistemic transformation*. Her knowledge of what something is like, and thus her point of view, changes"(16).

For example, Paul urges that the experience of having a child—the experience of "gestating, producing, and becoming attached to that child"—is epistemically transformative: you can't know antecedent to the experience of having a child, what it's like for you to have a child (77-78; see also Paul 2015). And since the knowledge you gain when you experience having a child "radically chang[es] your point of view," the experience is also personally transformative.

Since you can't know what an epistemically transformative experience is like before having the experience, you can't assign value to what it's subjectively like antecedent to the experience. This, in turn, Paul thinks, raises a problem for standard forms of normative expected utility theory, which, she notes, presupposes that the rationality of a choice is determined by the values of the alternatives. If your choice is whether to have an experience you can't assign a value to having, then you can't rationally choose it on the basis of its value: normative expected utility theory seems to break down.[8] Paul also makes the further point that since you are personally transformed, the preferences you have before the experience are different from those you have after the experience and so, she urges, there is the problem of determining which set of preferences should be the ones on the basis of which to evaluative the alternatives. Since this latter aspect of Paul's view is essentially the same as the central point of Edna Ullman-Margalit's account of transformative choices, which we discuss below, we here focus on what is most distinctive about Paul's account—viz., her claim that epistemic transformation entails that you can't know the value of the experience and so can't assign it a value, which you need to do in order rationally to choose whether to undergo it.

Interestingly, although Paul thinks that epistemically transformative experiences lead to a breakdown in rational choice, she doesn't conclude that we should reject normative expected utility theory. Instead, she suggests that we reconceive epistemically transformative choices in a way that no longer poses a challenge to the standard approaches. We should think of the choice about whether to have a child, for example, not in terms of what it's like to have a child but instead as a choice about whether to gain a certain kind of knowledge—viz., the knowledge of what it's like to have a child.[9] That reframing of the choice, Paul suggests, saves standard approaches from the challenge posed by epistemic experiences because we can assign value to knowing what an experience is like.

I doubt, however, whether epistemically transformative experiences raise any difficulties for any plausible forms of normative expected utility theory or, indeed, for 'standard approaches' to rational choice more broadly understood here. First, it seems that genuinely epistemically transformative experiences as Paul strictly understands them are very rare; we can know what most experiences are like antecedent to having the experience, and

[8] As Paul writes, "To apply a normative decision-theoretic model for ignorance to a decision about whether to perform an act, you need to know the values of the relevant outcomes, including their relative strengths, and you must be able to compare the values of the outcomes in order to determine the overall structure of the value space. But in the case of a decision involving a[n epistemically] transformative experience, you cannot know what it is like to have that kind of experience until you've had it" (2014, 32). Note too that Paul's worry applies to the subjective reading of the first assumption.

[9] Paul: "[In transformative choices] we choose between the alternatives of discovering what it is like to have the new preferences and experiences involved, or keeping the status quo" (2014, 122). Thus what it's subjectively like drops away as relevant to the choice.

we can know enough about them in order to assign them at least some rough value. Second, even if we can't know what an experience is like before having it, it doesn't follow that we can't assign it a value. This is because the value relevant to rational choice isn't simply the value of what it's subjectively like. The objective value of experience matters too and typically matters in a way that allows us to assign a rough value to the experience. And third, quite generally, the focus on *experience* is misplaced; transformative choices aren't typically about what *experiences* to have but are choices about whether to undergo certain transformative *events* that go well beyond how things subjectively feel to an agent. Transformation isn't typically about how we subjectively experience things but about how things *in the world* change us. While an interesting phenomenon in its own right, epistemic transformation is not, I think, the right key to understanding transformative choices.

2.1.1 *Epistemic Transformation Is Rare and Atypical of Transformative Choices*

What sort of experiences are most plausibly ones about which we could have no antecedent knowledge of what they're like? I suggest that, as a first cut, we work with the idea that epistemically transformative experiences as the de novo exercise of a 'basic' capacity. Trying to give a proper account of what makes something a basic capacity would take us too far afield, but we might give an intuitive gloss of them as follows: a basic capacity is a sui generis capacity that belongs to a set of capacities from which all others can be derived for a type of creature; basic capacities are ones that are not the exercise of other capacities but are atomic capacities for that type of creature. And since our interest is in rational choice for humans, we focus our attention on the basic capacities of human rational agents. Seeing, hearing, tasting, touching, smelling—the five traditional senses—are plausibly basic physical capacities. But so too is the normative capacity to recognize and respond to reasons. If you've never exercised those capacities before, it's plausible to think that you can't know in advance of experiencing their exercise what it's subjectively like to experience their exercise.

What about seeing red, tasting Vegemite, or having a child for the first time? Are these cases, which Paul uses to illustrate epistemically transformative experiences, really experiences about which we could have no knowledge antecedent to having the experience?

Paul uses as her touchstone case that of Frank Jackson's Mary, who, after living in a black and white room all her life, emerges and sees red for the first time. Jackson points out that antecedent to the experience of seeing red, Mary can't know what it's like to see red. The point of Jackson's (1982) example is to show that there are some phenomenal properties that require experience in order to know them (and that physicalism was therefore in

doubt).[10] 'What it's like' for Jackson, however, is a simply a matter of phenomenological feel. As we've seen, 'what it's like' for Paul goes beyond mere phenomenal feel and includes attitudes and emotional responses to that feel (2014, 12, 27). Paul is right to insist upon this broader notion of 'what it's like' since phenomenal feel mostly isn't relevant or all that's relevant in transformative choices. When we are talking about 'what it's like' to have a child, for instance, we aren't concerned simply with the phenomenological feel of that experience. So when Paul argues that just as Mary can't know 'what it's like' to see red, the child-free you could not know 'what it's like' to have a child, we must be careful that there is no equivocation. So we might ask, Can Mary know what it's like to see red in the broader, beyond-mere-phenomenological-feel sense that is relevant to transformative choice, antecedent to the experience of seeing red? As we'll see, there is good reason to doubt that Paul's touchstone case is a case of epistemic transformation, and a fortiori, that 'higher-level,' more complex experiences of tasting something new and becoming a parent are cases of epistemically transformative experiences. As it turns out, genuinely epistemically transformative experiences are rather hard to come by.

But first, we need to make a clarification. Paul says that an epistemically transformative experience as an experience that you *could not know what it is like* to have without having the experience.[11] There is a stronger and a weaker interpretation of this claim. On the stronger interpretation, you can't have *any knowledge whatsoever* what the experience is like antecedent to having the experience. It would then follow straightforwardly that you can't assign a value to it on that basis. On the weaker interpretation, although you can have *some* knowledge of what it's like, that knowledge is not sufficient for you to assign a value to it based on what it's like. Paul's text seems implicitly to endorse the stronger reading since she raises objections to claims about how you could have *some* knowledge of what it's like as a way of bolstering her claim that you can't know what it's like and since she doesn't give an account of what sort of knowledge of what it's like would be insufficient to assign a relevant value to the experience. But it's worth considering both interpretations in turn.

Could Mary, who is born and raised in a black and white room, have *no knowledge whatsoever* about what seeing red is like? I think the answer is 'no'; although Mary may not be able to know the phenomenal feel of seeing red, she might nevertheless have *some* knowledge of what seeing red is like beyond its phenomenal feel. I think this is true for three reasons.

First, seeing red isn't for Mary simply the de novo exercise of basic capacities. Seeing red is, after all, like seeing black and white in some

[10] See also Nagel 1974 and Lewis 2004.

[11] Paul: "[I]n the case of a decision involving a[n epistemically] transformative experience, you cannot know what it is like to have that kind of experience until you've had it" (2014, 32). The modality of 'cannot' here is significant in evaluating the scope of epistemic transformation but I am unclear as to what she has in mind.

respects. While seeing is a basic capacity, seeing red is not—it involves the exercise of the basic capacity of seeing, which by hypothesis, Mary has exercised before. Since it involves the familiar exercise of a capacity, Mary can have *some* knowledge of what it's like. Thus, one reason to think that seeing red isn't epistemically transformative is that it involves the exercise of capacities that have been exercised before.

Seeing red might also be said to belong to a *type* of experience, where 'types' are individuated by 'what it's like' for humans. All humans have a certain range of subjective responses to the exercise of their capacities. But since there is a wide variation of human subjective responses to experiences, an experience will belong to many different 'types'—different 'what it's like's.' So for example for some humans, seeing red will fall under the 'emotionally neutral' type of experience, while for others it may fall under the 'thrilling' type of experience. The point here is that, given human capacities and variations, a matrix of 'types' of experiences could be constructed so that every possible human experience could be classified as belonging to a range of 'types'—thrilling, emotionally neutral, boring, etc. Since there's no a priori reason to think that such an individuation of experiences wasn't possible, Mary could look up the experience of seeing red in the matrix and see the range of types of experience in which seeing red falls for humans. She would see, for instance, that seeing red falls under many of the same types of experiences that other experiences she's had fall under. If seeing red always falls under a type of experience that includes other experiences she's had, then she would thereby have some knowledge—disjunctive though it may be—of what seeing red would be like for her. It would be like one of those other experiences that fall under the types that seeing red falls under for humans generally. She doesn't need to know *which* type of experience seeing red would fall under *for her*; it's enough that she knows that whatever seeing red will be like for her, it will be like one of the experiences she has already had that fall under the types of experience that seeing red falls under for humans generally. So a second reason to think Mary could have some knowledge of what it's like to see red is that she could know what it's disjunctively like for her to see red.

Finally, Mary can get testimonial evidence about what it's like to see red from those similarly situated—or in the ideal case, from those who share the same physiological and psychological properties that subvene her own experience of seeing red and are otherwise similarly situated. Of course, Mary would need also to gauge the reliability of the testimony and be aware that the testimony given might reflect an ex-poste shift in the way one comes to view an experience after having it. But after confirming reliability, such testimony gives her at the very least knowledge of what an experience is thought to be like after having had it—even if that involves a change in her preferences—by those that share her subvening properties. That is some knowledge about what it's like—it's the kind of experience which, ex ante, those similar to her—or in the best case scenario, her

Doppleganger—experience in such-and-such way, and ex poste, report in such-and-such way. If the testimony is from those merely similar to her but not exactly like her in the relevant respects, it is evidence of what it might be like for her to the extent that she is like those whose testimony she has.[12]

This is not to say that if Mary is to make a rational choice about whether to see red, she must blindly follow the testimony of others or that her choice about whether to see red is no longer first-personal or 'authentic,' a worry Paul raises about the appeal to testimonial evidence (2015, 19 and footnote 33; 2014, 105–107). Taking account of evidence about what the experience is like for those similar to ourselves does not entail that when we make a decision about whether to undergo an experience, we don't act autonomously or authentically or from the first-personal perspective. None of those features of agency require that we act simply from our own subjective preferences, uninformed by external facts, including the testimony of others.[13] Mary's Doppleganger might reliably inform Mary that seeing red is just fine. But Mary might nevertheless prefer not to take the risk, just in case the testimony isn't one-hundred-percent reliable or is ex-poste corrupted. Thus appeal to testimonial evidence doesn't undermine Mary's agency. She can have some knowledge of what seeing red is like through testimony.

It's worth noting that the last two ways in which you can get knowledge about what an experience is like also hold for an agent's de novo exercise of basic capacities that have already been exercised by other humans. A congenitally blind person can have some knowledge of what seeing is like by knowing, disjunctively, what seeing is like for the human species, and by reliable testimony from those who share her subvening properties. She won't be able to know its phenomenal feel, but that's not to say that she can know nothing about what it's like.[14] On the strong interpretation of epistemic transformation, then, the only genuinely epistemically transformative experiences are those involving *only* the de novo exercise of basic

[12] The modality of 'can't' in Paul's claim that you can't have knowledge is important here, but since Paul does not elaborate, I leave open the sort of testimonial and matrix evidence to which any agent might have access. If, for example, everyone could always have reliable testimonial evidence from her Doppleganger, that would make short work of epistemic transformation. Insofar as Paul has a very restricted sense of 'can't' in mind, this reduces the scope—and interest—of her claims accordingly.

[13] Paul seems to assume a view of authentic agency whereby an agent simply models possible futures without external, non-self-generated data, and then consults her subjective preferences about those futures. I believe that there are more plausible views of authentic agency that even standard normative expected utility theorists can help themselves to. See Paul 2014, 105–107, 112, 130. We shouldn't elide authentic, autonomous, and first-personal choice with choice based solely on our preferences about which subjective experience to have.

[14] Later, I'll be giving another reason to doubt whether the choice of whether to become sighted involves an epistemically transformative experience: what transforms you in such a choice isn't the experience of being sighted but the goods and bads of the objective fact of being sighted.

human capacities for which no matrix can be constructed and no reliable testimony can be given. It's difficult to imagine such experiences. Maybe evolved new basic capacities unlike any capacity ever exercised by humans before, such as teletransporting oneself by thought alone, would count. But the paradigmatic transformative choices that are of interest to us don't involve such fanciful experiences.

What goes for Mary goes too for the person tasting Vegemite for the first time and, most importantly, for the person having experiences that are typically thought to be personally transformative, such as the experience of having a child. The experience of "gestating, producing and becoming attached to a child" involves the exercise of many capacities you've exercised before. So on the basis of your experience in exercising those capacities, you could have some knowledge of what the experience of having a child will be like for you. Moreover, given our matrix, you can know that the experience of having a child falls under a certain range of types of experiences for humans, and you've very likely had experiences that fall under those types before. Maybe the experience of having a child falls under types that include the experience of being in a family, passing a kidney stone, having a pet, and so on. Since you've had experiences that fall under the same types before, you will know *something* about what it's like to have a child. And, finally, you can get reliable testimonial evidence from parents who share the properties—including the moral and social ones—that subvene what it's like for you to have a child (Harman 2015). That testimonial evidence will give you *some* knowledge of what it's like for you to have a child, especially since gestating, producing, and becoming attached to your child is a multifaceted experience that involves many different 'sub' experiences about which you can know something in one of our three delineated ways. None of this is to deny that the experience of having a child will involve many new experiences or that you can know *exactly* what it would be like for you before actually undergoing the experience. But our point is only that it will involve experiences that either you have had before or are like experiences you've had before, or be amenable to reliable testimonial report. And so you can have *some* knowledge of what it is like. If you can have some knowledge of what it is like, we can't simply assume, with Paul, that you *can't* assign some, perhaps rough, value to the experience.

The strong interpretation of Paul's claim, then, doesn't hold of the kind of experiences that seem most relevant to transformative choices. You not only can know *something* of what it's like to see red for the first time, but you can also know *something* of what it's like to try Vegemite or to have your first child. We haven't shown that you can, therefore, assign value to the experience—we'll make a suggestion about that in a moment—but have argued only that since such experiences aren't epistemically transformative in the strong sense, it doesn't follow that we can't assign a value to them.

What about the weak interpretation of epistemic transformation? When Paul claims that we can't know antecedent to an experience what it's like,

she may mean that to be consistent with the claim that we know something of what it's like, or know what it's like to some extent (though, as I've noted, this interpretation does not sit easily with the text). What she may mean by 'knowing what it's like,' in particular, is 'knowing what it's like sufficient to assign it a value on that basis.' That might be consistent with knowing *something* of what it's like.

But what could 'knowing what it's like sufficiently to assign it a value on that basis' amount to? One important question is: If an experience involves the de novo exercise of a basic capacity, is that sufficient to *block* assigning it a value on the basis of what it's like? To answer this question, we need a theory of the ways in which various bits of knowledge of what it's like contribute to the value of what it's like. We want to know, in particular, whether there are 'organic unities' that form between different pieces of knowledge of what an experience is like so that if we lack a bit of knowledge, we lack the knowledge needed to assign a value on the basis of what it's like. Paul doesn't offer a theory and we don't have space to try out a theory here, but we can consider some cases and draw a tentative conclusion.

There are some cases in which the de novo exercise of a basic capacity is clearly such an insignificant contributory factor to what an experience is like that you can nevertheless assign a rough value to what it's like. Suppose for example, that Mike May, while blind, is hit by a car. Pre-sight-restoring-operation, May knows enough about what it would be like to be hit by a car while sighted to assign it a rough value, even though the experience of being hit by a car while sighted would involve the de novo exercise of the basic capacity of seeing. What it's like to be hit by a car is primarily about the hitting, not about the seeing.

There are other cases, however, in which the experience so centrally involves the de novo exercise of a basic capacity that it seems that you can't know what it's like without knowing what it's like to exercise that basic capacity. The experience of listening to Beethoven's Fifth Symphony for a deaf person might seem to be like this. Although that experience involves the exercise of capacities the person has, suppose, exercised before, for instance, the capacity to take sensory input to form an artistic interpretation, what the experience of listening to Beethoven's Fifth is like is so centrally about hearing, that without familiarity with what it's like to hear, it seems that you can't know what it's like to listen to Beethoven's Fifth.

But note that it's yet a further question whether *other* knowledge you could have of what listening to Beethoven's Fifth is like is sufficient for you to assign a rough value to the experience nonetheless. This is, as I've said, a matter of substantive argument, but for my own money, I would bet that knowing the disjunctive range of what it could be like for her to hear Beethoven's Fifth à la our matrix, and having testimonial evidence from similarly situated friends, is sufficient knowledge to assign a rough value to the experience. My own suspicion is that any plausible theory

of what knowledge is sufficient to assign value on the basis of what it's like will, at best, show that *only* experiences involving *solely* the de novo exercise of a basic capacity that hasn't been exercised much by humans before, will preclude the assignment of a value to an experience based on what it's like. Experiences, such as listening to Beethoven's Fifth by a deaf person, which involve not only the de novo exercise of a basic capacity but the familiar exercise of other capacities as well, are experiences about which we can have both 'matrix' and 'testimonial' knowledge, knowledge which is, I believe, sufficient to allow us to assign a rough value to what it's like to have the experience. You can know, for instance, that what it's like to listen to Beethoven's Fifth will be better than what it's like to be skinned alive. The same goes, I believe, for the experience of seeing, hearing, tasting, and so on, for the first time. If this is right, we need to revise our first thoughts about epistemically transformative experiences. They don't include the de novo exercise of any basic capacity, but only those that humans in general have not exercised before. This reduces the interest of epistemically transformative experiences significantly.

Regardless of whether these substantive musings are correct, we can reasonably doubt that paradigmatic transformative choices, such as whether to have a child, involve epistemically transformative experiences since, as we've seen, either, on the strong interpretation, such experiences are recherché and don't figure in paradigmatic transformative choices, or, on the weak interpretation, it doesn't follow from the fact that we lack some knowledge of what an experience is like that we cannot assign a value to what it's like. You can know that what it's like to have a child isn't as bad as being boiled in hot oil or slowly dismembered without anesthesia.[15]

None of this is to say that there couldn't be a transformative choice that involves having an epistemically transformative experience. Choosing whether solely to exercise a new basic capacity that has never been exercised by humans before plausibly involves such an experience.[16] Transformative choices that are both epistemically transformative and personally transformative in Paul's sense, then, are I think best restricted to just these cases. Paul's case of choosing whether to become a vampire, though purely fictional, is plausibly such a case (though it has the added complication that it might involve abandoning our status as humans, though if Stephanie Meyer of Twilight fame is right, vampires and humans can interbreed). Becoming a vampire would presumably involve the de novo exercise of basic capacities never before exercised by humans, and so it's plausible to think that we can't know, as mere humans, what it's like to be a non-human

[15] Another way to put the point is that we can do a bit of cognitive modeling to determine the rough value of an experience. When you simulate what it's like to have a child, you may be unsure what it will be like. But you can be sure that what it will be like is better than what it would be like to be slowly dismembered without anesthesia.

[16] It seems likely that neither matrix nor testimonial knowledge will be of much help in knowing what the exercise of a new-to-humankind basic capacity is like.

vampire. But they don't include choices about whether to become sighted, to taste Vegemite for the first time, or to have a child. Such experiences always seem to involve the exercise of familiar capacities, be representable in a matrix of human responses to such experiences, and be amenable to testimonial evidence about what they're like.

It is at best doubtful that we have many epistemically transformative experiences. And paradigmatic cases of transformative choice—such as whether to have a child, change careers, give most of your wealth away to charity—don't seem to involve epistemic transformation. You can have *some* knowledge of what an experience is like, even if that experience involves the de novo exercise of a basic capacity, because such experiences also involve the exercise of capacities you've exercised before, you can know disjunctively what the experience is like, and you can gain suitable testimonial evidence of what it is like. We should be cautious about overgeneralizing from fictional, recherché cases, like choosing whether to become a vampire, to the kind of experiences typically involved in a transformative choice.

2.1.2 Epistemically Transformative Experiences Can Be Evaluated

If the arguments of the last section are correct, the only epistemically transformative experiences there are involve the de novo exercise of a basic capacity that has not been exercised by humans before—not the experiences involved in paradigmatic transformative choices. But now let us set those arguments aside and assume, for the sake of argument, that transformative experiences are typically epistemically transformative as Paul suggests. Does it follow, as Paul argues, that we can't assign a subjective value to such experiences?

Suppose, arguendo, that you can't know in advance of suffering a violent trauma what it would be like to suffer it. But you can nevertheless know that what it will be like will be bad. So you can assign a subjective value to the experience even if, pace our arguments of the last section, you can't know in advance what the experience will be like. Or suppose that you can't know in advance of rollicking with the angels in heaven what being in heaven is like, but you can know that what it will be like will be good—or at least better than burning in the fires of hell and damnation. So why does Paul think that epistemically transformative experiences can't be assigned a subjective value?

Paul herself acknowledges that "you can know that being eaten by a shark will be horrible" (2014, 27). But she says that such experiences are ones about which "there is no need to deliberate by cognitively modeling in order to assess the subjective value of the relevant outcomes" (27) and so she "will be setting aside decisions like [these]" (27). As she puts it, she wants to focus on epistemically transformative experiences "you're not sure how you'd respond to" (28).

What Paul appears to be claiming is that an epistemically transformative experience can't be assigned a subjective value unless it can be. If it can be, she's not interested in it. And whether an epistemically transformative experience can be assigned a subjective value depends wholly on whether "you're not sure how you'd respond to" it. The idea here seems to be that some epistemically transformative experiences can be assigned a subjective value independently of having to run a simulation of what they're like, that is independent of deliberation by cognitive modeling. Even though you can't know, by hypothesis, what being eaten by a shark is like, you can assign it a subjective value because its subjective value doesn't turn on running a simulation of what it's like. If, on the other hand, you're not sure what an experience is like, you do a bit of cognitive modeling, and if you're still stuck, if "you're not sure how you'd respond to" it, the experience counts as epistemically transformative in the sense Paul is interested in, the kind to which you can't assign a subjective value on the basis of what it's like. So from not being sure how you'd respond to an experience, it seems to follow, as Paul suggests, that you can't assign a subjective value to it on that basis. But not knowing how you'd respond to an experience is one thing, not being able to assign it a subjective value another. What's needed, I think, is an account of why some epistemically transformative experiences can evidently be assigned a subjective value while others putatively can't. That epistemically transformative choices split in this way suggests that being epistemically transformative isn't the feature that is really supposed to be causing problems for expected utility theory in the first place.[17]

It might also be wondered why Paul thinks that in order to be able to assign a subjective value to an epistemically transformative experience, we have to *know* what it's like. Standard approaches to rational choice generally recognize that we rarely *know* what will happen in the future. So the value we assign to future outcomes is based in part on the probability that that outcome will come to pass. Why can't Mike May simply have some probabilistic expectation of what seeing is like and assign a subjective value to the experience of seeing on that basis? Seeing might be like many different things, and May could have a probability distribution over the many different ways seeing might be like. He can then assign a value to each way seeing might be like, however varied these values might be. So, a toy illustration: the experience of seeing for the first time might have a 50% chance of being emotionally thrilling to him, a 35% chance of being confusing and depressing to him, a 10% chance of being scary to him, and so on. Having been thrilled, confused and depressed, and scared before, he can assign a rough value respectively to having a thrilling, confusing

[17] There is a large discussion that could be undertaken here about what feature of experience can play the role Paul envisions if, as we've suggested, being epistemically transformative isn't it. But this would take us too far astray. In any case, as we argue below, any event downstream from choice, whatever it's features, won't raise a problem for standard approaches to rational choice as they are understood here.

and depressing, and scary experience by coming to see even if the way it is thrilling, etc., might not be exactly the same way some of his past experiences have been thrilling, etc.

May's uncertainty may not only be nonnormative—about what the experience is like—but normative—about what utility to assign to an experience *like that*. So even assuming that he *knows* what seeing will be like, he may be uncertain as to what utility it has. In this case, he can do an expected value calculation on the utilities, though there are more complex and plausible approaches to handling normative uncertainty.[18] Of course, he can't know exactly what the value of the experience will be—maybe he can only assign a rough value—and nor can he know the precise probabilities of the experience being one way rather than another—and that's why he can only assign a probability distribution among the possible outcomes. But having collected all the rough probabilities and rough utilities, he can then tot up the expected utility of seeing by multiplying the rough probability of seeing being like this rather than that and the rough value assigned to its being like this. All of this should sound familiar, because it is essentially a roughed-up version of standard expected utility theory under normative and nonnormative uncertainty.[19] If we can assign probabilities to what an experience will be like for us, rather than *knowing* what an experience will be like for us, and we can assign rough utilities to each way the experience might be—however varied those utilities might be—we are squarely within the framework of expected utility theory. It doesn't matter how varied the possible utilities might be because we can still maximize the expected utility whatever the utilities—however rough—may be. It seems clear that we can assign such probabilities and utilities—*unless* of course the experience is wholly unlike anything we—or any other human—has ever experienced before. That's how we return full circle to the idea that genuinely epistemically transformative experiences are those involving only the de novo exercise of a capacity never before exercised by humans.

But let's grant Paul's assumption that in order to assign a subjective value or utility to an epistemically transformative experience, you have to *know* what it's like—probabilistic information won't do.[20] And let's also grant

[18] For views about how expected utility theory can deal with normative uncertainty, see Sepielli 2009 and Ross 2008.

[19] Things are more complicated than this sentence suggests, but we don't have to attend to those complications for our purposes. The key point is that it is unclear whether assigning subjective value requires *knowing* both what the experience is like and what its utility is if it's like that. Something short of knowing could well suffice for assigning subjective value.

[20] Some of Paul's assumptions seem to turn on the very particular form of normative expected utility theory that is her target. The target seems to be a view that requires for rational choice 1) knowledge of what an experience is like through cognitive modeling of what it's like, where 2) experience is all that matters for rational choice, 3) the subjective value of the experience—what the experience is like—crucially matters for the rationality of choice such that if you don't know what it's subjectively like, you can't assign it a value relevant to choice,

that a rough range of subjective value assignments won't suffice as an assignment of subjective value. Let's also grant, for the sake of argument, that there are epistemically transformative experiences beyond merely fictional cases involving the de novo exercise of a new-to-human-kind basic capacity, such as the vampyric capacity to conduct one's life without sleep. Because our interest is in transformation in the real world, let's also just grant, for the sake of argument, that *any* de novo exercise of a basic capacity is one you can't know what it's like antecedent to having the experience. So seeing for the first time, hearing for the first time, and so on, will be, by hypothesis, epistemically transformative (even though, recall, we've seen good reasons to doubt this). Even granting all this, pace our reasons to think otherwise, we can, as I'll now suggest, nevertheless assign a relevant value to an epistemically transformative experience. So even being as concessive as I think we can be, it seems doubtful whether epistemically transformative experiences raise any problems for the assumption of standard approaches that the rationality of a choice is based on the value of the options. This is because the value of an option that is relevant to rational choice goes well beyond the subjective value of what an experience is like.

Central to Paul's argument about transformative choices is the idea that such a choice "essentially involves your subjective values" (2014, 18) and that in such choices there is no "external reason that trumps or dominates your choice, making subjective deliberation irrelevant or unnecessary" (19).[21] But she seems to think that so long as the subjective value of an experience is necessarily a part of its value, if you can't assign a subjective value, you can't assign a value to the experience.[22] But this doesn't follow.

Return to Mike May who must choose whether to have an operation that will give him the experience of being sighted. Let's suppose, again for the sake of argument, that the experience of being sighted is epistemically transformative for him—he couldn't know what it is like to be sighted

4) for your choice to be 'first-personal,' it must be based primarily on the subjective value of what an experience is like, and 5) the value of an alternative can never be rough—that is, it can only be represented by a standard utility function and not by anything more sophisticated, such as a vector or probability distribution. Moreover, for there to be any reasonable scope for epistemically transformative experiences, the "can't" in the claim that they are experiences the agent 'can't' know antecedently what they're like must be read broadly to include merely contingent factors that may prevent a particular agent from knowing what an experience is like. This further narrows the import of her conclusion. See Paul 2014, passim. As we've noted, our understanding of 'standard approaches' is much more wide-ranging.

[21] It's unclear to me whether by 'external reason' Paul means something like a reason of morality—for example a moral prohibition—that overrides the reasons that would otherwise determine the rational choice or whether she means to include by 'external reason' the objective values of having the experience. I take the conservative interpretation and so argue that the objective values of having an experience are relevant to transformative choices.

[22] Recall that we are assuming, for the sake of argument, though we have queried it, that in order assign a subjective value, you have to *know* what the experience is like. So if you don't know what an experience is like, it follows that you can't assign it a subjective value.

antecedent to the experience of being sighted. Let us further suppose that because he can't know what it's like to be sighted, he can't assign a subjective value to the experience of being sighted. Must it follow that he can't assign a value to the experience of being sighted overall?

Note that it would be irrational to think that what it is subjectively like for May to be sighted is *all* that matters in the choice of whether to undergo the operation. The choice about whether to see is not a choice about which subjectively best experiential feel to have. At the very least the objective values of the experience matter too. For example, by having the experience of being sighted, May can gain certain objective goods that improve his well-being, such as, say, greater intimacy with his wife and children, less dependence on others in the execution of quotidian tasks, deeper and more varied social connections with strangers, and so on. Even if we assume, arguendo, that May can't assign a subjective value to seeing, he can, we can suppose, know and evaluate the objective goods he will gain by having that experience. This evaluation of the objective value of the experience can be sufficient for assigning a rough overall value to the experience: given the objective value of the experience, he knows that the experience of seeing has greater value overall—taking into account both subjective and objective value—than the experience of having a hot poker in his eye. This is not because the objective values of having the experience 'trump' what it's like, in the way that rights might trump or be lexically prior to utility, or because they 'dominate' the subjective value of what it's like in the sense that one item might dominate or be pareto superior to another if it is at least as good in all respects and better in at least one. In May's case, the subjective and objective values of the experience weigh against one another in the ordinary way, but because the objective values are such important and significant contributors to the overall value of the experience, a rough overall value can be assigned to the experience even without knowing its subjective value. The subjective value of the experience might then affect the value of the experience within some rough range.[23] So if we can know the objective value of having an experience, even if we don't know its subjective value, we might nevertheless be able to assign it a rough overall value.

What about the case of choosing whether to have a child? Gestating, producing, and becoming attached to a child has a certain subjective feel. But it would be the height of irrationality to think that the choice of whether to have a child is simply a matter of getting the subjectively best experience. The objective values of the experience also matter, and they matter significantly. You can know, for instance, that the experience of

[23] The same goes, I believe, for experiencing life as a non-human vampire or bat or squirrel. You can have knowledge of the objective value of that experience. So, for instance, the experience of life as a squirrel is probably objectively worse, roughly speaking, than the experience of life as a human, at least from where you stand now. It may be less clear in the fictional case of experiencing life as a vampire. The issue of evaluation from different standpoints is one I take up in the discussion of Ullman-Margalit's view below.

having a child will objectively enrich your life in significant ways, be instrumental in bringing a valuable human life into the world, help create in the world a loving family bond between you and your spouse, etc. On the basis of these objective values of the experience, you can know that the experience of having a child will be better than, say, the experience of being skinned alive. So even if you can't know the subjective value of the experience, you can assign a rough value to the experience on the basis of its objective value.

How the objective and subjective values of an experience interact to determine the overall value of an experience is a substantive matter for axiological theorizing and will include investigation of any organic unities among them with respect to the value of the experience overall. But the key point is that we cannot simply assume that if we don't know the subjective value of an experience but know its objective value, that we can't assign it a rough value overall. We might be able to.[24]

None of this is to deny that there could be some cases in which ignorance of the subjective value of an experience blocks knowing its overall value, even if you know its objective value. The overall value of *some* experiences, such as that involved in having a sumptuous meal, might turn mostly on the subjective value of how we experience them.[25] So if you don't know the subjective value of the experience, you can't assign an overall value to it. But the kinds of experiences that most plausibly figure in paradigmatic transformative choice are not like this. When you choose whether to gain sight or hearing, or whether to become a parent, *how it feels to you* is only one small factor in determining the rationality of the choice. What matters significantly more is the way things are, not how you experience them.

Two points help to show why this is so.

Suppose that the experience of having a child will make your life go objectively great—significantly better than it would go if you were to remain child-free. An evil demon, however, plays with your brain and makes the experience of having a child for you a drudge. Paul explicitly says that her interest is in veridical experiences only. But if you experience having a child as drudgery, *and* if that experience is veridical, then there is drudgery in your life—an objective disvalue that is entailed by your veridical experience. One way the subjective badness of an experience might seem to be important to the value of the experience overall, then, is by surreptitiously assuming the objective badness that subjectively bad *veridical* experiences entail. In order to isolate subjective value *per se*, we should assume that your experiences are *nonveridical*, and then ask whether ignorance of the intrinsic badness

[24] Just as the question of how some knowledge of what an experience is like might contribute to knowledge of what an experience is like overall is a substantive matter, as we saw in the last section.

[25] This can also be disputed; the value of the sumptuous meal may be primarily a matter of how its enjoyment objectively conduces to your well-being. But we don't need to take a stand on that question here.

of your nonveridical experience could block assigning an overall value to the experience on the basis of the objective values of that experience. Once we suppose that an experience does not reflect reality, but is like a dream of hallucination, then, I believe, the import of its subjective value significantly diminishes. If the experience of having a child will, objectively speaking, make your life significantly better, and if what matters to the choice is your well-being, then it seems plausible that you can assign a rough value to the experience despite not knowing the subjective value of your experience.

But what if the evil demon is especially cruel and makes your nonveridical experience of having a child not boring but hellacious torture? Again, the experience isn't veridical, but doesn't the possibility that the experience is torturous show that the subjective value of the experience can make a significant contribution to the overall value of having it? If you can't know whether the experience of having a child will be torturous, even if the experience will be nonveridical, doesn't that block assignment of a rough overall value to the experience on the basis of its objective value? To answer this question, we need once again to isolate the intrinsic subjective badness of the experience. The experience of torture, even if nonveridical, makes a significant contribution to the *objective* badness of your well-being. This is true for every subjectively good or bad experience. Suppose you have traumatic nightmares every night. Even if they don't reflect anything in reality, the unpleasantness of the experience can make your life go objectively worse. Similarly, an experience of hellacious torture, even if fabricated by an evil demon, can take an objective toll on your well-being. As many normative philosophers have pointed out, the objective goodness or badness of something can depend—causally, for instance—on your subjective experiences. So the subjective badness of an experience may amount to nothing more than the objective badness of it, where its objective badness depends in part on how the experience feels. Why shouldn't we think of the subjective badness of hellacious torture in this way? It is bad, not because of any intrinsic subjective badness but because of the way it objectively can harm your well-being.

More concessively, we might imagine a way to isolate the intrinsic subjective badness of an experience. Perhaps after the evil demon makes you experience, nonveridically, hellacious torture, God comes along and immediately expunges the experience from your life so that there is no objective harm from the experience. In this case, we might think that although there is no objective badness of the experience, the subjective badness of the experience remains. Now the question becomes, does the subjective badness of the experience block your assigning a rough overall value to the experience when what matters to the choice is your well-being? It would be strange to think that it does. By hypothesis, having a child will make your life objectively much better than it would otherwise be. What matters to the choice is how well your life goes. Even if what partly contributes to your well-being is the subjective value of your nonveridical

experiences, once we isolate this subjective value from the objective value that such subjective experiences can have, it's hard to see how the intrinsic subjective badness of hellacious torture—understood apart from the way it might objectively harm you—can block your assigning a rough overall value to the experience. Note that this is not simply to assume that well-being cannot be partly a matter of subjective value. It is only to underscore what many other philosophers have persuasively argued that a theory of well-being according to which the *only* thing that makes your life go well is having subjectively good experiences is wrongheaded.[26]

We can leave our theorizing there because, as I'll be arguing in the next section, in any case, the focus on *experiences*—whether their subjective or objective value—is misplaced to begin with. If that's right, then the issue of whether we can assign a rough value to having an experience based on its objective value alone becomes a side issue.

I've argued that transformative choices aren't about choosing the subjectively best 'what it's like' you can get. They are about choosing what you have most reason to choose, and what you have most reason to choose isn't simply about getting the subjectively best experience. Insofar as the value of the experience matters, it is not only the subjective but more significantly the objective value of the experience that matters. Since the objective value of an experience matters more than its intrinsic subjective value, you can plausibly assign a rough value to an experience even if you don't know its subjective value. And if you can assign an epistemically transformative experience a rough value, you can rationally choose it on that basis.

Nor should we think that the roughness of the value you can assign raises any special difficulties. Normative rational choice theory has the tools to represent rough value (e.g., Pettigrew 2014, Hsieh 2005). And even if two alternatives both have only rough value assignments, there are many ways to understand rational choice on the basis of those rough values. You might rationally treat them as roughly equally good or 'on a par.'[27] Since you can assign a value to epistemically transformative experiences,

[26] One reading of Paul's target is the view that the rationality of transformative choices turns solely on the subjective value of experiences. I think this is an untenable view of well-being and that even standard forms of expected utility theory need not embrace it. See, e.g., Broome 1991. Another reading is that subjective value is a contributor to the overall value of the experience. This is a more charitable reading of standard expected utility theory and is the reading I suppose in my argument that ignorance of subjective value does not entail ignorance of overall value. Note that this more charitable reading is compatible with preference-satisfaction accounts of well-being since preference-satisfaction accounts do not implausibly presuppose that it's only the subjective feel of preference satisfaction that makes your life go well but the fact of preference-satisfaction.

[27] There is a separate set of issues I don't have space to discuss here about how an assignment of rough values to each option or rough relative value to the set of options determines rational choice. So, for example, if we understand rough value as a closed interval range of reals, and two options are represented by overlapping intervals, which is it rational to choose? Economists and decision theorists have offered different answers to this and related questions.

the putative problem they pose for the assumption that rational choice is based on the value of the alternatives disappears.

2.1.3 Experience Isn't Primarily What Matters in Paradigmatic Transformative Choice

Epistemically transformative experiences, we've argued, don't pose any threat to standard approaches to rational choice. Strictly, they include only the sole de novo exercise of a new-to-human-kind basic capacity, not experiences typical of paradigmatic transformative choices. But even if we allow, for the sake of argument, that epistemically transformative experiences include de novo exercises of basic capacities that have been exercised before, such as seeing for the first time, and higher-level complex experiences, such as becoming a parent, we find that those experiences—assuming that they are ones we can't know what they're like—are nevertheless experiences to which we can assign a value relevant to rational choice. Since we can assign a value to epistemically transformative experiences, such experiences don't undermine the assumption that rational choice is based on the value of the alternatives.

There is a more serious worry about the focus on epistemically transformative experiences. Is transformative *experience* the right phenomenon on which to focus inquiry into transformative choices? Choices that can change who we are—e.g., a change in careers, divorce, giving a significant portion of our wealth to charity, and so on—aren't primarily about choosing the option that we think will deliver the best experience, whether 'best' is understood objectively or subjectively. As with any choice, a transformative choice is one in which you should do what you have most reason to do, and what you have most reason to do isn't typically a matter of how an *experience* will be for you. The objective and subjective value of an experience may of course be one relevant factor in determining what you have most reason to do, but it is arguably typically of only modest significance. To think otherwise would give experience a distorted importance in understanding choice.

Return to Mike May. At the outset of the paper, we suggested that May's choice was about whether to have an extended transformative experience—to experience life as sighted. But we can now see that there is a better understanding of his case. What mattered in May's choice was not simply the experience of being sighted—the subjective and objective value of that experience—but the goods (and bads) he would thereby gain (and lose) in his life, not just from the experience of being sighted but from *the fact of being sighted*. What transformed him was not the *experience* of being sighted but the fact of being sighted and its many upshots. Of course May could have been the sort of person who made his life a matter of chasing the best subjective experiences. But he would then have misunderstood the nature of his choice about whether to regain his sight.

The same goes for choosing whether to have a child. What matters in such a choice is not getting the objectively and subjectively best experience. Having a good experience is relevant to the choice, but the choice about whether to have a child is primarily one about whether to bring a being into your life and into the world, not about what you are to *experience*. Nozick taught us long ago, experiences aren't what matters in human life. What matters is how things are, not how we experience them. We choose between different ways things are to be, not between different experiences we might have. The focus on experience misses what is important in transformative choice—transformation isn't concerned with how we experience life but with how are lives actually are. The focus on experience makes transformative experience a phenomenon of creatures in Nozick's experience machine or of brains-in-vats. We should broaden our understanding of transformation so that not only experiences, but events in the world can change who we are.

For related reasons we might question Paul's suggested 'solution' to the putative problem raised by epistemic transformation. Recall that, according to Paul, since, by hypothesis, you can't assign value to an epistemically transformative experience, you can't rationally choose whether to have it. If such choices can be rational, then the natural conclusion to draw is that standard approaches to rational choice need revision. But Paul instead suggests that we reconceive the choice of whether to have a child not as one about whether to have an epistemically transformative experience but about whether to have a certain kind of knowledge, in particular, knowledge of what it's like to have the experience of having a child. Since we can assign value to knowing what something is like, such choices would raise no difficulty for standard approaches.

We might doubt Paul's otherwise interesting suggestion because, however odd it might seem to conceive of the choice of whether to have a child as a choice about whether to have a subjective experience, it's odder still to conceive of it as the choice of whether to have knowledge of what something is like. Recasting the choice in these terms seems to misunderstand the nature of the choice. Again, none of this is to deny that some agents might mistakenly think of their choices in these terms. But we should not build a theory of transformative choices on misunderstandings of what such choices involve. At any rate, it isn't a solution to a problem to find a related phenomenon in the neighborhood that doesn't raise the problem.

There are undoubtedly transformative choices about whether to have a transformative experience. You might be poised to ride Full Throttle at Six Flags, an experience that will transform you into someone who is no longer afraid to try activities many would consider terrifying, or be contemplating whether to try Vegemite, an experience that will transform you into a Vegemite fanatic. In these cases, the choice might plausibly be

one about whether to have a transformative experience.[28] And for some people, the experience of having a child can be transformative, changing them from a me-first person into someone who can care for another for her own sake. But this isn't to say that their choice of whether to have a child is a choice about whether to have a transformative *experience*, even if it is an experience that transforms them. Transformative choices aren't typically about which subjective feel would be best for you but about ways the world—including you as an agent—are to be. What you are choosing between when you make a transformative choice isn't typically experiences but ways your life might go.

2.2 Event-Based Transformative Choices More Broadly Understood

Transformative choices aren't typically choices between different experiences. Indeed, what matters in a transformative choice isn't simply getting the best subjective feel; what matters is the value of events in the world downstream from the choice, which may include experiences but need not.[29] What matters in May's choice about whether to see again are not only the objective and subjective values of the *experience* of seeing but also the objective *goods* (which we can characterize in terms of events) he will have in his life if he is sighted. Indeed, it makes sense to think that it was not the experience of seeing that primarily transformed him but other events, like communing with his wife over a beautiful sunset, responding to visual feedback from his children, and learning new skills that gave him greater opportunities that did the transforming work. So we should allow not only that events, broadly understood, are relevant to assessing the value of an option but that they can do the transformative work in a transformative choice.[30]

Once we move to events, however, we must abandon Paul's argument that transformative choices raise a problem for standard approaches because her argument crucially turns on the idea that certain kinds of *experiences*—epistemically transformative ones—preclude evaluation and therefore rational choice on the basis of their value. We turn instead to events more broadly, including experiences that aren't epistemically transformative and ask: Can the choice of an alternative that has as a downstream effect a transformative event—experiential or not—pose a problem for standard approaches to rational choice?

[28] As I've argued above, if these choices involve one's well-being—that is debatable—then they probably aren't choices *simply* about whether to have an experience but rather choices about whether to make one's life go a certain way. Choices simply about whether to have a certain experience are quite limited indeed.

[29] Paul herself sometimes slips into talk of nonexperiential events: e.g., "the process of having a child changes people" (2014, 90 and passim).

[30] Thanks to Louis Philippe for urging me to clarify the connection between what matters in a transformative choice and what events might do the transforming work.

Edna Ullman-Margalit (2006; 2007) proposes an event-based view of transformative choice.[31] She argues that a transformative choice is a choice to do something that changes your utility function. In particular, Ullman-Margalit thinks that events downstream from choice change your 'rationality base,' the beliefs and desires that form the basis of your reasons, so that your utilities after these downstream events are discontinuous with your utilities before them. You are transformed—you change who you are—by events downstream from choice that change your utility function (2006, 167–168).[32] Transformative choices, she thinks, are ones in which "the old 'rationality base' is replaced by a new" one (168). Paul shares Ullman-Margalit's view of transformation; she holds that when an epistemically transformative experience is personally transformative, your utilities before the experience are discontinuous with your utilities after it. So, like Ullman-Margalit, she thinks that personal transformation occurs when your utility function changes.

By way of example, Ullman-Margalit, like Paul, tells the story of a person contemplating whether to have a child. At one point in time he doesn't "want to become the 'boring type' who has children." But then he decides to have a child. The story continues: With time, "he did adopt the boring characteristics of his parent friends—but he was happy!" Prior to his having a child 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" (167, footnote 10). Transformative choices, Ullman-Margalit explains, "are choices that straddle two discontinuous personalities" (2007, 60). So a narcissist, for example, might be transformed by having a child because she has never before cared for anyone else for her own sake. Before having a child, the narcissist's utility function would value self-interested pursuits above all else, but after having a child, her utilities may reflect appreciation of the greater intrinsic value of the well-being of others over some of her pursuits. She is personally transformed by having a child since "what is rational for [her] to do beyond this point [of having a child] is different from the basis for the rationality assessment of [her] actions prior to that point" (2006, 168).

Ullman-Margalit goes on to suggest that such choices raise a problem for standard approaches to rational choice, and in particular, normative expected utility theory, because there is no stable utility function from which to determine the value of the alternatives. Before having a child,

[31] Strictly speaking, Ullman-Margalit is interested in cases she calls 'opting,' cases in which we 'make a leap of faith' in choice. But her discussion of 'opting' focuses on cases that have as their first feature that they are 'transformative.'

[32] Ullman-Margalit doesn't distinguish between the choice and its downstream effects, but it's clear from what she says that she assumes that what changes your utility function are effects downstream from choice.

the alternative of having a child has less value than being child-free, but after having a child, it has more value. Since there is no single, correct value of alternatives but only value relative to a set of preferences, utility function, or 'personality,' standard approaches to rational choice break down (167–168).[33] In response to the problem posed by such choices, Ullman-Margalit suggests that all the agent can do is make "a leap of faith," that is, arbitrarily 'opt' for one option over the other (169).

The challenge Ullman-Margalit poses, however, is a one that standard approaches—and normative expected utility theory in particular—can handle. Perhaps the most famous case in the neighborhood comes from Jon Elster's (1979) discussion of Ulysses and the sirens. Prior to hearing the sirens' song, Ulysses values his life and that of his men more than hearing the song. After hearing the song, his preferences support the reverse valuation. Elster suggests that Ulysses should exogenously bind his future irrational self to the ship's mast so that he cannot wreck his ship on the rocks when going mad from the sirens' song. Rational choice theory, broadly understood, has no difficulty with the case.

Of course Ulysses's case, since it involves future *irrational* preferences, is not strictly analogous to the challenge that Ullman-Margalit poses. Ullman-Margalit's choice of whether to have a child involves different sets of putatively *rational* preferences, that is, two perfectly rational but discontinuous and incompatible utility functions before and after the event of having the child. A closer analogy might be provided by Parfit's (1984) Russian nobleman. As a young man, Parfit tells us, the nobleman is a socialist who wants his old, richer, future self to distribute all of his wealth among the peasantry. But his older, conservative self prefers to keep most of his wealth for himself. It's not irrational—and we can suppose, not immoral—for the aged nobleman to keep most his wealth, after having discharged whatever duties he might have to be charitable, but nor is it irrational for him to give it all away as his younger self would wish. We just have two perfectly rational but discontinuous and incompatible utility functions before and after life's intervening events. Parfit suggests a solution akin to Elster's—the young nobleman should exogenously bind his future self, in this case, Parfit suggests, by creating a contract that only his wife can revoke and getting his wife to agree never to revoke it. This is because, Parfit seems to suggest, the younger self is the 'real' self just as the 'real' Ulysses is the one not driven to madness by the siren's song.[34]

[33] Paul makes the same claim about transformative choices, except that her concern is with epistemically transformative experiences that lead to a change in your utility function while Ullman-Margalit has a broader view of events that may lead to a change in your utility function. Since Paul's claim is an instance of Ullman-Margalit's, my discussion of Ullman-Margalit is intended also to be a discussion of Paul's similar claim.

[34] Christine Korsgaard (2009, 202) criticizes Parfit's solution as requiring the nobleman to put his wife in an "impossible position" because she must wrong either her young or her older husband.

But when you choose to have a child, your child-free self needn't be more 'you' than your parent self. You have two different sets of preferences, both of which reflect different 'you's' at different stages in your life. And both, we can suppose, are rational both at the time that they are had and from the time of choice.[35] So how are you rationally to choose whether to have a child when your preferences may radically change after the events involved in having a child from what they are before you have a child?

One solution is to take your present preferences as the basis for determining your rational choice. It needn't be that your present self is the 'real' you; instead it could just be a default rational principle that you must choose according to your present utility base rather than one you expect to have in the future. Note that privileging one's present preference profile would be compatible with those preferences reflecting knowledge of forseeable future facts, including your future preferences.

Suppose that, without thinking about how having a child might change your preferences, you now prefer to remain child-free, primarily because you are concerned about the impact being a parent might have on your career. You then read Ullman-Margalit and Paul, and you start to think about whether your preferences will reverse themselves if you end up having a child. You talk to your parent friends who didn't want children but ended up having them and take note of their post-child preferences, look up empirical data on the number of parents whose attitudes toward their work changed after having a child, research neurological studies claiming that post-child-birth, you will likely have a hormonal imbalance that contributes to a desire to have yet another child, and so on. While you don't know whether your preferences will change after having a child, you know that it might. You can then take this fact into account in forming your current preferences. The fact that you might be ecstatic about having a child after having one might be grounds now for you to prefer to take the risk of having a child even though the thought, now, of having a child fills you with anxiety and dread. Or that fact might simply change the strength of your current preference to remain child-free, so that you are closer than you were before to preferring to have a child, and that preference could later reverse as new events unfold in your life. Or the fact that you might be miserable with a child, but self-delusional, might strengthen your preference to remain child-free, if, for example, you now strongly prefer never to be self-delusional.

[35] The claim that both sets of preferences can be rational at the time of choice might be doubted. In discussing Parfit's Russian nobleman, for instance, Christine Korsgaard argues that the young nobleman must treat the preferences of his future self as irrational (2009, 202–204). I have argued elsewhere that two incompatible evaluative orderings according to 'given' values or reasons can both be rational if they are 'on a par' (e.g., Chang 2013b). If you don't believe in parity, then the two incompatible orderings can also be understood as different 'sharpenings' of a (nonsemantic) indeterminacy in which one option is better supported by rationality. In any case, the more interesting version of Ullman-Margalit's and Paul's case assumes that both can be rational at the point of choice.

Appealing to your fully-informed present preferences—informed by your knowledge of your future preferences—to determine rational choice is in fact the default view of standard rational choice theory. After all, very few people live and die with continuous utilities over the course of their lives. What Ullman-Margalit and Paul usefully point out is that, although typically a person's utility function changes slowly over time, it can also change quickly and dramatically by an event downstream from a choice. And if you know that your utilities might change dramatically in this way, the question arises, on what basis should you determine what it is rational for you to do? Standard rational choice theory has an answer: choose on the basis of your present preferences—informed by what you reasonably believe your future preferences will be. Of course the kind of normative weight you give to your future preferences when forming your present ones will depend on what your present preferences are. In the choice of whether to become a parent, for example, you might presently be extremely risk averse or you might be a daredevil. Your current attitudes will affect what normative role information about your future preference profile will have for you. But that is par for the course.

There are other ways rational choice theory could deal with choices about whether to undergo events that will change your utility function. Such theories might posit a 'master' utility function—the function your ideal self would have if it knew all the facts, present and future, relevant to any choice you might make in your life. This master utility function could then order the two sets of preferences with respect to any given choice. If, for example, the master utility function favors the ordering, (have a child, remain child-free), then, even if your present preferences favor remaining child-free, since the preferences you would have after you have a child are better—presumably with respect to your well-being—than your present preferences, the rational thing to do is to follow not your present preferences, but your future preferences. This is the rational thing to do because your future preferences reflect the preferences of your master utility function.

Or, rational choice theory could posit a principle of rationality according to which when you know that one of the options for choice will change your utility function, you should just 'wait and see.' Such a principle might counsel that you put off the choice if possible, or take an incremental approach to the choice by breaking it down into smaller sub-choices that can made over time, which could have the effect of turning transformative choices into ordinary choices by which one is transformed over time.[36]

The important point for our purposes is that standard approaches to rational choice have ways of dealing with choices with downstream effects that lead to a change in your utility function, personality, or point of view

[36] See also Ullmann-Margalit 2006 who suggests that one strategy for dealing with cases of transformative choice is to break them down into smaller choices.

without compromising either of its two fundamental assumptions. While your personality may change, the rationality of your choice is based on the value of the alternatives. And while the value of the alternatives may be relative to different personalities, the rationality of the choice is based on the value of the alternatives nonetheless. Until these approaches are shown to be untenable, the problem supposedly posed by transformative events downstream from choice in evaluating the value of the alternatives is one that standard approaches to rational choice can solve.

3 Choice-Based Transformative Choices

As we've seen, the first assumption of standard approaches to rational choice, namely, that rational choice is determined by the value of the alternatives, can be kept intact even in the face of choices about whether to undergo events that will change our utility functions, personalities, or point of view. While it's plausible that many transformative choices will be event-based, such choices don't seem to raise any genuine challenge to standard approaches. This is not to say that there are no such choices, but only to suggest that perhaps they aren't the most interesting kind of transformative choice around.

In the rest of this paper, I want to begin to lay the groundwork for an account of choice-based transformative choices, choices in which *the choice itself* does the transforming work. As we'll see, the account of choice-based transformative choices challenges the second assumption of standard approaches, namely, that choice cannot be *that in virtue of which* we have reasons or it's rational for us to choose something.

But first we need to try to clarify what we mean by 'transformation.' Both Ullman-Margalit and Paul suggest that transformation is a change in your utility function, personality, or point of view. I agree, but I want to tweak and expand their expected-utility-theory-based idea of transformation, in part so that it is not wedded to any particular substantive view about how normativity is to be modeled. I'll use the term 'reasons' to indicate considerations that count in favor of an alternative, whatever those considerations might be—preferences, evaluative facts, duties, excellences, and so on. (The points I want to make can also be put in terms of 'values' and even pro tanto 'duties' (with appropriate bells and whistles) but I will stick with 'reasons' since, at least to my ears, that is the most neutral-sounding normative term.) As I'll be suggesting, you are transformed—change who you are—whenever your reasons change in a way that alters your normative character. And in choice-based transformative choice, you alter your normative character in a distinctive way: you change 'who you are'—change the reasons of your normative character—*in virtue of* your choice.

3.1 Transformation

A transformative choice changes 'who you are.' But what might this involve?

First, we ask, Who is transformed? A *person* gets transformed in some way, but not into a nonperson—not into a bat, or squirrel, or vampire (if vampires aren't people, teen fiction notwithstanding). So a person must remain before and after a transformative choice. Furthermore, *you*, whatever you metaphysically are, must remain before and after a transformative choice. That's how we can intelligibly say that *you* have undergone a transformation. Transformative choices don't alter your numerical identity or your personhood; they merely change the way *you*, the person, are from how *you*, the person, were before the choice. A transformation that meets both these conditions is what we might call a *personal* transformation. We can understand the transformation of choice-based transformative choices as personal in this sense.

This way of understanding transformation ensures that transformative choices include paradigmatic cases of transformative choices such as whether to regain one's vision, have a child, change careers, and so on, that are part of ordinary human life. Excluded are 'radical transformative choices,' choices that either transform you from a *person* into a non-person or *you* into something that isn't you—such as choices of whether to turn into a tree or cockroach or ghostly spirit. Such choices no doubt raise interesting questions for the rationality of choice but aren't relevant here.[37] If being human is inessential both to being you and to being a person, transformative choices will include those about whether to undergo some significant enhancement, such as one that would enable you to live for 500 years or would allow you to jump to the moon, that might transform you from being human to being 'superhuman.'

Second, we ask, What is the feature of *you, the person* that gets transformed in a transformative choice? As we've seen, both Paul and Ullman-Margalit understand personal transformation as a change in your utility function—your preferences over alternatives is different before and after a transformative event. Both of them plausibly understand transformation as a *normative* phenomenon. Although May, upon regaining sight, changes dramatically in nonnormative ways, these nonnormative changes aren't transformative in the sense of interest unless they subvene or ground a *normative* change. As I will put it, you change 'who you are' when you change your *reasons*, which includes a change in the strength of your reasons as well as coming to have new reasons you didn't have before. After

[37] See Kemp 2015 for a defense of the idea that what I am here calling 'radical' transformation is not something you can undertake or will but is something that happens to you. The transformations of interest here are not radical in this way since there is a you—whatever that entails—that is a person, to whom we can attribute and who can, in principle, undertake the transformation.

his transformation, May's reasons to read in Braille are weaker than they were before, and he has normative reasons to go to museums that he didn't have before. The transformation of transformative choices, I suggest, is essentially a matter of changing your normative reasons.

The transformation of interest is also objective; the reasons you come to have are reasons you *in fact* come to have, not reasons you merely believe you have. Sometimes we can be transformed and realize it much later. You might choose to become a parent and only after a few months of changing diapers realize that you are no longer the self-centered, me-first person you were before. Similarly, you can falsely believe that you are transformed—say, by attendance at an awesome rock concert—but in fact you are much as you always were. This implies that in a transformative choice, you may not know that an alternative you choose will transform you.

Now, for transformation to take place, the change in your reasons can't be a change in any old reasons you happen to have. The choice needs to change 'who you are.' I suggest that 'who you are' is understood normatively, as your normative character or normative identity. 'Who you are' is who you are *qua* normative agent, and the reasons that belong to 'who you are' are the reasons that make up your *normative character*, roughly your normative personality, or 'the sort of person' you are, normatively speaking. Transformation, then, is a matter of changing your normative character, by having reasons that make the normative you different from how you were before.

We'll have more to say about normative character shortly. But one possible misunderstanding should be put aside here. It might be thought that the normative you is your 'deep self' or, as it is sometimes put, 'who you really are, deep down inside.'[38] But I think identifying your normative character with your deep self, assuming that this idea is even coherent, would be a mistake. This is because transformation may be a much shallower phenomenon than changing who you 'really are, deep down inside.'

Suppose, for instance, that your deep self is given by the reasons that play an organizing, structuring, executive, or some higher-order role vis-à-vis your other reasons. They might be general Bratmanian 'self-governing' policies such as 'be consistent' or 'always do what's best for me,' or substantively thick Aristotelian 'master ends' such as happiness or flourishing, that structure your other reasons.

But your higher order policies aren't plausibly the reasons that determine your normative character. This is for two reasons. First, we need a story as to why your actual policies determine your normative character (or for that matter, your deep self) as opposed merely to being the policies under which you happen—perhaps because of some brainwashing—to be

[38] I assume for the sake of argument that the idea of 'who you really are, deep down inside' is coherent.

operating. The Bratmanian story, according to which one's self-governing policies have special status because they ensure one's metaphysical identity over time as an agent, won't help us here because we have put aside 'radical' transformation, the transformation that alters your metaphysical identity. In any case, it isn't plausible to think that whatever policies determine you to be a single metaphysical agent acting over time are the same policies that make up your normative character, since your normative character is rather more specific.

Second, since transformative choices change your normative character, the reasons that determine that character should be reasons that are changed in a transformative choice. But paradigmatic transformative choices don't require changes in your general policies, master ends, or higher-order normative principles. You might have a general policy to do whatever you believe is morally right, to love thy neighbor, or even to look out only for yourself and your loved ones. When you choose to have a child, live in the countryside or change careers, your general policies and master ends may well remain intact. It's not that, having moved to the countryside, you become morally reprobate. And yet a move to a rural life can change your normative character. You are no longer the always-overcommitted multi-tasker who never stops to smell the roses. You're now the sort of normative character who sits on a porch swing, doing nothing while enjoying the sunset. In short, can change who you are, normatively speaking, by becoming a parent, or a country-dweller, or a graveyard-shifter, by changing your reasons without changing the reasons of your deep self.

There are other possible suggestions about the reasons that determine your deep self, but I think all of them will fail as reasons that determine your normative character. This is because your normative character is shallower than 'who you really are, deep down inside.' Some people seem much the same, deep down, throughout their lives. Nevertheless, they can have different normative characters at different point in their lives. Parfit's Russian nobleman might be such a person. Nelson Mandela and Mother Teresa might be others.

When you make a transformative choice, you can make yourself into a person who would make you cringe at some other point in your life. Such transformations might involve changes in the deep self. But they need not.

So far we've said that transformative choices involve changing your reasons, and in particular, the reasons that determine your normative character. But there's a third, crucial question we should ask about transformation. How does a transformative choice change your reasons? Since the most interesting change in your reasons is coming to have new reasons you didn't have before, I'll focus on that case.

Standard approaches to rational choice explain how you come to have reasons by appealing to something essentially independent of your choice—typically some relation between an alternative and your mental states, a normative fact about the goodness of alternative, or the normative fact

that whatever is the reason is a reason. To keep things simple, I'll focus on explanations of how you come to have reasons that appeal to normative facts, though the same points can be made in terms of preferences and other mental states.

An example will help. Suppose you could have a burger or salad for dinner. You choose to have the burger. As a consequence of choosing to have it, you order a burger. Your choice of the burger, and the subsequent action of ordering it, changes the facts going forward. You've now ordered the burger, you'll be having red meat for dinner, and so on. The fact that you've ordered a burger (along with pre-existing facts about your tastes and so on) gives you a reason you didn't have before, a reason, say, order a cold Heineken (and not a hot eggnog) as a dinner beverage.

Why does the fact that you've ordered a burger give you a reason? That is, *in virtue of what* does the fact that you've ordered a burger give you a reason to order a beer? According to standard approaches, the answer is given by a normative fact, viz., *If you order a burger (in such-and-such circumstances), you have a reason to order a cold beer.* A normative fact that connects downstream effects of your choice with reasons is that in virtue of which you have those reasons. We might call such facts 'grounding' normative facts. When you satisfy the antecedent of such facts, you thereby 'trigger' the consequent reasons of such facts.

There are many other examples. There is a normative that, *If you punch someone in the nose (in such-and-such cirumcumstances), then you have a reason to make amends.* So if you fulfill the antecedent condition of this normative fact, that is, if you punch someone in the nose, you have a reason to make amends. You have that reason in virtue of the grounding normative fact according to which punching someone in the nose 'triggers' a reason to make amends. If you chose to have a child, as a downstream effect of so choosing you may have a child. If you have a child, then you have a reason to nurture and care for her that you didn't have before. You have this reason in virtue of the normative fact that, *If you have a child (in such-and-such circumstances), then you have a reason to nurture and care for her.* Again, your having a child 'triggers' a reason to nurture and care for her. According to standard approaches, you come to have reasons not in virtue of your choosing anything, but in virtue of a grounding normative fact that connects downstream effects of your choice with reasons.

Note that according to this standard explanation of how you come to have reasons, choice *per se* may be normatively irrelevant to your having new reasons. Even if you have a child by accident and not by choice, the fact that you have one triggers reasons to nurture and care for her in virtue of the same (or similar) normative fact.[39] If transformative choices change your reasons in this way, then choice may be irrelevant in explaining both

[39] Whether it's the same grounding normative fact turns on a matter of substance, viz., whether how you come to have a child—by choice or not—affects which reasons having a child triggers. In the same way, it's a substantive matter whether the fact of choice is a relevant fact that

what transforms you and how you are transformed. All the work is done by effects downstream from choice—the events, broadly construed—that trigger reasons in virtue of a grounding normative fact connecting those events to those reasons. Event-based transformative choices, then, can give you new reasons in the standard way. If we understand transformative choices on the model of event-based transformation, we take the 'choice' out of 'transformative choice.'

Choice-based transformative choices, by contrast, put the 'choice' back into transformative choices. Choice is not only what transforms you but how are you transformed; that is, choice is *that in virtue of which* you come to have the reasons of transformation. You change your normative character through choice itself.

3.2 An Account of Choice-Based Transformative Choices

The idea that we come to have new reasons in virtue of choice itself may seem unpromising. After all, standard approaches assume that choice itself can't ground your having reasons for obvious reasons: if choice can ground your having reasons, you can choose your way into any reasons you want. It would seem, then, that rejecting this assumption of standard approaches leads to intolerable bootstrapping.

But choice-based views of transformative choices can help themselves to a general theory of the grounds of normativity that avoids this result. Elsewhere, I've proposed that reasons can have one of two different grounds; for any consideration that is a reason there are two different sorts of consideration that can make it a reason, one a normative fact and the other an act of will. Correspondingly, because a given consideration can have two different sources, it can be two different reasons.[40] The view is 'hybrid voluntarism,' so called because it understands the grounds of practical normativity as a hybrid of two kinds of considerations that can make something a reason. Some reasons are grounded in something other than our wills—they are 'given' reasons because they are given to us—while other reasons are grounded in our wills—they are 'will-based' or 'voluntarist' because we create them through an act of willing.[41]

---

triggers reasons. In some cases, surely it is. But choice in such cases is not a ground but a triggering fact.

[40] Alternatively, we might say that the overall normative strength of a single—and single kind of—reason may have two different sources. To draw the starkest contrast between the two sources of normativity, I assume that reasons are individuated by both 'content' and source and so we have two 'kinds' of reasons, given and will-based. The substance of the view, however, can be put equivalently in terms of the normative strength of a single kind of reason having two different sources.

[41] See Chang 2009, 2013a,b. Talk in terms of new will-based reasons helps to underscore the fact that the source of the normativity of the reason is in the will, but the view can also be equivalently stated in terms of the will being that in virtue of which a given reason has greater (or lesser) strength than it had before.

Suppose, for instance, that your child wants a new toy. Why is this fact a reason to buy her one—in virtue of what is it a reason, assuming that it is? One answer is that satisfying her desires makes her feel safe and happy, which is good or valuable in some way.[42] The fact that satisfying her desire is good is that in virtue of which the fact that she wants a new toy is a reason for you to buy her one. Notice that this explanation makes no appeal to your will. What makes something a reason is something other than your will—its being a reason is 'given' to you, not willed by you. It's given by the goodness of buying her a new toy. So the fact that she wants a new toy is a *given* reason for you to buy her one.

A reason can also be a reason because of an act of your will. Suppose that you *will* that your child's needs and interests are normative for you, or, as I will put it, you *commit* to her needs and interests by putting your will—your very self as a normative agent—behind those needs and interests. By committing to her in this way, your commitment can be *that in virtue of which* her desire for a new toy is a reason for you to give her one. Satisfying her desire for a new toy will serve her needs and interests by making her feel safe and happy. Your commitment to her needs and interests is thus what makes her desire for a new toy a reason for you. By committing to something, you can make something a reason for you to perform some action.

Notice that the same fact—that your child wants a new toy—can be the 'content' of two different reasons, distinguished by what makes the fact a reason. A fact can be a reason to do something in virtue of a normative fact, such that doing that thing is good or that the fact is a reason to do it, or it can be a reason to do something in virtue of your act of will—your commitment to something. You can have a reason to give her a new toy in virtue of the fact that doing so would be good, but you can also have a reason to give her a new toy in virtue of your commitment to her needs and interests.

Now if we left things there, will-based reasons would indeed lead to intolerable bootstrapping. Perhaps buying your child a new toy would be bad for her, but you could nevertheless bootstrap your way into having reasons to buy her a cornucopia of toys by committing to satisfying her every desire.

Hybrid voluntarism maintains that there is a hierarchy among your given and will-based reasons. It holds that your given reasons, so long as they don't 'run out,' always determine what you have most reason to do. So if you have most given reason not to buy your child a new toy—if it would be bad for her, for instance—you shouldn't buy her one, all things considered. When your given reasons 'run out,' however, your will-based reasons can determine what you should do, all things considered.

[42] Or that buying her a new toy will satisfy your desire that she feel safe and happy. Kate Manne offers an interesting twist on this answer: buying her a new toy will satisfy *the child's* desire to feel safe and happy. See Manne Forthcoming.

Sometimes your given reasons will 'run out.' They can 'run out' in one of two ways, but for simplicity's sake we can focus on just one of them. Your given reasons 'run out' if the options are comparable but neither option is better than the other and nor are they equally good. That is, they are *on a par* with respect to what matters in the choice. Or, put equivalently in terms of reasons: the strengths of the reasons are comparable, and they don't favor one option over the other and nor are they of equal strength. Your reasons are *on a par*. Two items are on a par if they are comparable and yet neither is better than the other and nor are they equally good.[43]

If your given reasons are on a par, they don't tell you what you should do. You now have the normative power to commit to one of the options, thereby creating a will-based reason in favor of it.[44] The silence of your given reasons is one sense in which when your given reasons 'run out,' you can create will-based reasons in favor of an alternative. Your will-based reasons can, in turn, make it the case that you have most all-things-considered reasons to choose one option over the other.

Return to the choice of whether to have a child. Suppose, for simplicity, that what matters in the choice is simply your well-being (perhaps because you are a single parent, an unborn child has no well-being, you know that your child will be carbon neutral, etc.). Suppose that, with respect to what would make your life go best, having a child is better in some ways, remaining child-free is better is some other ways, and having a child is neither better nor worse than remaining child-free, overall. There are just different tradeoffs to be made whichever course you take. The options are on a par with respect to your given reasons. Typically options that are on a par will bear very different values while nevertheless being in the same neighborhood of overall relevant value.

According to hybrid voluntarism, if your given reasons are on a par, you can *commit* to some feature of being a parent that counts in favor of being a parent and thereby create for yourself a new will-based reason to have a child. Your new will-based reason, then, may make it true that you have most reason, all things considered—considering both given and will-based reasons—to have a child.

This hierarchy between given and will-based reasons eliminates the worry that you can bootstrap your way into any reasons you like. You can create will-based reasons only when your given reasons have run out. If you have most given reasons to remain child-free, say, because having a child will make you suicidal, then you can't will yourself reasons that make it the case that, all things considered, you should have a child. Hybrid voluntarism holds that your given reasons have 'first dibs' in determining what you should do.

[43] See Chang 2002. I don't have space here to discuss the possibility that the options are incomparable. I address that possibility in Chang 2012, Forthcoming.

[44] You can also create a will-based reason against choosing an option by committing to not having some feature of one of the alternatives in your life.

When you commit to some feature of having a child, such as forming a parent-child bond, you thereby choose to become a parent in a *thick* sense of 'choose': you select an alternative by putting your very agency behind it. By committing to forming a parent-child bond, you now 'stand for' forming such a bond—you, your very self as an agent, are *for* forming such a bond.[45] You can also choose to become a parent in the *thin* sense of 'choose'; you choose in a *thin* sense whenever you select an alternative or merely intend to go for it without throwing your very agency behind the alternative. The thick sense of choice involves an act of will, a commitment of your normative self; the thin sense doesn't. Crucially, choice-based transformative choices involve choice in the thick sense: when you choose an option, you are committing to that option by putting your agency behind its features.[46]

We can now see how in choice-based transformative choices your choice can be both what transforms you and that in virtue of which you are transformed. In deciding whether to have a child, by hypothesis, the given reasons are on a par. You have the normative power to commit to one of the options or one of its features. You might commit to forming a parent-child attachment. That commitment just is choosing to have a child in the thick sense. That commitment then creates new will-based reasons for you to have a child, that is, your commitment is that in virtue of which you now have a new will-based reason to have a child. Your new will-based reason then interacts with your other, given, reasons and guides your choice in the thin sense. You may now have most all things considered reasons to choose to have a child. Your new will-based reason transforms you because it is a reason that determines your normative character. You are now the sort of person who has most all things considered reasons to have a child. Before the choice you were the sort of person for whom the reasons for having a child and remaining child were on a par. By choosing, you change the reasons that determine your normative character.

Your normative character then, that is, the normative you, is given by all the will-based reasons you have before the transformative choice. Those will-based reasons are the reasons that make you 'who you are,' normatively speaking. When you create a new will-based reason for yourself, you transform yourself into someone who has reasons she didn't have before. These new reasons are reasons you have created for yourself through your commitment to something. In this way, through choice, you transform yourself from one normative character into another.

[45] For further discussion of this form of internal commitment that can be the source of reasons, see Chang 2013a.

[46] Commitment needn't be a conscious, deliberate steeling of the will. We can be unaware of and even deny commitments we have in fact made. Think of the man who has a self-conception as a swinging bachelor but who would sacrifice his life to protect the needs and interests of the person he's made a life with for the last several decades.

The way you come to have new will-based reasons on this account, is incompatible with the second assumption of standard approaches to rational choice. You come to have reasons in virtue of choice itself, not in virtue of something essentially independent of choice. And to think that the choice itself could play the role of an event downstream from choice—that is, as something that fulfills the antecedent condition of a normative fact in virtue of which when you make the choice, you have a reason, would lead again to intolerable bootstrapping; there are no normative facts of the form, *If you choose x, then you have a reason to pursue x*. Thus, choice-based transformative choices, as I've proposed we understand them, require fundamental revision of standard approaches to rational choice.

Now, when faced with a transformative choice in which the alternatives are on a par, you might *not* commit to one alternative or the other. You might choose in the thin sense without choosing in the thick sense. If you choose without committing, you either 'drift' into an alternative or 'plump' for it. Drifting encompasses many different ways of selecting an alternative. You could drift by selecting by omission (you never seriously consider whether to have a child), of by taking the path of least resistance (all your friends are becoming parents, so it might be easiest for you simply to follow the crowd), or by allowing some emotion, such as fear, determine what you will in fact do (you are terrified that when you are on your deathbed that you will profoundly regret never having a child). When you drift into an option, you don't put your agency behind it or its features but let reasons of the world cause you to take one path rather than another.

You can also, instead of committing or drifting, 'plump' for an alternative, that is, arbitrarily select one option over another for no reason at all.[47] Choosing in the thin sense by drifting or plumping can, of course, lead to downstream events that are transformative. Importantly, however, both drifting and plumping are ways transformation can *happen to you* rather than be *by you*. When you choose in the thick sense, by committing to an option, you transform yourself by creating for yourself new reasons you didn't have before. Your choice does the transforming work.

It's worth noting that according to Ullman-Margalit and Paul's event-based views of transformative choice, at the point of choice, there is no comparative fact about the merits of the alternatives. Paul suggests that the alternatives are what I have called 'noncomparable' (2014, 102, footnote 55). Two items, such as fried eggs and the number nine, are noncomparable with respect to tastiness when there is a formal precondition for the possibility of comparability that is not met. In the case of fried eggs and the number nine, the formal condition is that the 'covering value,' that

[47] Plumping is not the same as 'picking' between two equally good alternatives; note that having a child and remaining child-free are, by hypothesis, not equally good or equally well-supported by reasons. When you pick between equally good alternatives, you choose for the reason that the options are equally good and it's better to select one than to remain poised between them like Buridan's ass.

is, the respect in terms of which the items are being compared, namely 'tastiness,' fails to cover both items: the number nine is not the kind of thing that can be tasty. Paul thinks that alternatives in a transformative choice are noncomparable because one of them involves an epistemically transformative experience (that is also personally transformative), which precludes assigning a value to it. Since there can be no value assigned to any alternative that is epistemically transformative, a formal precondition for the possibility of comparing them, namely that each alternative has some value, isn't met. We've seen some reasons to doubt, however, whether epistemically transformative experiences preclude assigning a value to an alternative.

Ullman-Margalit, by contrast, seems to think that there is no comparative fact about the merits of the alternatives at the point of choice because their value is relative to the utility function that exists at the relevant time. Each alternative has value, but the value is different depending on when the evaluation takes place. Before you have a child, it's better for you to remain child-free since your utility function favors that option. After you have a child, it's rational for you to have the child—your utilities after the event of having a child favor your having the child. But it's puzzling why we should relativize the rationality of our actions to the utility function that exists at the time. As a general principle, holding that the rationality of an action is relative to the utility function held at the time leads to untoward consequences. It would make, for example, a course of action where you undermine all your other aims and projects rational so long as your utility function at the time favored doing so. And if we tried to constrain your utility function, such as by restricting the ways in which you can come to have it, we will end up denying as rational certain intuitively rational choices to undergo certain transformations.

In any case, there is a more worrying puzzle. Both Ullman-Margalit and Paul seem to think that at the point of choice, both alternatives are *rational*. If they didn't think this, then transformative choices would amount to an instance of Elster's Ulysses case: at the point of choice, one alternative is irrational, and so one should bind oneself in a way so as not to choose it. Standard approaches have no difficulty accounting for the rationality of such choices. If, instead, the alternatives are both rational at the point of choice, then how could there be no comparative fact about how the alternatives relate? Paul explicitly denies that the alternatives are *incomparable*, and I think she is right to do so.[48] Ullman-Margalit doesn't say much about how the alternatives in a transformative choice relate, but she seems, like Paul, to assume that they aren't equally good. If the alternatives are neither incomparable, nor equally good, and one is not better than the other but both are rational, then what relation holds between them? They are, I suggest, on a par.

[48] See Chang 2012, Forthcoming.

The options involved in transformative choices can evaluatively relate in one of three ways.[49] First, one option might be better (or good enough), in which case it is rational for you to choose it. You might, for instance, choose not to wander into a dangerous neighborhood where it is likely that you will be set upon by a gang of thugs. You know that being beaten up will be traumatic and transformative—and *worse* than not being beaten up. So you rationally choose not to have the transformative event in your life because having it is worse than the alternative. Or, to take the reverse case, you might face a choice of whether to claim your lottery winnings. Winning the lottery is a mixed bag, but overall, it's better to have the money than to forgo it. If you choose to claim your winnings, you'll undergo transformative events downstream from that choice. You rationally choose to have a transformative event in your life because it's better than the alternative.

Second, your options might be equally good with respect to what matters in the choice. This will, however, be rare in cases of transformative choice. If one option isn't better than the other, then it's unlikely that they will be equally good. It might be that for you, for example, having a child is neither better nor worse than remaining child free with respect to you well-being. Does it follow that they are equally good? We can run a test to determine whether they are. If we improve (or detract) from one of the options a bit—perhaps we throw in a part-time nanny if you have a child—does it necessarily follow that having a child is now better than remaining child-free. We can imagine a scenario where it doesn't necessarily follow. If this is right, then neither option is better than the other and nor are they equally good. For if they're equally good, an improvement in one, even if small, must make the improved option better (Chang 2002). But this doesn't plausibly hold for transformative choices between options, neither of which is better than the other. Moreover, if the options are equally good, it would be intrinsically rational for you to flip a coin between them. It seems odd to think that it's intrinsically rational for you to flip a coin in deciding whether to become a parent (Chang 2012).

Third, the options might be on a par. Typically, items are on a par when, with respect to some things that matter in the choice, one option is better, with respect to other things that matter, the other option is better, and yet neither is better than the other overall. For many people, remaining child-free and becoming a parent will be on a par. I've had more than one philosopher-friend say to me: 'Having a child might never allow me to achieve what I want to in my work, on the other hand, it might allow me to have a life enriched in ways that I can't now forsee. I know that whichever option I choose, I will be transformed into a different sort of person than how I am now.' If the alternatives are on a par, it would be a mistake to

[49] There are also the options that they are incomparable or indeterminately comparable due to vagueness in language. I argue against this options elsewhere. See Chang 2002, 2012, Forthcoming.

continue search for reasons in the world that might make it the case that one is better; if they're on a par, there are no such reasons. Instead, since your given reasons are on a par, you have the normative power to choose one of the options, in the thick sense, by committing to one of its distinctive features. When you commit to an option, you create a will-based reason for you to choose it. You've now changed your reasons—you used to be the sort of person for whom having a child and remaining child-free were on a par. After committing to remaining child-free, you thereby become the sort of person for whom there is most reason to remain child-free. Your commitment changes your normative character.

Choice-based transformative choices capture paradigmatic cases of transformative choice. We've already looked the case of choosing whether to have a child. What about choosing whether to have a cochlear earplant or an operation that will restore your vision? Return to Mike May. Being unsighted involves having certain goods that are precluded if you are sighted, and vice versa, and perhaps those goods are, overall, in the same neighborhood of value although very different in value.[50] Perhaps May faced a choice in which his options were on a par. On the choice-based view of transformative choice favored here, May could transform himself by committing to the goods of being sighted, and thereby give himself most reason to undergo the operation that would restore his sight. He could transform himself by his choice by creating for himself new reasons he didn't have before in virtue of that choice, and perhaps thereby making it true that he is someone who now has most reasons to become a sighted person.

Thus when the Hollywood plastic surgeon gives up her luxurious lifestyle to volunteer in a war torn region, she commits to features of being a volunteer which then creates new will-based reasons to live differently than she has before. These new reasons change her normative character. When Gauguin chooses to abandon his family for his art, he creates a will-based reason to pursue his art and changed his normative character decisively into what it was. When Sophie, faced with the choice about which of her children to send to the gas chambers, creates a will-based reason to save Jan rather than Eva. It's this commitment that arguably tortures her for the rest of her life. And when the philosophy chairperson of the story with which we began this paper chooses wisdom over truth and money, the goods are on a par. But by committing to wisdom, she now is the sort of person for whom wisdom matters more than truth or money. What makes

[50] It might be difficult for sighted people to understand the depth of the goods gained by being congenitally blind. One grows up with different capacities that sighted people cannot share. Of course, once you are sighted, it seems much worse to be blind, especially if you cannot then achieve the special capacities of the congenitally blind. May was born sighted, became blinded at a very young age, and had the opportunity to become sighted at a late age, so his case is correspondingly more difficult.

the story droll is the suggestion that for such a person, money should be chosen instead of wisdom.

Again, this is not to say that there aren't transformative choices of the event-based kind. You can be transformed from a fun-loving, happy-go-lucky person into someone withdrawn and fearful by a traumatic event downstream from a choice to, say, pursue a dangerous career. Some transformative choices aren't between options that are on a par, but between options one of which is better than the other. In such choices, there is no problem for standard approaches since the rational thing to do is to choose the better option. Indeed, I suspect that many paradigmatic transformative choices are 'mixed'—involving both choice-based and event-based transformation. My suggestion is only that in cases in which it might appear that an event does the transforming work, the choice itself may also transform the agent. Focusing only on event-based transformative choices leads us to overlook choice-based transformation. And only the latter raises problems for standard approaches to rational choice.

I'll end the paper by pointing out two upshots that the account of choice-based transformative choices favored here have for our thinking about rational choice. First, the account of transformative choices I've offered is neutral as to the nature of options for choice. What you choose between, in a transformative choice, may not itself be transformative. It is often assumed that transformative choices must be 'big' choices, choices between options that, if pursued, will transform your life. But on the choice-based account, transformative choices can be made between options that don't themselves transform you. You can change your normative character in small, mundane ways. A beach vacation and a mountain retreat might be on a par. Neither option will transform you. But now suppose you commit to some feature of the beach vacation, thereby making it true that you have most reason to go on the beach vacation. You change 'who you are' in this small way by creating for yourself a new reason you didn't have before. You are now, to that small extent, a beach person rather than a mountain person. The same goes for choosing between desserts that are on a par. If you create for yourself a will-based reason to go for the chocolate mousse over the fruit cup, you change your normative character to the extent that before the choice, those desserts were on a par, and after your choice, the chocolate one is better. In short, you always have the opportunity to transform yourself whenever you face options that are on a par. That's because when options are on a par, we have the normative power to create reasons for ourselves that may then change us from people for whom two options are on a par to people for whom one of them is better. We can change who we are in both big and small ways.

The most important and far-reaching upshot of the account proposed here is that transformative choices, so understood, require us to reject standard approaches to rational choice. Transformative choices, then, suggest that rational agency is not simply a matter of *recognizing* and then

responding to reasons given to us by the world. Instead, part of what it is to be rational is to *create* reasons for yourself, that is, put your agency behind a consideration by committing to it. By creating reasons for yourself, you change who you are—you transform yourself—through an activity of your own will. This is transformation in the deepest sense—not something that happens to us but something we do ourselves.

Ruth Chang
E-mail : ruthechang@gmail.com

References:

Anderson, Elizabeth. 1993. *Value in Ethics and Economics*. Cambridge, MA: Harvard University Press.

Barnes, Elizabeth. 2015. "Social Identities and Transformative Experience." *Res Philosophica* 92 (2): 171–187. http://dx.doi.org/10.11612/resphil.2015.92.2.3.

Broome, John. 1991. *Weighing Goods*. Oxford: Blackwell.

Carthcart, Thomas and Daniel Klein. 2007. *Plato and a Platypus Walk into a Bar . . . : Understanding Philosophy Through Jokes*. New York, NY: Abrams.

Chang, Ruth. 2002. "The Possibility of Parity." *Ethics* 112: 659–688. http://dx.doi.org/10.1086/339673.

Chang, Ruth. 2009. "Voluntarist Reasons and the Sources of Normativity." In *Reasons for Action*, edited by David Sobel and Steven Wall, 243–271. New York, NY: Cambridge University Press.

Chang, Ruth. 2012. "Are Hard Choices Cases of Incomparability?" *Philosophical Issues* 22 (1): 106–126. http://dx.doi.org/10.1111/j.1533-6077.2012.00239.x.

Chang, Ruth. 2013a. "Commitments, Reasons, and the Will." *Oxford Studies in Metaethics* 8: 74–113. http://dx.doi.org/10.1093/acprof:oso/9780199678044.003.0004.

Chang, Ruth. 2013b. "Grounding Practical Normativity: Going Hybrid." *Philosophical Studies* 164 (1): 163–187. http://dx.doi.org/10.1007/s11098-013-0092-z.

Chang, Ruth. Forthcoming. "Comparativism: The Grounds of Rational Choice." In *Weighing Reasons*, edited by Errol Lord and Barry MaGuire. Oxford: Oxford University Press.

Elster, Jon. 1979. *Ulysses and the Sirens*. Cambridge: Cambridge University Press.

Harman, Elizabeth. 2015. "Transformative Experience and Reliance on Moral Testimony." *Res Philosophica* 92 (2): 323–339. http://dx.doi.org/10.11612/resphil.2015.92.2.8.

Herman, Barbara. 1996. *The Practical of Moral Judgment*. Cambridge, MA: Harvard University Press.

Hill, Thomas. 2001. *Human Welfare and Moral Worth: Kantian Perspectives*. Oxford: Oxford University Press.

Hsieh, Nien-he. 2005. "Equality, Clumpiness, and Incomparability." *Utilitas* 17: 180–204. http://dx.doi.org/10.1017/S0953820805001512.

Jackson, Frank. 1982. "Epiphenomenal Qualia." *The Philosophical Quarterly* 32: 127–136. http://dx.doi.org/10.2307/2960077.

Kemp, Ryan. 2015. "The Self-Transformation Puzzle: On the Possibility of Radical Self-Transformation." *Res Philosophica* 92 (2): 389–417. http://dx.doi.org/10.11612/resphil.2015.92.2.11.

Acknowledgements This paper was presented at the *Transformative Experience* conference at St. Louis University, September 19–20, 2014. Thanks to Hille Paakkunainen for helpful commentary and Jon Jacobs for his invitation and editorship. Thanks also to various audiences at the University of Texas and Princeton for discussion, especially Johann Frick and Alec Walen, and to two anonymous referees for their comments.

Korsgaard, Christine. 1996. *The Sources of Normativity*. Cambridge: Cambridge University Press.
Korsgaard, Christine. 2009. *Self-Constitution: Agency, Identity, and Integrity*. Oxford: Oxford University Press.
Kurson, Robert. 2007. *Crashing Through: A True Story of Risk, Adventure, and the Man Who Dared to See*. New York, NY: Random House.
Lewis, David. 2004. "What Experience Teaches." In *There's Something About Mary: Essays on Phenomenal Consciousness and Frank Jackson's Knowledge Argument*, edited by Peter Ludlow, Yujin Nagasawa, and Daniel Stoljar, 77–103. Cambridge, MA: MIT Press.
Manne, Kate. Forthcoming. "Democratizing Humeanism." In *Weighing Reasons*, edited by Errol Lord and Barry MaGuire. Oxford: Oxford University Press.
Nagel, Thomas. 1974. "What Is It Like to Be a Bat?" *Philosophical Review* 83 (4): 435–450. http://dx.doi.org/10.2307/2183914.
Parfit, Derek. 1984. *Reasons and Persons*. Oxford: Oxford University Press.
Paul, L. A. 2014. *Transformative Experience*. Oxford: Oxford University Press.
Paul, L. A. 2015. "What You Can't Expect When You're Expecting." *Res Philosophica* 92 (2): 149–170. http://dx.doi.org/10.11612/resphil.2015.92.2.1.
Pettigrew, Richard. 2014. "L. A. Paul on Transformative Experience and Decision Theory I." http://m-phi.blogspot.de/2014/08/l-paul-on-transformative-experience-and.html.
Ross, Jake. 2008. *Acceptance and Practical Reason*. Ph.D. thesis, Rutgers University.
Scanlon, Thomas. 2004. "Reasons: A Puzzling Duality." In *Reason and Value: Themes from the Moral Philosophy of Joseph Raz*, edited by R. Jay Wallace, Phillip Pettit, Samuel Scheffler, and Michael Smith, 231–246. Oxford: Oxford University Press.
Sepielli, Andrew. 2009. "What to Do When You Don't Know What to Do." *Oxford Studies in Metaethics* 4: 5–28.
Ullman-Margalit, Edna. 2007. "Difficult Choices: To Agonize or Not to Agonize?" *Social Research* 71 (1): 51–78.
Ullmann-Margalit, Edna. 2006. "Big Decisions: Opting, Converting, Drifting." *Royal Institute of Philosophy Supplement* 58: 157–172. http://dx.doi.org/10.1017/S1358246106058085.

Neophobia

John Collins

Abstract: L. A. Paul argues that epistemically transformative choice poses a special problem for standard theories of decision: when values of outcomes cannot be known in advance, deliberation cannot even get started. A standard response to this is to represent ignorance of the nature of an experience as uncertainty about its utility. Assign subjective probabilities over the range of possible utilities it may have, and an expected utility for the outcome can be figured despite the agent's ignorance of its nature. But this response to Paul's challenge seems inadequate. Decision theory should leave conceptual room for rational neophobia. A decision theory like Isaac Levi's, which allows for indeterminacy in utility, might accomodate the phenomenon. Levi's discussion of indeterminate utility has focused on examples of risk aversion like the Allais problem and on situations in which there are conflicts of value. Cases of unknowable value arising in transformative choice problems might be handled similarly.

L. A. Paul defines a transformative experience to be one which is both epistemically transformative and personally transformative. An experience is epistemically transformative when there is no way of knowing in advance what the experience will be like, because actually having the experience is the only way of coming to know what it is like. An experience is personally transformative when it is "life-changing in that it changes what it is like to be you, that is, it changes your point of view, and by extension, your personal or subjective preferences" (2015, 16).

Paul argues that such experiences "constitute a class of experiences that raise a special problem for rational decision-making" (2015, 17). And in fact this seems straightforwardly to be the case. Suppose that one is deliberating about whether or not to undergo a transformative experience. Following Paul, let's call such a decision problem a transformative choice problem. Then you are deliberating about what sort of person to become in the future, and in particular you are deliberating about what sort of preferences your future self should have. But some of these possible future preferences might be quite different from your present preferences. They might disagree with your present preferences in various ways. In fact they

Res Philosophica, *Vol. 92, No. 2, April 2015, pp. 283–300*
http://dx.doi.org/10.11612/resphil.2015.92.2.6

might conflict with your present preferences on the very question of whether it is good to be, or to become, a person with such preferences. In such a case, where there is a clash between prior- and envisaged post-choice preferences, it is far from clear which should rationally prevail. That is, it doesn't seem right to say that such conversions can always be justified *ex post facto* from one's transformed point of view. After all, the person one has become might have views that from one's former standpoint seem completely reprehensible. But neither does it seem correct to say that one's prior preferences should always win out either. Mightn't there be genuine cases of enlightenment where your later self thinks quite rightly: I'm a better person now for having undergone that change? And mightn't one add: And I'm better in ways that I simply wouldn't have appreciated beforehand?

For those reasons I think that Paul is absolutely correct in thinking that standard accounts of rational decision-making have a deep difficulty in accounting for choices concerning personally transformative experience.

My interest here, however, is with a parallel problem for decision theory that Paul sees as arising in decision problems involving options that are merely epistemically transformative, like, for example, the decision whether or not to try a new and unfamiliar type of food. This forms one of the major threads running through Chapter 2 of her book, the chapter entitled: "Transformative Choice." I find Paul's argument curiously compelling but also quite elusive. It is my aim in the present paper to explain what I find difficult about Paul's line of thought about epistemically transformative decision problems, and also to attempt to explain why, even when certain distracting side issues are cleared up, there remains a significant core truth here that Paul is sensitive to. I would like to try to display that truth in a way that is free from what to my mind are the distracting side issues.

Paul writes:

> The key to understanding the problem that transformative experience raises is to recognize that the standard models for ignorance can only function if they can represent the structure of the value space of the outcomes for a decision problem. . . .
>
> As a result, in order to use these models for a decision made under conditions of ignorance, *you must be able to know the values of the of the relevant outcomes*. You do not need to know the probabilities that the outcomes, given the acts, will occur, but you do need to know how to value the relevant outcomes. A way to put this is that you must be able to describe the state space of your outcomes, and you must have a suitably defined value function for these outcomes. If you cannot know the values of the relevant outcome or if the values are not yet determined, so that you cannot describe the state space or assign values that

> will remain constant to outcomes, you do not have the information you need to use these types of models to represent your decision. For without an adequate description of the space and without a suitable defined value function for the outcomes, you cannot know if the structure of any particular model adequately represents the structure of the actual situation. (2015, 30–31)

We might summarize the line of argument like this. Deliberation cannot even get started unless the decision maker knows the values of the possible outcomes. When options are epistemically transformative, their values cannot be known in advance. Hence in epistemically transformative choice problems deliberation cannot get started.

There's a picture of decision making behind all this that we might call the simulation model of deliberation. In other passages Paul is quite explicit about this picture:

> When you are considering your options, you evaluate each possible act and its experiential outcomes by imagining or running a mental simulation of what it would be like, should you act, for each relevant possible outcome of each relevant act. You simulate the relevant possible outcomes for yourself, that is, you simulate what it would be like for you to have each of these experiences.
>
> After you run each cognitive simulation, you assign each outcome a subjective value. . . . [O]nce you've determined the overall subjective value of each outcome, you can compare the expected values of different possible acts to determine which one you should perform. (2015, 26–27)

This simulation model of deliberation assumes what Philip Pettit has called the idea of decision theory as a *calculus* for decision making (1991). In order to understand this idea, we shall need to focus a little on the details of the standard theory.

Common to all the standard accounts of decision theory is the idea that rational choice is choice that maximizes expected value. The agent is supposed to have a subjective probability function that assigns credences to all the various possible states of the world, and a subjective utility function that assigns real number values to possible outcomes. This utility function is only unique up to positive affine transformation, in other words both the choice of unit size, and the location of the zero point are arbitrary. (This kind of scale dependence is familiar to us from the case of temperature measurement. Degrees Fahrenheit can be obtained from degrees Celsius by the following affine transformation: multiply by 9/5 and add 32.)

Then the expected value of each of the agent's options can be calculated as a credence-weighted average of the utilities of each of the possible

outcomes of that option. The picture of decision theory as a calculus for decision-making is the natural idea that the process of deliberation mimics this formalism; it is the idea that when a rational decision-maker deliberates, she engages in something like this calculation of expected utilities as subjective probability weighted averages of the utilities of the possible outcomes, where the utilities of the individual outcomes have been arrived at prior to all this by the method of mental simulation.

If this is one's picture of rational deliberation, then it is difficult not to agree with Paul's claim that deliberation cannot even get started unless the decision-maker already knows the values of outcomes.

This picture of rational deliberation goes hand-in-hand with a *psychological realism* about utility and credence. (See, e.g., Buchak 2013, 17.) According to the psychological realist, utility and credence are real mental states. Think of them as degrees of desire and degrees of belief respectively.

For present purposes I'm happy to assume this realist picture. But it will be useful for our purposes to follow Jamie Dreier in drawing a further distinction between two kinds of psychological realism about utility and credence. (See Dreier 1996 and Buchak 2013, 17–18.) Let's focus on the case of utility. Dreier distinguishes between a *constructive* and a *non-constructive realism* about utility. At issue is whether or not facts about an agent's utilities go beyond the facts about the agent's preferences between options. The constructive realist is someone who believes that they do not. According to the constructive realist, all the facts about the agent's utility function supervene on the facts about her preferences. So, for example: the fact that outcome y lies exactly halfway between outcomes x and z on the agent's utility scale is simply the fact that the agent is indifferent between y and a gamble that gives her a fifty percent chance of outcome x and a fifty percent chance of outcome z. For the constructive realist, such facts about preference are constitutive of what it is to have a particular utility function.

A non-constructive realist, on the other hand, thinks that it is possible, in principle, for there to be facts about the utility function that outstrip the facts about what the agent prefers. So according to a non-constructive realist about utility, it might be possible, for example, to access the facts about one's own utility function by direct introspection, or, perhaps, by the method of mental simulation of outcomes that Paul describes.

Now Paul notes that that "for simplicity" she is assuming that "values or utilities assigned to outcomes are psychologically real for the agent (even if, for example, utilities turn out to be partially constituted by their role in preferences)" (2015, 21, fn. 25). But it seems to me that it is only on the assumption of a non-constructive realism about utility that her "deliberation cannot even get started" argument can be made to seem at all plausible.

Suppose that one adopts a constructive realism about utility. Then the whole idea of direct access to one's utilities for outcomes via introspection and mental simulation will seem completely implausible. In fact the whole

idea of decision theory as a calculus for decision making will seem misguided. For the constructive realist, decision theory will be better viewed as what Pettit calls as a *canon* rather than a calculus for good decision making. For the constructive realist, preference is conceptually prior to utility. Any agent whose preferences are coherent in the sense that they satisfy the axioms of formal decision theory can be seen as choosing rationally so as to maximize expected utility, that is, so as to best serve her desires according to her beliefs, where those desires and beliefs, construed as admitting of degree, are quite real, but are nothing over and above that coherent pattern of preference to which she is disposed.

From this viewpoint it seems quite clear what the decision theorist should say about cases of epistemically transformative choice. Since it is impossible to know what an outcome of such a choice will be like in advance of actually having made the choice and experienced the outcome, the method of simulation is unavailable. But so what? In such a situation an "outcome" will in turn be a risky prospect that delivers, with subject probabilities determined by the agent's coherent preferences, various possible utilities if the world turns out to be one way, or another, with respect to how it would turn out to feel like to be the agent experiencing that outcome.

That the precise phenomenological character of each of these "refined" outcomes cannot be anticipated is neither here nor there. Remember: we are working in a decision theoretic framework according to which all that is relevant to the rationality of an agent's choices are the utilities she assigns to outcomes and the credences she gives to possible states of the world. Nothing else is relevant. In particular: further facts about the particular phenomenological character of the outcomes are not relevant. Once one gives up on the non-constructive realist idea that utility is conceptually prior to preference, and thinks instead of the utility function as contructed out of facts about coherent preference, there is nothing at all paradoxical or puzzling about this picture of things: in-principle ignorance as to the precise value of an outcome of an epistemically transformative choice problem simply gets represented, in the usual and obvious way, as a gamble that might yield any one of a range of possible utility values, depending on how things things turn out to be.

So far this all sounds as though I am unsympathetic to Paul's claim that epistemically transformative choice poses a problem for standard decision theory. But that's actually not the case. As I said earlier, I think there's a core of truth to what Paul is claiming. The rest of the paper will be devoted to explaining one way of starting to make good on this claim. It's offered as a friendly amendment to the argument of the second chapter of Paul's book, and she is welcome to accept it or reject it as she sees fit.

Here's the rough idea. Various critics of standard decision theory have argued that decision theory is lacking in that it allows no room for a rational aversion to risk. Similarly—I think—reflection on Paul's epistemically transformative choice examples might lead one to think that the standard

theory is impoverished in another important respect. It's impoverished in that it leaves no conceptual room for what one might call rational *neophobia*. Then in so far as neophobia should not be seen as irrational-in-principle, it will follow that Paul's examples do offer a new and serious challenge to standard accounts of rational choice.

'Neophobia' is a term used in the psychological literature to refer to an abnormal fear of anything new. (Sometimes this is referred to instead as *cainophobia* or *cainotophobia*.) One particularly common form is food neophobia, as many parents of young children well know. Here I will use the term in a neutral way that is not intended to suggest that there is anything abnormal, or pathological, or irrational about this kind of preference structure. Neither do I want to suggest that neophobia is either more or less common than, or more or less reasonable than, the opposing tendency: neophilia (nor, for that matter, to a ceteris paribus indifference toward outcomes that are new and alien in Paul's sense of being epistemically transformative).

Now let's consider what the preferences of a neophobic agent might look like.

In particular, let's consider situations in which a person is confronted with a choice problem in which one of the options has outcomes with which she is experientially unacquainted. For the sake of simplicity I will focus on just the kind of example that Paul introduces: a situation in which an available option is to try some sort of food of a kind that the agent has never previously tasted and in which it might be reasonable to think that the experience of trying it for the first time might be radically unlike any kind of taste experience the agent has ever had in the past. To be definite: let's imagine that the agent, having never previously eaten durian, is now faced with a choice situation in which one of the options is to taste it for the very first time. For the uninitiated: durian is a kind of fruit native to Southeast Asia. Reported opinions about it vary wildly. It has a distinctive smell that some find pleasant, while others find completely disgusting. All agree, however, that the distinctive aroma and taste of the durian fruit are impossible to convey to someone who has never experienced eating it.

Part of the reason for choosing this kind of example is the fact that it seems fairly safe to say, with Paul, that opting for such an outcome will be epistemically transformative for the agent without being personally transformative. Once I've tasted durian for the first time I'll have learned something that I could not possibly have learned in any way other than by actually having had the experience. But at the same time it seems fairly safe to say, in advance, that whatever that experience turns out to be like, it's not going to change in any deep, or important, or fundamental way, the kind of person that I am. It's not, for example, going to result in any change to my core values or preferences, and this is something which, in turn, I can be fairly sure of ahead of arriving at a decision.

So there's the radically unknown, and unknowable option of durian, say, available on the menu. How, according to a standard theory of choice, is the agent supposed to evaluate this option?

The standard proposal, rehearsed earlier, is to represent the agent's ignorance about what the experience of tasting durian will be like as ignorance over a range of possible outcomes in which the experience of tasting and smelling the fruit turns out to be more or less pleasurable (or unpleasant). Now even though the particular felt qualities of the possible experiences in this range cannot be described or anticipated in advance, the idea is that that should not matter, because all of that unattainable information is going to be filtered through the lens of the agent's utility function anyway. Ultimately—so the orthodox story goes—all that is going to end up mattering to the theory of rational choice are the utilities that the agent would assign to each of those possible experiential scenarios were they to turn out to be actual. If that is correct, then the particular, and ungraspable, felt quality of various of those experiences simply falls out of the picture. The adequacy of the standard theory is defended by representing all of that ignorance as simply ignorance as to what the utility of the experience will actually turn out to be.

Now another reason for favoring an illustrative example of this fairly trivial sort is that it also seems fairly safe to say at this point that whatever the experience of tasting the fruit turns out to be like, its utility can be anticipated to fall within a certain range of possible values, so the agent can confidently place upper and lower bounds on how good or bad the experience will turn out to be. So let's assume that we have good evidence that enables us to set aside, for example, such possibilities as that the fruit will turn out to be poisonous, or that it will send the agent into anaphylactic shock, or that it will trigger some other kind of allergic reaction. Similarly, at the other end of the scale, let's suppose that the agent can safely assume in advance that the experience is not going to be *so good* that it will turn out to be "off the charts" in the sense of being better, and of course in an unanticipatible way, than some value set in advance as the maximum possible utility.

Once we have this upper and lower bound to the possible utility of the unknowable experience set, then the idea will be that we can, in principle, go about the task of constructing a kind of *synthetic lottery* over a range of quite familiar outcomes, a synthetic lottery that can then go proxy for the outcome that involves the epistemically transformative experience.

We need not suppose that this lottery have a continuum of possible prizes corresponding to all of the real numbers that are the possible utilities in the interval between the minimum and maximum values. We may suppose that what I'm calling the synthetic lottery has only some finite number of outcomes or prizes. The important thing, however, is that all of those outcomes must involve experiences that are quite familiar to the agent, and that the known utility of each outcome must lie somewhere on the closed

interval between the upper and lower bounds, and that those utilities be sufficiently well distributed, or uniformly spread, over the interval so that whatever the epistemically transformative experience turns out to be like, it will also turn out to have a utility, for the agent, that is very close to the utility of one of the prizes in what I'm calling the synthetic lottery. Again: what one means here by "very close" can simply be adjusted, if required, by increasing the finite number of prizes.

We are now in a position to see what the preferences of what I'm calling a neophobic agent might be like.

Suppose that an agent confronts a decision problem in which some option A is epistemically transformative. Construct a synthetic lottery corresponding to the experientially unknowable option A so that:

(1) For any possible utility value x that the epistemically transformative experience may turn out to have for the agent, there is a possible outcome to the lottery that is both (a) experientially familiar to the agent and (b) has a utility that is (arbitrarily) close to x.
(2) The chances of the various possible outcomes to the lottery are weighted so as to correspond to the agent's subjective probability distribution over the range of possible utilities that the epistemically transformative option A may turn out to have, whatever that subjective probability distribution happens to be.

Then such a synthetic lottery will have, for the agent, an expected utility that is equal to the agent's expected utility for option A.

But now suppose that, despite this equality in expected utilities, the agent nevertheless prefers the prospect of the synthetic lottery to the epistemically transformative option A.

If competing explanations of the pattern have been ruled out, then the remaining preference for the synthetic lottery over the prospect of the epistemically transformative experience may be taken, I think, as an indication that the agent is neophobic. And, of course, the opposite preference pattern, that is, a preference for the radically unfamilar option over the corresponding synthetic lottery constructed so as to have the same expected utility, would be an indication of neophilia.

My feeling is that there need be nothing at all irrational about either of these possibilities. We should have a normative theory of decision liberal enough to allow for cases of rational neophobia. And of course the same goes for the opposed phenomenon of rational neophilia.

It will be helpful here, I think, to compare what I'm calling neophobia with other patterns of preference that orthodox decision theory cannot accommodate and yet which seem perfectly rationally permissible. The first kind of example I have in mind involves an agent who is averse toward risk. Just as many have argued that a theory of decision making should allow for the rationality of various attitudes other than indifference toward risk,

so—it might be argued—such a theory should be just as permissive when it comes to attitudes other than indifference towards what is new.

So let's approach this task by first reviewing the problem that risk poses for standard accounts of decision theory.

> **Example 1: The Allais Problem.** The agent is to win a prize determined by drawing a ticket in a fair lottery with one hundred tickets. Consider the four options A_1 to A_4 displayed in the table below.

	0.01	0.10	0.89
	Ticket 1	Tickets 2-11	Tickets 12-100
A_1	\$1M	\$1M	\$1M
A_2	\$0	\$5M	\$1M
A_3	\$1M	\$1M	\$0
A_4	\$0	\$5M	\$0

Option A_1 guarantees the agent one million dollars no matter which ticket is drawn. Option A_2 is somewhat riskier: it yields five million dollars instead of a million if a ticket numbered 2 through 11 is drawn, but it also leaves the agent with a one percent chance of getting nothing at all. Faced with a choice between these first two options, many agents report a preference for A_1 over A_2. Now this might be taken as evidence that such an agent has a diminishing marginal utility for money: getting the first million dollars makes a lot more difference than getting the next four million dollars would. And, in fact, if the utility difference for the agent between the outcomes Win \$1M and Win \$0 is more than ten times the utility difference between Win \$5M and Win \$1M then a preference for A_1 over A_2 is exactly what expected utility theory prescribes.

The problem is, however, that many of those same agents—apparently perfectly rational people, I'm one of them—also report a preference for A_4 over A_3. That is, they prefer a ten percent chance of five million dollars to an eleven percent chance of one million.

But now agents like us have run foul of standard expected utility theory. For there are simply no utilities that may be assigned to the three outcomes \$0, \$1M, \$5M that can rationalize that pair of preferences as maximizing expected utility. The agent's preferences are in violation of Savage's Sure-Thing Principle, one of the axioms of the standard theory. If you cover over the third column of the table, the pattern of outcomes on what remains is the same for A_1 and A_2 as it is for A_3 and A_4, so the Sure-Thing Principle requires that an agent's preference for comparison for the first pair match that for the second.

What is going on here?

Many decision theorists, going back to Allais himself, have taken this example to be a reductio of any normative theory of choice which rules out as irrational the kind of aversion to risk that characterizes the preference

for A_1 over A_2 and for A_4 over A_3. If these preferences express a perfectly rational attitude toward risk, then standard expected utility theory will have to be liberalized in some way to yield a more reasonable set of norms.

But how might the standard theory be adjusted to accommodate the possibility of rational risk aversion? I will describe two possible answers to that question in what follows. The first of these is a particularly well-worked-out and elegant proposal due to Lara Buchak, developed and defended in her recent book *Risk and Rationality*. I'll approach Buchak's account via an example of the sort she uses to motivate the project in the first chapter of the book. (The version of the example I present here is due to Rachael Briggs.)

> **Example 2: The Pizza Problem.** Confronted with a choice between the following two options:
> (A) One pizza for sure.
> (B) A gamble that yields two pizzas if the toss of a fair coin lands heads and nothing if the coin lands tails.
> My friend and I share a preference for (A) over (B).

But now let's stipulate that the explanation of my preference for (A) over (B) differs from that of my friend's preference for (A) over (B). In particular, let's suppose that I prefer the certainty of one pizza to a toss-up between two pizzas and nothing, because one pizza is just about all that I can eat. I'm full after a single pizza, and as a result, the value I assign to getting a single pizza lies more than half way along the interval on my utility scale from no pizza to two pizzas. As a result of the fact that I have this kind of diminishing marginal utility for pizza, I prefer (A) to (B).

Things are quite different, on the other hand, in my friends's case. My friend, let's suppose, is insatiable. For him, the utility of the second pizza is undiminished by the fact that he has already eaten the first. So for my friend:

$$U(\text{two pizzas}) - U(\text{one pizza}) = U(\text{one pizza}) - U(\text{no pizza})$$

Yet my friend, like me, prefers (A) to (B). Why? Because he is risk averse. He simply does not want to take the chance of getting nothing.

There's another possibility here too, which I will only mention and then set aside. An agent with an insatiable appetite for pizza might prefer (A) to (B) out of *pessimism* rather than risk aversion. That is, the agent might judge that the probability of the fair coin landing heads is less than one half when his dinner depends on the outcome of the toss.

But let's set that further possibility aside. Let's suppose that we are satisfied that my friend assigns subjective probability $\frac{1}{2}$ to the coin's landing heads, whether or not his dinner depends on the outcome, and let's suppose further that we are satisfied that, for him, the utility of one pizza is exactly half way between the utilities he asssigns to two pizzas and that he assigns to nothing. Then by the lights of standard decision theory, my friend's

preference for (A) over (B) is irrational, since, for him, the expected utilities of (A) and (B) are equal to one another.

Yet—to many of us at least—this seems to be the wrong thing to say about my friend's preferences. To many of us it seems as though it is perfectly rationally permissible to be averse to risk taking in this kind of way. From this viewpoint, standard expected utility theory seems unduly harsh or over-restrictive for deeming such patterns of preference irrational.

But perhaps one ought to be suspicious of what I have stipulated above in setting out the details of this second example. By stipulating that we are satisfied, somehow, in a way that is independent of his preference for (A) over (B), that my friend's subjective probability for the coin landing heads is $\frac{1}{2}$ and that his utility gain from the second pizza is equal to the utility gain from the first, we might be thought to be committing ourselves to a non-constructive realism about utility and begging the question against the constructive realist.

Now, certainly, if there *are* further features of my friend's psychological state that we can point to and identify as those psychological features that ground the facts about his utility function stipulated in the second example, then that would demonstrate the inadequacy of any theory that left no room for that possibility. But suppose that there are no such further features to be found. Then a defender of the standard theory might simply reply that the apparent distinction stipulated in Example 2 between my friend's situation and mine is really a distinction without a difference. That is, the claim a defender of the standard theory might make is that this apparent distinction between my friend's situation and mine is precisely the consequence of that incorrect, non-constructive, conception of utility.

This still seems wrong to me. However if the defender of the standard theory adopts this strategy the mistake now seems to be not that a certain rationally permissible set of preferences is being ruled out incorrectly as irrational, but rather that the standard theory is leaving no room at all for a preference structure that in fact is perfectly possible.

Buchak develops and defends a theory of *risk-weighted expected utility* in which the choice-worthiness of an act is determined by three factors, not two. In this risk-weighted theory, the traditional roles of subjective probability and utility are augmented by a third factor, namely a *risk function*

$$r : [0, 1] \longrightarrow [0, 1]$$

that is non-decreasing and such that $r(0) = 0$ and $r(1) = 1$. The function r is intended to capture the facts about an agent's attitude to risk, and, crucially, does so in a way that can be elicited from a pattern of preferences that is coherent in an appropriate technical sense quite independently of the elicitation of probability and utility.

In order to see how this tripartite risk-sensitive scheme works it will help first of all to reformulate the standard account of expected utility in a kind

of stepwise fashion that proceeds from an initial monotonoic rank-ordering of outcomes from worst to best.

The most general case need not detain us here. The basic idea can be grasped by looking at a simple case where there are two possible states of the world s and t and two possibles outcomes x and y ordered so that the latter is at least as good as the former.

In that case the standard expression for the expected utility of an option $f = \{s,x\ ;t,y\}$, that is, of the act that delivers outcome x in state s and outcome y in state t is:

$$\mathrm{SEU}(f) = p(s).U(x) + p(t).U(y)$$

which, since x, y have been listed in order of increasing goodness, can be re-written in stepwise fashion as:

$$\mathrm{SEU}(f) = U(x) + p(t)(U(y) - U(x))$$

Now that we have this equivalent step-wise reformulation of standard expected utility, we can adjust it, via the risk function as follows to obtain Buchak's risk-weighted expected utility REU.

$$\mathrm{REU}(f) = U(x) + r(p(t)).(U(y) - U(x))$$

To get a sense of how this works, let's see how it might be applied to make sense of the distinction between my attitude and my friends's attitude toward pizza in Example 2 above.

Here the two relevant states of the world are H and T, the two possible results of the toss of the fair coin, and the outcomes, ranked for both of us in order from worst to best are no pizza, one pizza, two pizzas.

Then the previously mentioned distinction between my friend's risk aversion and insatiable desire for pizza, and my own risk neutrality and diminishing marginal utility for pizza can be captured by, for example, the assumption that my utility function for pizza is U_1 where

$$U_1(n) = \sqrt{(2n)}/2$$

where n is the number of pizzas received, while my friend's utility function is

$$U_2(n) = n$$

And, furthermore, my risk function r_1 is the identity function

$$r_1(x) = x$$

while my friend's risk function is

$$r_2(x) = x^2$$

Note that for both of us:

$$p(H) = p(T) = 1/2$$

since he and I agree that the coin is a fair one.

Plugging these utility and risk functions into the expression for risk-weighted expected utility we see that for me the value of the gamble g that delivers nothing on heads and two pizzas on tails is:

$$\mathrm{REU}(g) = 0 + r_1(p(T)).(U_1(\text{two pizzas}) - U_1(\text{no pizza}))$$

in other words:

$$\mathrm{REU}(g) = 0 + 1/2.(1-0) = 1/2$$

and this is less than the utility I assign to receiving a single pizza, that is,

$$U_1(1) = \sqrt{2}/2 \approx 0.707.$$

For my friend, on the other hand:

$$\mathrm{REU}(g) = 0 + r_2(p(T)).(U_2(\text{two pizzas}) - U_2(\text{no pizza}))$$

and so for him:

$$\mathrm{REU}(g) = 0 + 1/4.(2-0) = 1/2$$

which is less than the utility he assigns to getting a single pizza, that is, $U_2(1) = 1$.

This indeed has the required result that both of us prefer one pizza for sure to the gamble that gives us a 50% chance of two and a 50% chance of nothing. But that pattern of preferences has a quite different explanation in his case, where it is due to risk aversion, and in my case, where it stems from my diminishing marginal utility for pizza.

An appropiately chosen risk function can similarly rationalize the characteristic pattern of preferences in the Allais problem.

However, there is another kind of example that raises a similar challenge to standard decision theory, and which also cannot be accommodated in Buchak's system. The problem is due to Daniel Ellsberg and it turns out, I think, to be even more helpful to us than the first two examples in seeing how the possibility of rational neophobia might be treated formally (1961).

> **Example 3: The Ellsberg Problem.** An urn contains balls of three colors: red, black, and yellow. You know that it contains exactly thirty red balls and that there are an additional sixty balls which are either black or yellow, but in a ratio that is not known to you. You are asked to compare first the pair of options E_1 and E_2 the outcomes of which are determined by the color of a ball drawn at random from the urn, as specified in the table below.

	Red	Black	Yellow
E_1	\$100	\$0	\$0
E_2	\$0	\$100	\$0
E_3	\$100	\$0	\$100
E_4	\$0	\$100	\$100

Then you are asked to compare option E_3 to option E_4. As was the case in the Allais example above, many apparently perfectly rational agents express a preference for E_1 over E_2, and for E_4 over E_3, despite the fact that there is no standard expected utility representation of that pair of preferences. The situation is strikingly similar to the Allais case in that once again we have a violation of Savage's Sure-Thing Principle: cover over the third column of outcomes on "Yellow," and the pattern of outcomes on what remains is the same for E_1 and E_2 as it is for E_3 and E_4.

But there the similarities end. Strikingly, the risk-weighted utility theory of Buchak cannot accommodate the rationality of the Ellsberg preferences, although as we have seen her account can deal perfectly well with the Allais phenomenon. This difference arises because Buchak drops the Sure-Thing Principle in her axiomatization of preference; part of its work gets down by an axiom she calls Strong Comparative Probability, and it is the Strong Comparative Probability axiom that separates the Allais and the Ellsberg problems. The Allais preferences satisfy it; the Ellsberg preferences do not. (For details see Buchak 2013, 98–100, and Machina and Schmeidler 1992, 762–763.)

The moral of all this seems to be that the pattern of preferences commonly elicited by the Ellsberg example should be seen as an expression not of an aversion to risk, but rather of an aversion to what Ellsberg called *ambiguity*. It seems as though what leads to the choice of E_1 over E_2, and the choice of E_4 over E_3, is a preference for gambling on options where the outcomes have known objective probabilities, rather than options where the situation is "ambiguous" in the sense that the agent does not know what the objective probabilities are.

Now Isaac Levi is a prominent example of a decision theorist who has argued that the Ellsberg preferences should be regarded as perfectly rationally permissible, and that the way to accommodate them in a formal theory of decision is to allow that an agent's subjective probabilities, that is, her degrees of belief, may be *indeterminate* (1986).

In Levi's account, an indeterminate belief state is represented not by a single sharp subjective probability function, but by a convex set P of probability functions. (To say that the set is "convex" is to say that whenever p and q are probability functions in P, then every mixture $\alpha.p + (1-\alpha).q$, where $0 < \alpha < 1$, is also a probability function in P.)

There are various different ways in which such indeterminate probabilities might figure in a formal decision rule. Here we will follow Levi's suggestion that the agent first reduce the set of available options to those that are *E-admissible*.

> **Definition:** If an agent's utility function is u and her indeterminate belief state is represented by the convex set P of probability functions, then an option A is *E-admissible* for the agent if and only if there exists a probability function

> $p \in P$ such that A has maximal expected utility among all her options when those expected utilities are calculated using p and u.

In the Ellsberg example the agent's indeterminate belief state is represented by the set of all probability functions that assign probability 1/3 to Red, probability x to Black where $0 \leq x \leq 2/3$ (and a multiple of 1/60), and probability $2/3 - x$ to Yellow. With these indeterminate degrees of belief, both elements of the option set $\{E_1, E_2\}$ are E-admissible in Levi's sense. If p is chosen from P so that $x = p(\text{Black}) \leq 1/3$ then option E_1 has maximal expected value. For any other choice of p the option E_2 achieves the maximum. So either may be chosen. We can see similarly that both elements of the option set $\{E_3, E_4\}$ are E-admissible.

We could leave it at that, or we could follow Levi in allowing that some second-round rule of choice be applied to further winnow down the options that have survived the first-round test of E-admissibility. For example, if the agent adopts the rule of choosing the option from the E-admissible set that has the highest "security level," that is, the maximin expected utility over all $p \in P$, then the agent will indeed choose E_1 over E_2 and E_4 over E_3. The security levels for the four options E_1-E_4 in that order are $100/3, 0, 100/3$, and $200/3$ respectively (taking the utility of money for the agent to be given by function $u(\$n) = n$.)

Now Levi also maintains that an agent's utilities might also be indeterminate, and this allows him to give a similar account of the rational permissibility of the Allais preferences.

We allow, that is, that an agent's utilities for outcomes be given by a convex set U of determinate utility functions. Since there is already a "choice of scale" indeterminacy in measuring utility—we noted earlier the fact that utilities, like temperatures, will only ever be unique up to a choice of zero point and unit—let's assume that there is a pair of options x, y between which the agent is not determinately indifferent and that are ranked in the same order, y preferred to x say, by every utility function in the agent's set U. Then we may "normalize" the set U by choosing the scale for each of its elements u so that $u(x) = 0$ and $u(y) = 1$

The earlier definition of E-admissibility is then naturally extended to this system that allows indeterminacy in both probability and utility:

> **Definition:** If an agent's indeterminate belief state is represented by the convex set P of probability functions, and her indeterminate value state by a normalized convex set of utility functions, then an option A is *E-admissible* for the agent if and only if there exists some probability function $p \in P$ and some utility function $u \in U$ such that A has maximal expected utility among all her options when those expected utilities are calculated using p and u.

The application of this idea to the Allais problem is quite straightforward. The agent determinately ranks \$0 below \$1M, which is in turn ranked below \$5M. We may choose \$0 and \$1M as the outcomes with respect to which all the utility functions in the set U are normalized, by setting $u(\$0) = 0$ and $u(\$1M) = 1$ for all $u \in U$. Suppose that the agent's value state is then represented by a convex set U of utility functions such that for some $u \in U : u(\$5M) < 1.1$ and for some other $u' \in U : u'(\$5M) > 1.1$, Then for such an agent the characteristic Allais preferences will be rationally permitted, since each of A_1, A_2 will be E-admissible choices from the set $\{A_1, A_2\}$ and each of A_3, A_4 will be E-admissible choices from the set $\{A_3, A_4\}$. And an agent who adopts, for example, the second-round rule of choosing from among the E-admissible options the one whose second-worst outcome is best, will consider the characteristic Allais preferences to be the uniquely rational ones.

I think we should accommodate the possibility of rational neophobia in exactly the same way that Levi treats the Allais problem. That is, I think we should approach it as a phenomenon that can arise when an agent has indeterminate utilities for certain outcomes. Faced with a choice problem involving an epistemically transformative option, an agent can find herself with no determinate attitude toward the goodness of that outcome, with no determinate utility for it. The situation is not one which resolves itself into an uncertainty over which of some set of more fine-grained sub-outcomes is true. It's simply a matter of a basic and irresolvable indeterminacy. That's why the orthodox decision theorist's suggestion that we elicit her utility for the transformative outcome by the method of constructing a synthetic lottery need not always work. It's not possible to elicit a sharp determinate value for the utility of an outcome when it is just a fact that no such unique value exists. The synthetic lottery may yield some unique number, but so what? It's providing an answer to a different question.

If the agent simply has no determinate utility for an outcome X because she is phenomenologically unacquainted with outcomes of that type, then she may recognize that both the outcome X and its synthetic lottery "equivalent" are E-admissible options. And then she might rationally opt for the synthetic lottery over the unknown outcome because she adopts a second-round rule of preferring the familar to the unknown. This is a neophobic preference structure, and it should not be ruled out by a normative theory of choice as irrational. So we should admit indeterminacy in utility, and we should allow for the possibility of rational neophobia.

Indeterminacy of utility can arise in various ways. One variety in which Levi has been particularly interested throughout his career is the kind of indeterminacy that stems from a conflict in values. An agent may recognize that two different and perhaps competing features of an outcome are relevant to establishing its utility. The agent may know that the utility of the outcome is to be figured as a tradeoff between these competing criteria—as some weighted mixture of the simple determinate utilities that would be

arrived at if only one or the other of the two factors were relevant. And yet the agent may be forced to admit that there is no fact of matter as to how the weighting of that mixture should get done. In such a situation the agent will assign no determinate utility to the outcome. The best she may be able to do is to assign it some interval of real-number values parametrized by the possible values of the weighting factor.

In some of the most fascinating, and elusive, passages of Chapter 2 of her book, L. A. Paul seems to be pushing just this kind of point. I have in mind those passages in which, for example, she stresses the richness and multi-dimensionality of the notion of value. To take that kind of criticism seriously might seem to be to reject the standard decision theoretic framework in a rather drastic and fundamental way. It might seem to require rejecting the very idea that rationality of choice could depend simply on facts about expected utility. I've previously resisted that idea strongly and argued it at length with L. A. Paul. But it now seems to me that the required revision to the standard theory need not be so drastic, and that the means for handling her cases of epistemically transformative choice are already well known from the work of Isaac Levi and others and might already be required to handle other well-known problems. That's how I now read those fascinating and elusive passages of the second chapter of Paul's book. I've come to see her discussion of epistemically transformative choice problems as identifying a new and very important role for the theory of indeterminate utility. It's one more reason to be grateful to Paul for having written such a rich and interesting book.

John Collins
E-mail : john.collins@columbia.edu

References:

Buchak, Lara. 2013. *Risk and Rationality*. Oxford: Oxford University Press.

Dreier, James. 1996. "Rational Preference: Decision Theory as a Theory of Practical Rationality." *Theory and Decision* 40: 249–276. http://dx.doi.org/10.1007/BF00134210.

Ellsberg, Daniel. 1961. "Risk, Ambiguity, and the Savage Axioms." *Quarterly Journal of Economics* 75 (4): 643–669. http://dx.doi.org/10.2307/1884324.

Levi, Isaac. 1986. "The Paradoxes of Allais and Ellberg." *Economics and Philosophy* 2: 23–53. http://dx.doi.org/10.1017/S026626710000078X.

Machina, Mark and David Schmeidler. 1992. "A More Robust Definition of Subjective Probability." *Econometrica* 60 (4): 745–780. http://dx.doi.org/10.2307/2951565.

Paul, L. A. 2015. "What You Can't Expect When You're Expecting." *Res Philosophica* 92 (2): 149–170. http://dx.doi.org/10.11612/resphil.2015.92.2.1.

Acknowledgements An earlier version of this paper was presented to the *Res Philosophica* Conference on Transformative Experience held at Saint Louis University on September 19-20, 2014. Thanks to Jon Jacobs and the other organizers of that conference, and to all the participants. I owe a special debt both to Laurie Paul for many helpful discussions of these matters, and to Sophie Horowitz, whose acute comments on the presented version of this paper led me to revise it extensively. Thanks also to two anonymous referees for helpful suggestions.

Pettit, Philip. 1991. "Decision Theory and Folk Psychology." In *Foundations of Decision Theory: Issues and Advances*, edited by Michael Bacarach and Susan Hurley, 147–175. Oxford: Blackwell.

Expecting the Unexpected

Tom Dougherty, Sophie Horowitz, and Paulina Sliwa*

Abstract: In an influential paper, L. A. Paul argues that one cannot rationally decide whether to have children. In particular, she argues that such a decision is intractable for standard decision theory. Paul's central argument in this paper rests on the claim that becoming a parent is "epistemically transformative"—prior to becoming a parent, it is impossible to know what being a parent is like. Paul argues that because parenting is epistemically transformative, one cannot estimate the values of the various outcomes of a decision whether to become a parent. In response, we argue that it is possible to estimate the value of epistemically transformative experiences. Therefore, there is no special difficulty involved in deciding whether to undergo epistemically transformative experiences. Insofar as major life decisions do pose a challenge to decision theory, we suggest that this is because they often involve separate, familiar problems.

1 Introduction: Is Becoming a Parent Rational?

Like most major life decisions, the decision whether to have children is fraught with uncertainty. How would your child turn out? What would your relationship with her be like? How much would you enjoy parenting? As such, it would seem a paradigm of a decision amenable to philosophers' favorite tool for making decisions under uncertainty—decision theory. Very roughly, according to decision theory, you should gauge how good or bad the possible outcomes of each option are, and weight these "utilities" by how likely you think these outcomes are, in order to calculate how much "expected utility" would result from each option. The decision-theoretic recommendation is that you choose the option with the highest expected utility. Since the purpose of this formal tool is to provide a lamp to guide us through the fogs of the future, we might have hoped that it would help with the decision about whether to have kids.

These would be false hopes, according to L. A. Paul, in an article (2015) that has perhaps had more impact outside of academia than any other

* Authors are listed alphabetically by surname.

Res Philosophica, *Vol. 92, No. 2, April 2015, pp. 301–321*
http://dx.doi.org/10.11612/resphil.2015.92.2.5

philosophy essay in recent years.[1] Paul argues that parenting decisions are intractable for standard decision theory. This would mean that insofar as we take decision theory to determine what it is rational to choose, we must conclude that it is neither rational to choose to become a parent nor rational to choose not to become a parent. The limits of reason have been reached, and any parenting decision would be a leap of faith.

These are bold and exciting claims. So what could justify them? In this essay, we will focus on a novel argument of Paul's that is based on the claim that becoming a parent is "epistemically transformative": it gives one new knowledge of what it is like to be a parent and to have experiences related to parenting. These epistemic discoveries are made only upon entering parenthood—too late to inform one's decision to become a parent. On these grounds, Paul argues that a childless person cannot determine how desirable parenting outcomes would be. But without rationally determining the utility of each parenting outcome, this person cannot rationally calculate the expected utility of having a child. Hence, decision theory's silence.

Although Paul's primary focus is parenting, this argument is powerful enough to apply more generally to all decisions that determine whether one undergoes an epistemically transformative experience. It would show that we cannot rationally decide to undergo any new experience, from tasting Szechuan peppercorns to experiencing one's first kiss. Could it really be that decision theory comes unstuck with all of these decisions?

Our answer will be that decision theory is posed no special problems by epistemically transformative experiences. To see this, we will draw a distinction between knowing *what it is like* to have an experience and rationally estimating *how valuable* that experience is. We agree that one cannot know in advance what it is like to have an epistemically transformative experience. But we disagree that someone cannot rationally estimate how valuable such an experience is. This is because direct experience is only one epistemic route to the value of experiences. Two of the other routes are testimony, and observing others' behavior. Moreover, in many cases, we have experiences that are in some respects similar to epistemically transformative experiences. These resemblances can yield us partial knowledge of what an epistemically transformative experience is like. This partial knowledge is often enough for us to be able to rationally assign credences about how desirable we would find the experience—our third method of estimating its value. In this way, we will argue that a more nuanced account of the epistemology of value can provide a firmer foundation for decision theory as a theory of practical reason.

Before proceeding, let us clarify what our target is in this article. Since our interest is in Paul's argument concerning epistemically transformative experiences, our primary focus will be on her first work on epistemic transformation, which is published in this special volume of *Res Philosophica*

[1] Media discussions include Burkeman 2013; Gopnik 2013; Lombrozo 2013a,b; Marshall 2013; Rothman 2013.

but has been available online in its finalized typeset form since at least 2013. This argument is novel, but, we argue, fallacious. Having addressed this argument, we turn to subsequent work of Paul's, which develops a more nuanced overall argument (2014). This argument is more successful in posing a challenge to decision theory, but only because it appeals to familiar problems for decision theory. The epistemically transformative nature of these experiences does no special work.

2 The Challenge of Epistemic Transformation

Since Paul's argument focuses on discovering what it is like to have a new experience, we will frame our discussion around one of philosophy's most famous characters, who also appears in Paul's essay: Frank Jackson's color scientist Mary. As you will recall, Mary "is confined to a black-and-white room, [and] is educated through black-and-white books and through lectures relayed on black-and-white television" (Jackson 1986, 291). We will assume that Mary has survived this social isolation psychologically unscathed by being provided with an ample supply of literature. This has nourished her imagination and allowed her to build up hopes of a future life outside her prison. One of the things she wonders about is whether to become a parent.

From reading glossy black-and-white magazines, Mary has discovered what Paul describes as our culture's "ordinary" way of making a decision whether to have a child: in order to decide whether or not to have a child, someone should consider what the experience of being a parent would be like and consequently "carefully weigh the value of . . . [these] future experiences" (2015, 2). Paul later characterizes the values of these future experiences as "centering on . . . the subjective value of *what it is like* to be the person who made the choice" (4, emphasis added). To decide whether or not to have a child is thus a choice between different "phenomenal outcomes that involve what it's like for her to have her own child" (4). This way of making parenting decisions is Paul's first target.

From reading less glossy black-and-white scholarly tomes, Mary has also discovered a way to formalize our ordinary decision-making: decision theory. This is Paul's main target, which will be of particular interest to philosophers, and hence the one that we will primarily focus on. She describes it as follows:

> To make a choice rationally, we first determine the possible outcomes of each act we might perform. After we have the space of possible outcomes, we determine the value (or utility) of each outcome, and determine the probability of each outcome's occurring given the performance of the act. We then calculate the expected value of each outcome by multiplying the value of the outcome by its probability, and

> choose to perform the act with the outcome or outcomes with the highest overall expected value. (2015, 3)

When this approach is applied to the decision whether to become a parent, one must assign utilities and probabilities to each possible outcome that would result both from becoming a parent and from remaining childless.[2]

Mary cannot wait to apply these approaches to her parenting decision! But then she picks up the Wall Street Journal, and reads an article (2013) reporting the bad news from Paul. According to Paul, these approaches are useless to Mary because they direct her to focus her decision-making on phenomenal outcomes, and yet Mary is phenomenally impoverished. Paul illustrates this point in terms of the epistemically transformative experience of seeing red for the first time:

> For our purposes, Mary's impoverished epistemic situation means, first, that since Mary doesn't know how it'll phenomenally feel to see red before she sees it, she also doesn't know what emotions, beliefs, desires, and dispositions will be caused by what it's like for her to see red. Maybe she'll feel joy and elation. Or maybe she'll feel fear and despair. And so on. Second, because she doesn't know what emotions, beliefs, desires, and dispositions will be caused by her experience of seeing red, she doesn't know what it'll be like to have the set of emotions, beliefs, desires, and dispositions that are caused by her experience of seeing red, simply because she has no guide to which set she'll actually have. And third: she doesn't know what it'll be like to have any of the phenomenal-redness-involving emotions, beliefs, desires, and dispositions that will be caused by her experience of seeing red. Even if she could somehow know that she'll feel joy upon seeing red, she doesn't know what it will be like to feel-joy-while-seeing-redness until she has the experience of seeing red. And these are all ways of saying that, before she leaves her cell, she cannot know the value of what it'll be like for her to see red. (2015, 7)[3]

Similarly, since Mary does not know what it is like to be a parent, Paul would argue, she cannot rationally place a value on becoming a parent. But if she cannot rationally assign utilities to parenting outcomes, then decision theory cannot guide her choice.

Although Mary is alone in her room, she is not alone in facing Paul's problem. The same considerations apply to anyone who is deciding whether

[2] Moreover, standard decision theory assumes that an agent's preferences are complete: for any two outcomes, she is indifferent between these outcomes, or strictly prefers one to the other. Further, it assumes that an agent's preferences do not form a cycle, e.g., it is not the case that an agent has intransitive preferences by, e.g., preferring *A* to *B*, *B* to *C*, and preferring *C* to *A*.

[3] See also Paul 2015, 10, 11, 12, 13, and 15.

to undergo an epistemically transformative experience. So in its generalized form, we can summarize Paul's argument as follows:

(1) There is a certain class of life decisions, including parenting decisions, in which an agent is deciding whether to perform an action that has some chance of resulting in an outcome in which she has a phenomenal experience that would be epistemically transformative for her.
(2) If a phenomenal experience would be epistemically transformative for an agent, then she does not antecedently know what the experience would be like.
(3) If an agent does not know what it is like to have an experience, and this experience is constitutive of a "phenomenal outcome," then she cannot rationally judge the subjective value of this outcome for her.[4]
(4) If an agent cannot rationally judge the subjective value of a phenomenal outcome for her, then she cannot rationally choose between options when one of these options would lead to this phenomenal outcome.
(5) Therefore, there is a certain class of life decisions, including parenting decisions, in which an agent cannot rationally decide what to do.

This formulation is broad enough to apply to both of Paul's targets. To target the argument specifically at decision theory, we could specify that in Premises 3 and 4 the talk of judging a value's outcome should be understood in terms of talk of assigning utilities to this outcome.

Is this a sound argument? Premise 2 is true by the definition of "epistemically transformative,"[5] Premise 4 is highly plausible, and it is trivial to see that the argument is valid. Thus, there are two premises that deserve further investigation—Premises 1 and 3. We will proceed to discuss each, organizing our discussion in terms of increasing importance. We will start with preliminary remarks concerning Premise 1. We will then offer our central criticism of the argument, arguing that we should reject Premise 3. After that, we will offer a diagnosis of why the conclusion might have seemed plausible, by noting familiar problems that arise for decision theory in these contexts.

3 The Broad Scope of Premise 1: The Pervasiveness of Epistemic Transformation

We will begin by noting how Paul's argument applies to her main target: decision theory. When applying decision theory to a decision, an agent

[4] Thanks to Paul for guidance on how to formulate this premise.

[5] Though as we note later, we think it is illuminating to draw a distinction between having complete knowledge and partial knowledge of what it is like to have an experience. In light of what we go on to say, Premise 2 is only true when it is read as concerning complete knowledge.

needs to consider whether she has any credence that an option results in outcomes that involve her having certain phenomenal experiences. If she has some credence that these outcomes will obtain, then she will need to consider how much utility to assign to these outcomes. But if she cannot rationally assign utilities, then she cannot rationally provide herself with the inputs necessary for the decision-theoretic cogs to start grinding.

This feature of Paul's argument means that it has much wider scope than it might at first seem. The argument does not simply concern decisions to become a parent. It also concerns any decision that an agent thinks might lead at some point to her becoming a parent. Suppose Mary has escaped from her colorless prison and gets asked on a date for the first time. If she has some credence, however small, that accepting the invitation will one day lead her into parenthood, then decision theory requires her to assign a utility to the outcomes in which she becomes a parent. This point becomes even more pressing when we consider how many epistemically transformative experience there are: seeing red, tasting a durian fruit, flying in an airplane, falling in love, falling out of love, suffering the ennui of a mid-life crisis, grieving over a loved one's death, climbing a mountain, riding a roller-coaster, fighting in combat, and so on. These are all experiences that are foreign to Mary. Insofar as Mary has some credence that leaving her monochrome prison may result in her undergoing one such experience, Paul's argument would imply that she cannot rely on decision theory to rationally decide whether or not to escape. And this point does not concern just poor Mary. For almost any practical decision we make, we should have some credence that one of our options will bring about an outcome in which we have at least one epistemically transformative experience. Thus, if sound, Paul's argument would show that we cannot appropriately assign a utility to this outcome, and that hence decision theory is stymied. So the argument does not just threaten decision theory's application to parenting decisions. It threatens its application to almost any decision at all.

4 Rejecting Premise 3: The Epistemology of the Value of Experiences

Looking more closely at Premise 1 showed that Paul's argument has considerably more breadth than one might first have thought. Before accepting such a revisionary conclusion, we should examine the argument's crucial step: Premise 3.

(3) If an agent does not know what it is like to have an experience, and this experience is constitutive of a "phenomenal outcome," then she cannot rationally judge the subjective value of this outcome for her.

In this premise we move from descriptive uncertainty about what a phenomenal outcome is like to evaluative uncertainty about the value of that

phenomenal outcome. In support, Paul argues that "the relevant values are determined by what it is like for you to have your child." Consequently, when deciding whether to have children, "the value of your act . . . depends largely on the phenomenal character of the mental states that result from it" (2015, 5).

But while there is plausibility to the claim that the phenomenal character of an experience is typically relevant to the value of the experience, we should still distinguish the experience's phenomenal character from the experience's value. This is because agents might differ in their attitudes toward the same phenomenal experience. For example, Mary may prefer the taste of sugar to the taste of salt, while her prison guard has the opposite preference. So, the experience of tasting sugar may be more valuable to Mary than her guard. Drawing this distinction allows us to also draw an *epistemological* distinction between awareness of an experience's phenomenal character and awareness of its value.

Once drawn, this epistemological distinction should make us suspicious of Premise 3. From the fact that an experience is epistemically transformative, it only follows that the agent is not antecedently in a position to know what the experience would be like. This is consistent with the agent being able to rationally estimate the experience's value. If you have not given birth, then you do not know what it is like to have the experience of prolonged labor. If you have not experienced a year of isolation in a super-max prison, then you do not know what it is like to be deprived of all human contact for an extended period of time. If so, these experiences would be epistemically transformative. But without having undergone these experiences, you can still judge the intrinsic value of the phenomenal aspect of these experiences.[6] (Hint: they contain intrinsic disvalue.) The same holds for positive experiences. Given the limited dating opportunities in her prison, Mary does not know what it is like to fall in love with someone who reciprocates her feelings. Nevertheless, her literary window on the world could enable Mary to rationally estimate the intrinsic value of this experience.

So how can Mary rationally estimate the value of epistemically transformative experiences? What kind of evidence could she have? In fact, there is not one single source of relevant evidence. There are at least three. We will illustrate each with examples of epistemically transformative experiences.

4.1 The Method of Receiving Testimony: The Mystery Closet

The first source of evidence concerning the value of an epistemically transformative experience is testimony:

[6] This is consistent with thinking that the experience has *extrinsic* value, e.g., because the labor instrumentally leads to the birth of one's valued child. We focus on the intrinsic value of epistemically transformative experiences, given that it is this type of value that Paul claims one cannot know. Our conception of intrinsic value follows that of Rae Langton. Langton holds that something's intrinsic value is the value that something has "in itself" which we take to be equivalent to the value it has in virtue of its intrinsic properties (Langton 2007).

> **Mystery Closet.** From a flyer, Mary learns that the funfair is in town outside her prison walls. She reads that one of its attractions is the "Mystery Closet," in which customers undergo an experience. The experience is incredibly rare, and so almost certainly customers will not have had the experience before. Out of thousands of customers, every single one has said that they greatly valued the experience. Even better, the organizers have made it free to enter the closet, hoping that it will draw people into attending the funfair.

Easy question: can inexperienced Mary rationally estimate whether the Mystery Closet experience would have intrinsic value for her? The answer is obvious: yes she can. The evaluative testimony of the other customers has given her excellent evidence that the epistemically transformative experience would be a valuable one. Another easy question: if Mary could enter the Mystery Closet, should she choose to do so? The answer is yes, again: it would be rational for her to choose to enter it, given her evidence about the value of the experience.[7]

As well as stylized examples, there are real world examples of uniformly positive testimony. The most obvious examples involve extreme pleasure or pain. We can put our hands on our hearts and say that we do not know what it is like to be high on heroin or crack cocaine. And yet we are still able to rationally assign credences about whether we would find intrinsic value in these experiences. Similarly, we are fortunate enough not to know what extreme torture is like. All the same, we are able to rationally estimate whether we would disvalue this experience. One reason why we are able to do so is that other people have had these experiences, and have testified as to whether these are valuable or not. Our estimates of these experiences' future value can then rationally guide our actions. If we were given a choice as to whether to undergo torture for a couple of dollars of reward or forgo both torture and reward, it would be rational for us to choose the latter.

These are examples of uniform testimony. But more commonly, testimony will be mixed. Consider:

> **Durian (Simplified).** Mary reads that 50% of people who eat durian say they quite like the taste, but the other 50% say that they find it slightly nauseating.

In this case, Mary's credence as to whether she would enjoy the experience of tasting a durian should be split: she should assign 0.5 credence to the possibility that she would find value in eating durian, and 0.5 credence to the possibility that she would not. Assuming that the intrinsic value of the gustatory experience for someone depends primarily on this person's

[7] See Harman 2015 (especially section 1.2) for a similar argument that we can reasonably rely on testimony to learn the value of transformative experiences.

enjoyment of this experience, this gives Mary split evidence about whether the experience of tasting durian would be valuable for her. Again, this evaluative evidence can guide her actions. It is plausible to think that if Mary is risk averse, then she rationally ought not eat durian, whereas if she is risk loving, then it is rational for her to eat it.

Paul is more pessimistic about the possibility of learning from testimony in these cases, but this pessimism is based on considering only the idea that testimony cannot tell us *what an epistemically transformative experience is like*.[8] This idea is undoubtedly correct; the hallmark of epistemically transformative experiences is that we cannot fully know what they will be like, by testimony or any other means. But all the same, testimony can tell us how valuable an experience is. Paul does indirectly tackle this evaluative testimony when she discusses the evidence provided by survey data about how satisfied parents are (2015, 17–20). Paul's central response is that this evidence might only provide an agent with "external" evidence about whether parenting would maximize utility for her, but that it is irrational to choose to maximize utility instead of consulting her "subjective . . . phenomenal preferences."[9] Paul writes:

> Imagine Sally, who has always thought that having a child would bring her happiness, deciding not to have a child simply because she knows not having one will maximize her utility. For her to choose this way, ignoring her subjective preferences and relying solely on external reasons seems bizarre. . . . Now consider Anne, who has always thought that having a child would bring her misery, deciding to have a child simply because she knows it will maximize her utility. Again, the decision procedure seems bizarre from our ordinary perspective. Choosing rationally requires a very different way of thinking about the decision than we ordinarily think it does—to be rational, we have to ignore our phenomenal preferences. (2015, 19)

Unlike Paul, we do not find Anne's behavior bizarre at all. It seems that, like Sally, she has simply revised her earlier beliefs about how good it would be for her to be a parent, in light of new evidence about other parents'

[8] "Perhaps you think that you can know what it's like to have a child, even though you've never had one, because you can read or listen to the testimony of what it was like for others. You are wrong" (Paul 2015, 12). In personal communication, Paul agrees that in cases like Mystery Closet and Durian (Simplified) an agent can be rational in accepting evaluative testimony. Nevertheless, she argues that epistemically transformative experiences pose a special problem: in many cases involving epistemically transformative experiences people vary widely as to which value they assign to a particular phenomenal outcome. We agree that when there is such variation, relying on evaluative testimony is more problematic. But we are skeptical that the difficulty here has to do with those experiences being epistemically transformative. We discuss this in more detail in section 5.

[9] Paul also raises the worry that this evidence is not enough to go on. We respond to this when we discuss sparse or messy evidence in section 5.

happiness. As Paul notes, Anne has received evidence that parenting would maximize utility for her. Anne's utility of course depends on the satisfaction of her preferences. So Anne has received evidence that her preferences will be satisfied by parenting. If we assume, with Paul, that parenting happiness depends on the satisfaction of phenomenal preferences, then Anne has received evidence that her phenomenal preferences will be satisfied. Moreover, insofar as these phenomenal preferences are Anne's own preferences, she has received evidence that her subjective preferences will be satisfied. Therefore, we conclude that external testimony can provide Anne with evidence about how parenting would satisfy her "subjective, phenomenal preferences." The dichotomy between consulting subjective preferences and relying on external reasons is a false one.

It may be helpful in this respect to recall the example of the Mystery Closet, in which customers have novel experiences. This is a paradigm case where prospective customers should care about whether the experience will satisfy their phenomenal preferences. Moreover, since they are making these decisions self-interestedly, they should consult their subjective preferences. Of course, the testimony of previous customers provides them with excellent evidence that they will be glad they went in the closet. In this way, testimony can provide them with external evidence that their subjective phenomenal preferences would be satisfied. Therefore, even if we should make these decisions on the basis of subjective phenomenal preferences, then this consideration is not a good reason for turning our back on evaluative testimony.

It is of course true that, by using testimony, someone is not using first-personal imaginative projection to learn about the satisfaction of her subjective, phenomenal preferences. But our point is that nonetheless the testimony does allow her to learn about the satisfaction of her subjective, phenomenal preferences. First-personal imaginative projection is not the only epistemic route available.

Should we worry that relying on testimony as evidence about the satisfaction of phenomenal preferences would be an "inauthentic" way of making decisions?[10] It is hard to say in the abstract, without a developed account of what authenticity of choice involves, but we suggest not. It may be plausible that authenticity requires one to aim at the satisfaction of one's own preferences (including subjective phenomenal preferences). But we see no intuitive case for thinking that authenticity constrains how one should acquire evidence about how one's own preferences would be satisfied. After all, it would not be inauthentic for someone to choose to enter the Mystery Closet on the basis of testimony, provided that this testimony had bearing on whether the Mystery Closet would satisfy her own preferences. This seems to us no less true in cases where the stakes are very high, or where the testimonial evidence is messy or inconclusive. As we discuss in more detail later, mixed evidence would make the choice more risky. But as

[10] Paul raises considerations of authenticity (2014).

a general point there seems nothing inauthentic about making gambles, when one does so on the basis of how likely, and to what extent, one's own preferences will be satisfied. We suspect the temptation to think that there is a tension between authenticity and testimony-based deliberation comes from running together two ways in which deliberation might be "first-personal." Deliberation might be first-personal in either of two ways: in the sense that it involves imaginative projection concerning what it is like to have experiences, and in the sense that it aims at the satisfaction of one's own desires. The first type of first-personal deliberation may pose problems for testimonial evidence, but this type of deliberation has no connection to authenticity. There is plausibly a connection between authenticity and the second type of first-personal deliberation, but this is a type of deliberation that we can conduct on the basis of testimony. Distinguishing these two senses of "first-personal" therefore removes the temptation toward thinking there is a conflict between authenticity and testimony.

4.2 The Method of Observation: The Dog on the Beach

Testimony is not our only source of evidence about the value of others' experiences. Often, this value is revealed in their behavior. This is what makes it possible for us to discover whether speechless animals are having valuable experiences:

> **Dog on the Beach.** Sparky bounds up and down the sand. He dives into the sea to retrieve a tennis ball, before returning to the shore where he vigorously shakes himself dry. He meets a new dog, whom he gives a good sniff, and then chases a seagull, with abandon but not success. Throughout, Sparky's eyes are bright, and his ears are perky; he is jumping up and down, his body is wiggling and his tail is wagging.

It does not take a dog-whisperer to realize that Sparky is a happy dog, who is greatly enjoying his experiences on the beach. We know that his behavior indicates that his experiences contain intrinsic value. This is the case even though we do not know what it is like to have these canine experiences—no more than we know what it is like to be a bat (Nagel 1974).

The same is true of our fellow human animals. We can observe people's facial expressions, their body language, and other forms of their bodily behavior. On this basis, we can discover whether their experiences have intrinsic value. Moreover, we can do so even when we ourselves have not had these experiences. Suppose Mary watches footage of a drunk person who is smiling, laughing, and uncharacteristically telling her friends how much she loves them. As a lifelong teetotaler, Mary does not know what this person's inebriated experiences are like, but she can tell that the drunk is having a pleasurable experience. Alternatively, suppose Mary observes

someone suffering from clinical depression, who is eating less, sleeping less, and is removing herself from social engagements. Even if Mary does not know what the experience of severe depression is like, she can still infer that this experience does not contain intrinsic value. By using her knowledge of the value of others' experience, Mary can make inferences about how much value that experience would have for her. In this way, observation provides Mary with evidence with which to rationally estimate the value for her of these epistemically transformative experiences.

4.3 The Method of Inference from Similar Experiences: Vegemite

So far, we have argued that we can rationally estimate the value of an epistemically transformative experience by considering how much value this type of experience has for others. But we often also have specific evidence bearing on what our own personal preferences are likely to be. Experiences fall into broader kinds. If someone has had some experiences that are members of a kind, then she can inductively come to know something about what the other members of this kind are like. Thus, our third source of evidence regarding the value of an epistemically transformative experience is to consider its resemblance to other experiences that we have had.

To illustrate this point, let us consider an example that Paul takes from David Lewis. According to Lewis, you cannot come to know what it is like to taste Vegemite without actually having tasted it:

> If you want to know what some new and different experience is like, you can learn it by going out and really having that experience. You can't learn it by being told about the experience, however thorough your lessons might be. . . . You may have tasted Vegemite, that famous Australian substance; and I never have. So you may know what it's like to taste Vegemite. I don't, and unless I taste Vegemite (what, and spoil a good example!) I never will. (1990, 292)

Quoting this passage, Paul endorses Lewis's claim that tasting Vegemite for the first time is epistemically transformative. Since it is transformative, she argues, we cannot rationally assign a value to tasting Vegemite.

But this overlooks the fact that even if we cannot have complete knowledge of the phenomenal feel of an epistemically transformative experience in advance, we can still have partial knowledge of this. This partial knowledge can be a basis on which to rationally estimate the value of the Vegemite-tasting experience. For example, Mary can read that the experience of tasting Vegemite is an experience of tasting something intensely salty and savory. This testimony is enough for Mary to know that tasting Vegemite has some similarity to the experience of tasting soy sauce, parmesan, or

anchovies.[11] If Mary has been revolted every time that she ate intensely salty and savory foods, then tasting Vegemite is unlikely to be an intrinsically valuable experience for her. More generally, awareness of these resemblances and of one's preferences can provide a guide to whether a new experience would be valuable.

As with testimony, Paul does consider resemblances: "Being around other people's children isn't enough to learn about what it will be like in your own case. The resemblance simply isn't close enough in the relevant respects" (2015, 13). And this is plausible, so far as it goes. Arguably, one cannot fully appreciate what it is like to be a parent by being around other people's children. But even so, we can have partial knowledge of what this is like. In turn, this partial knowledge can provide a rational guide for our estimates concerning the value that parenting would have for us. Suppose a childless kindergarten teacher takes great pleasure in being around children, caring for them, and seeing them develop and flourish, and does not particularly mind the associated unpleasant tasks. This person clearly has some grounds on which to form credences concerning how much he would value the experience of parenthood.[12]

Indeed, if we filled in the details of the Mary case in the right way, we might even imagine that Mary is able to make predictions along these lines about her experience of seeing red. Because this would be an epistemically transformative experience, Paul argues that Mary cannot know whether she would value it (2015, 14). But we can imagine the case in such a way that it is plausible that Mary can justifiably have high credence that she would value it. Suppose that Mary's aesthetic sensibility is heavily biased toward finding sights beautiful; she finds value even in sights that are not conventionally beautiful. Further, Mary burns with a deep yearning to understand all aspects of the human experience—she wants to feel what others feel, as she values the insight this brings her of their lives. Moreover, Mary's curiosity knows no bounds; she is an adventurous sort who loves novelty for its own sake, and is never ruffled by the exotic. Now, consider the fact that seeing red for the first time is a member of the kinds, "visual experience," "experience that has been had by many other humans," and of course, "epistemically transformative experience." In light of this fact, if Mary is aware of her aesthetic sensitivity, her interest in other humans and her yen for the new, then she is in a position to rationally estimate the value of seeing red for the first time.

[11] Since testimony of qualitative resemblances is different from evaluative testimony, the third epistemic method of making inferences from similar experiences is distinct from our first epistemic method of receiving evaluative testimony. Receiving qualitative testimony that Vegemite is intensely salty does not by itself allow one to estimate the value of eating Vegemite. By contrast, receiving evaluative testimony that torture is intensely disvaluable does allow one to estimate torture's value.

[12] Harman (2015, section 1.1) also argues that having similar experiences can give us good evidence about what it is like to be a parent.

4.4 Summary: Why We Should Reject Premise 3

In light of these considerations, we conclude that there are counterexamples to Premise 3:

(3) If an agent does not know what it is like to have an experience, and this experience is constitutive of a "phenomenal outcome," then she cannot rationally judge the subjective value of this outcome for her.

The plausibility of Premise 3 relies on a restricted view about what counts as the admissible evidence concerning the value of experiences: the premise is true only if, as Paul suggests, the only admissible evidence is complete knowledge of the phenomenal character of the experience. Our arguments in this section aimed to show that this restricted view is false. We can use testimony, behavioral observation and inference from similar experiences to rationally estimate the value of new experiences.

5 Familiar Epistemic Problems for Would-be Parents

We have argued that Paul's argument fails: from the fact that an experience is epistemically transformative, it does not follow that one cannot make a rational decision about whether to undergo it. In more recent work, Paul offers a more restricted version of the argument. Paul has narrowed her interest to high-stakes cases.[13] In high-stakes cases, the transformative experiences that purportedly create trouble for decision theory involve not just phenomenal ignorance, but also conflicting and inconclusive testimony about what it is like to undergo them, as well as changes in the agent's core preferences. In this section, we agree that in these more restricted cases it may well be tricky to employ decision-theoretic reasoning to guide one's decision. This, however, can be traced back to some familiar, and more general, challenges for epistemology and decision theory. The fact that these experiences are epistemically transformative is irrelevant. At the same time, we offer an alternative explanation of why Paul's original argument may have seemed compelling. We will start by discussing problems that arise from the kind of evidence that we have available when making life-changing decisions. We will then discuss problems raised by preferences in life-changing decisions.

At several points in her discussion, Paul emphasizes how hard it is to know what one's future experience is like. She characterizes this problem as one of qualitative ignorance:

[13] Paul clarifies this in her comments on this essay at the 2014 Bellingham Summer Philosophy Conference. Similarly, in Paul 2014, 18, she focuses on "decisions about whether to undergo an experience that will change your life in a significant new way." We take this to be a refinement of her earlier argument in Paul 2015, the scope of which more broadly included low-stakes decisions to see red for the first time or to taste Vegemite for the first time.

> **Qualia-Ignorance:** Of one specific experience, not knowing what it is like to have this experience.

After all, this is why epistemically transformative experiences are meant to pose a special problem for decision-making.

But at key points, Paul also appeals to another type of ignorance.[14] Recall her discussion of Mary seeing red, quoted here in full in section 2 (2015, 7).[15] The intuition elicited by this discussion is that Mary's ignorance leaves her unable to assign a value to her experience. But why? We suggest that the main part of the explanation is that Mary is unsure whether her experience would be a frightening experience, a stressful experience, a satisfying experience, and so on.[16] This is simply an instance of a more general type of ignorance:

> **Which-Ignorance:** Of many specific experiences, not knowing which of these experiences one will undergo.

Which-ignorance is independent of qualia-ignorance.[17]

In her more recent work, it also looks as if Paul appeals to which-ignorance as posing difficulties for decision-theoretic reasoning. In discussing the transformative choice of becoming a vampire, Paul argues:

> What if it turns out, given your delicate sensibilities, that once you've transformed, you can't stand chicken blood—all you'll want to drink is human blood, in particular, the blood of male virgins. (One of your vampire friends confides that he is actually quite finicky now that his palate has been educated about platelet terroir.) But contemporary vampire society frowns on drinking human blood, since it isn't good for public relations. And so, if you become a vampire, for the foreseeable future, you'd have to eat food that absolutely disgusts you, and you'd have to constantly confront and overcome your repulsive urge to attack innocent little boys. . . . The problem here is that you can't predict how your preferences will change. Something that seems disgusting now might seem preferable to the finest of wines once you've been vampirically rewired. (2014, 45)

[14] See Paul 2015, 7, 9, and 14.

[15] See also Paul 2015, 11, 12, 13, and 15.

[16] We pass over a more minor point in the quoted passage where Paul notes that Mary "doesn't know what it will be like to feel-joy-while-seeing-redness until she has the experience of seeing red." We find this consideration to have no intuitive appeal: it should be clear to Mary that feeling-joy-while-seeing-redness will have positive value for her.

[17] We can see this by noting two points. First, there can be which-ignorance without qualia-ignorance: when the sky is gray, one can be unsure whether one will undergo the familiar experience of walking home in the rain or another familiar experience of walking home dry. Second, there can be qualia-ignorance without which-ignorance. Suppose that there is a single qualitative experience corresponding to what it is like to be a bat using echolocation to find an insect (Nagel 1974). Since it is a single experience, we do not have which-ignorance about it, but we do have qualia-ignorance about it.

Again, the problem Paul points to is that you cannot know which preferences you will acquire once you turn into a vampire. And so you cannot know which experiences you will be having: one of relishing chicken blood or one of being disgusted by it.

It is important to note that which-ignorance by itself poses no problem at all for decision theory. In fact, it is exactly this kind of ignorance that gives decision theory its purpose. Decision theory is a formal tool for acting when one is unsure about the causal consequences of various options; it guides these choices in light of one's credences in these causal consequences obtaining. Decision theory does not guide our actions by assuming we have knowledge of the actual utility of the outcomes that will in fact obtain as the result of our actions. Instead, it guides us to perform the actions that have the highest expected utility, which is based on how likely we consider various outcomes to obtain. All we need in such situations is the ability to assign rational credences to various outcomes' obtaining, and to assign utilities to those outcomes. And, as we have argued, this is something we can do when making decisions regarding epistemically transformative experiences.

But situations involving which-ignorance may prove tricky for decision theory in other ways. To have rational inputs with which to apply decision theory we need to be in a position to assign rational credences to various outcomes of our action. But our world is often extremely epistemically uncooperative. For one, it is often ungenerous with the evidence that it provides us. Paul brings this out when discussing the possibility of making use of survey data about other parents' happiness in order to inform our decisions about whether to have children. One of her objections is that "[t]here just isn't enough evidence available to support this sort of reasoning"; so, we should "hold off on deciding, due to lack of conclusive evidence" (2015, 19). Similarly, Paul argues that if "we assign values and credences based on insufficient evidence, and calculate the expected value of our acts using such assignments, our decision does not meet the normative standard for rationality" (2014, 23). In addition, the world sometimes provides us with different pieces of evidence that are so messy that it is unclear what the evidence supports. Even if we have plenty of survey data and detailed testimony from many friends who are parents, how should we evaluate this evidence to form our overall credences? As Paul points out, this is particularly a problem when agents vary widely as to which value they assign to the outcome in question (2014, 28).

In light of these challenges, we might say that there is a problem of sparse or messy evidence: either the evidence is too sparse to support any rational assessment at all, or the evidence is too messy to support the type of reasoning required for the precision of decision theory.

How to use sparse or messy evidence to form credences is a challenging problem for epistemology. To illustrate, consider the following case of Adam Elga's:

> **Stranger.** A stranger approaches you on the street and starts pulling out objects from a bag. The first three objects he pulls out are a regular-sized tube of toothpaste, a live jellyfish, and a travel-sized tube of toothpaste. To what degree should you believe that the next object he pulls out will be another tube of toothpaste? (2010, 1)

This case nicely illustrates the difficulty of assigning credences when our evidence is sparse and messy: you have not got much to go on, and it is unclear how to put together the scant pieces of evidence that you have. In these respects, Elga's case is similar to the kinds of evidential situations that we often find ourselves in when considering epistemically transformative experiences like becoming a parent. We have observed our friends and others becoming parents (or not). We have heard or read all kinds of testimony. But how should we weigh all of this evidence together?

Indeed, this problem is particularly likely to arise with the epistemic methods that we discussed earlier. Take testimony. Our previous durian example was artificially simplistic. A more realistic variant would be:

> **Durian (Complex).** Mary reads in the Lonely Planet Guide to Asian Fruits that many people consider durian a delicacy, while a minority find the taste disgusting. Her internet pen pal says that he considers it the "king of fruits." Her prison guards say that it is not such a big deal either way. Knowing this, Mary is deciding whether to eat a durian for the first time on her release.

In this case, it is much harder for Mary to estimate how much value she would get from eating durian. One problem is that it is hard to tell how much value is derived from tasting "a delicacy" or the "king of fruits." But more pressingly, it is hard for her to estimate how likely it is that her experience will be like that of the majority or that of the minority. How many people were consulted by the Lonely Planet before it judged what the majority and minority preferences were? And just how major is the majority: 90%? 70%? 50.01%? While this is in doubt, it is hard for Mary to use this evidence to estimate the value she would get from eating a durian. In this type of case we might think that the evidence is simply too sparse or too messy to license precise reasoning.

One might conclude that in cases of messy or unclear evidence we are not licensed to form any kind of doxastic attitude.[18] Taking this line would mean throwing out much of epistemology as well as decision theory. One might think that this goes too far: after all, we do have some information in situations like Elga's. It is just not clear exactly how it adds up. In light of this observation, some people—though not

[18] Just as one might think that in cases of extreme ignorance, where we have no evidence bearing on a proposition, one should not form any doxastic attitudes at all toward this proposition.

Elga—have taken this type of case to call for a partial revision of standard approaches in formal epistemology and decision theory. They argue that messy cases show that, sometimes, epistemic rationality does not require us to assign *precise* credences. Rather, in some cases we are rationally required to assign "mushy" credences, which are understood either as a range of precise credences, or sometimes as some other kind of coarse-grained doxastic state. But these problems are not unique to decisions about epistemically transformative experiences; they arise across the board.[19] For most of us, the possibility that Elga's stranger has another tube of toothpaste in his bag does not involve any epistemically transformative experiences—just familiar ones that are hard to assess under the particular circumstances. Whether a particular sparse or messy body of evidence concerns epistemically transformative experiences is doing no special work here.

These points bear on parenting decisions. Alas, our epistemically uncooperative world has furnished us with evidence that is less helpful than we should like. There are two key issues in this regard. First, there is a plurality of possible parenting outcomes that might obtain: postpartum depression, the parental pride that floods social media with baby photographs, and so on. Someone can have sparse or messy evidence about whether each outcome would obtain. (This leads to the aforementioned which-ignorance of the outcomes of parenting decisions.) Second, someone can have sparse or messy evidence about the value that she would get from a particular parenting experience. For example, if Mary's only testimony about a particular experience is limited to some rather abstruse poetry, then it will be hard for her to estimate how valuable the experience would be for her. But although these issues surface with parenting decisions, no special work is done by the fact that parenting experiences are epistemically transformative. So in these respects, parenting is simply an interesting new example of a familiar epistemic problem.

In addition to epistemic problems, there are also problems concerning preferences. We will end by noting two of these. The first is that decisions such as whether or not to have children may involve incommensurable preferences. To see this, suppose for simplicity that you have good evidence that whichever choice you make, you would be happy and fulfilled. But you would be happy and fulfilled in very different ways: you are deciding between the freedom to pursue your own projects and the joy of watching your child grow and develop. As such, your preferences may be incommensurable, and there may be no way of assigning precise utilities to each experience in a way that adequately captures your attitudes. Since decision

[19] For criticism of mushy credences, see White 2009. For a defense, see Schoenfield 2012. Also see Sturgeon 2008 for further discussion of when different types of evidential situation might warrant different types of doxastic attitude. See Carr Unpublished for an argument that we can accommodate intuitions supporting mushy credences without abandoning the standard Bayesian framework.

theory requires precise utilities,[20] decisions involving incommensurability present a challenge to standard decision theory—a challenge that is the subject of ongoing debate.[21] But this challenge is orthogonal to the issue of epistemic transformation—we can have incommensurable preferences about things we have already experienced, and we can also have commensurable preferences about epistemically transformative experiences.

The second difficulty is that life-changing decisions often involve a shift in one's preferences or desires regarding the outcomes in question. For example, the experience of becoming a parent may change one's preferences about being a parent, or the experience of becoming a vampire (we might suppose) may involve changing one's preferences about whether to be vampire or human.[22] This raises the difficult question about how practical rationality requires you to act when your current preferences diverge from the future ones. (This question is of course the close cousin of the familiar problem about whether future desires give one present reasons for action.)[23] Since standard decision theory tells you only which actions are rational in light of your current preferences and credences, it is indeed silent about what rational significance your future preferences have for you. And so if future preferences are rationally significant for present choices, then this means that one would have to either concede that decision theory is not fully comprehensive as a theory of practical reason or to find a way to extend decision theory so it provides guidance about how to act in light of preference-shift.[24] While there are genuine philosophical problems here, these challenges are again independent of epistemically transformative experiences. Though life-changing choices may involve both epistemic

[20] Formally, the problem is that incommensurable preferences are likely to be negatively intransitive—we strictly prefer A+ to A, we do not weakly prefer A+ to B, and we do not weakly prefer B to A– and incomplete: it is neither the case that we are indifferent between A and B, strictly prefer A to B, nor strictly prefer B to A. As we mentioned earlier in footnote 2, an assumption of standard decision theory is that rational agents have complete and acyclic preferences over all outcomes

[21] See Hare 2010 for a defense of prospectivism. See Bales et al. 2014 for criticism and an alternative proposal.

[22] "Your effort to evaluate testimony is complicated by the fact that even people who seemed quite anti-vampire beforehand can change their minds after being bitten, suggesting that some sort of deep preference change is indeed occurring. Although your friends, as vampires, report that they are happy with their new existence, it isn't clear that their pre-vampire selves would have been happy with the change. For example, your once-vegetarian neighbor who practiced Buddhism and an esoteric variety of hot yoga now says that since being bitten (as it happens, against her will), she too loves being a vampire. . . . Which preferences matter more? Your current, human preferences, or the preferences you'd have if you were bitten? How can you rationally choose to ignore your current preferences when making your choice? If you choose to become a vampire simply because you think that the fact of becoming a vampire will make you into a being who will be happy with the choice you've made, you are not choosing by considering your own (current) preferences" (Paul 2014, 46–47).

[23] See, e.g., Nagel 1970, Parfit 1984, Harman 2009, Brink 2010.

[24] Discussions of prefence-shift and decision theory include Weirich 1981, Ullmann-Margalit 2006, Arntzenius 2008, Briggs 2010.

transformation *and* preference shift, it seems to us that the challenges these choices pose to decision theory are just the familiar ones; epistemic transformation does not pose an additional challenge.

6 Conclusion

Life-changing decisions, such as the decision of whether to become a parent, are indeed difficult. They pose serious challenges for decision theory. And they often involve epistemically transformative experiences, too. But we have argued that, contrary to Paul, the challenges these choices pose for decision theory do not arise *because* they involve epistemically transformative experiences. Rather, life-changing experiences present us with a tangle of well-known difficulties for decision theory: the fact that our evidence about the value of future experiences is often sparse or messy, that our preferences may be incommensurable, and that these preferences may change in the future. Thus, when it comes to life-changing decisions, there are many factors that make it hard—or perhaps even impossible—to rationally decide what to do. But the fact that these decisions involve epistemically transformative experiences is not one of them.

Tom Dougherty
E-mail : tjsd3@cam.ac.uk

Sophie Horowitz
E-mail : sophie.horowitz@rice.edu

Paulina Sliwa
E-mail : pas70@cam.ac.uk

References:

Arntzenius, F. 2008. "No Regrets, or: Edith Piaf Revamps Decision Theory." *Erkenntnis* 68: 277–297. http://dx.doi.org/10.1007/s10670-007-9084-8.

Bales, Adam, Daniel Cohen, and Toby Handfield. 2014. "Decision Theory for Agents with Incomplete Preferences." *Australasian Journal for Philosophy* 92 (3): 453–470. http://dx.doi.org/10.1080/00048402.2013.843576.

Briggs, Rachael. 2010. "Decision-theory Paradoxes as Voting Paradoxes." *The Philosophical Review* 119 (1): 1–30. http://dx.doi.org/10.1215/00318108-2009-024.

Brink, David. 2010. "Prospects for Temporal Neutrality." In *The Oxford Handbook of Philosophy of Time*, edited by C. Callender. Oxford: Oxford University Press.

Burkeman, Oliver. 2013. "This Column Will Change Your Life: Transformative Experiences." *The Guardian*. http://www.theguardian.com/lifeandstyle/2013/apr/06/this-column-change-life-transformative-experiences.

Carr, Jennifer. Unpublished. "Imprecise Evidence Without Imprecise Credences."

Acknowledgements For helpful comments and discussions, we would like to thank Andy Egan, Brian Hedden, Chris Meacham, Laurie Paul, Josh Schechter, Miriam Schoenfield, Eric Swanson, the audience of the 2014 Bellingham Summer Philosophy Conference, and anonymous reviewers for *Res Philosophica*. For epistemically transformative experiences during the process of writing the paper, we would like to thank Theodore Swift and Martha Nunnenkamp.

Elga, Adam. 2010. "Subjective Probabilities Should be Sharp." *Philosophers' Imprint* 10 (5): 1–11.

Gopnik, Alison. 2013. "Is It Possible to Reason About Having a Child?" *The Wall Street Journal*. http://online.wsj.com/news/articles/SB10001424127887324432404579052901271445142.

Hare, Caspar. 2010. "Take the Sugar." *Analysis* 70 (2): 237–247. http://dx.doi.org/10.1093/analys/anp174.

Harman, Elizabeth. 2009. "'I'll Be Glad I Did It' Reasoning and the Significance of Future Desires." *Philosophical Perspectives* 23: 177–199. http://dx.doi.org/10.1111/j.1520-8583.2009.00166.x.

Harman, Elizabeth. 2015. "Transformative Experience and Reliance on Moral Testimony." *Res Philosophica* 92 (2): 323–339. http://dx.doi.org/10.11612/resphil.2015.92.2.8.

Jackson, Frank. 1986. "What Mary Didn't Know." *The Journal of Philosophy* 83 (5): 291–295. http://dx.doi.org/10.2307/2026143.

Langton, Rae. 2007. "Objective and Unconditioned Value." *Philosophical Review* 116 (2): 157–185. http://dx.doi.org/10.1215/00318108-2006-034.

Lewis, David. 1990. "What Experience Teaches." In *Mind and Cognition: A Reader*, edited by William Lycan, 499–519. Oxford: Blackwell.

Lombrozo, Tania. 2013a. "Is Having a Child a Rational Decision?" *NPR*. http://www.npr.org/blogs/13.7/2013/03/11/173977133/is-having-a-child-a-rational-decision.

Lombrozo, Tania. 2013b. "Is it Rational to Have a Child? Can Psychology Tell Us?" *Psychology Today*. http://www.psychologytoday.com/blog/explananda/201303/is-it-rational-have-child-can-psychology-tell-us.

Marshall, Richard. 2013. "Metaphysical (Interview with L.A. Paul)." *3:AM Magazine*. http://www.3ammagazine.com/3am/metaphysical/.

Nagel, Thomas. 1970. *The Possibility of Altruism*. Princeton, NJ: Princeton University Press.

Nagel, Thomas. 1974. "What Is It Like to Be a Bat?" *Philosophical Review* 83 (4): 435–450. http://dx.doi.org/10.2307/2183914.

Parfit, Derek. 1984. *Reasons and Persons*. Oxford: Oxford University Press.

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

Paul, L. A. 2015. "What You Can't Expect When You're Expecting." *Res Philosophica* 92 (2): 149–170. http://dx.doi.org/10.11612/resphil.2015.92.2.1.

Rothman, Joshua. 2013. "The Impossible Decision." *The New Yorker*. http://www.newyorker.com/books/page-turner/the-impossible-decision.

Schoenfield, Miriam. 2012. "Chilling out on Epistemic Rationality." *Philosophical Studies* 158 (2): 197–219. http://dx.doi.org/10.1007/s11098-012-9886-7.

Sturgeon, Scott. 2008. "Reasons and the Grain of Belief." *Noûs* 42 (1): 139–165.

Ullmann-Margalit, Edna. 2006. "Big Decisions: Opting, Converting, Drifting." *Royal Institute of Philosophy Supplement* 58: 157–172. http://dx.doi.org/10.1017/S1358246106058085.

Weirich, Paul. 1981. "A Bias in Rationality." *Australasian Journal of Philosophy* 59 (1): 31–37. http://dx.doi.org/10.1080/00048408112340021.

White, Roger. 2009. "Evidential Symmetry and Mushy Credence." In *Oxford Studies in Epistemology*, edited by Tamar Szabo Gendler and John Hawthorne, 161–186. Oxford: Oxford University Press.

Transformative Experiences and Reliance on Moral Testimony

Elizabeth Harman

Abstract: Some experiences are transformative in that it is impossible to imagine experiencing them until one experiences them. It has been argued that pregnancy and parenthood are like that, and that therefore one cannot make a rational decision whether to become a mother. I argue that pregnancy and parenthood are not like that; but that if even if they are, a woman can still make a rational decision by relying on testimony about the *value* of these experiences. I then discuss an objection that such testimony will be unreliable because parents will reflect on their *being glad* that their children exist, and will not realize that it's reasonable to be glad their children exist even if the parents' lives are thereby worse. I argue that despite this possible route to unreliable testimony, in general it is reasonable to rely on others' testimony about the value of their lives.

Experiences can be transformative in a number of different ways. One way an experience can be transformative is that it can be a new kind of experience, one which a person has not experienced before and cannot accurately imagine experiencing before having it first hand. Such an experience transforms, by expanding, a person's knowledge of what various experiences are like. A person who has never had any color experience before, who sees red for the first time, has an experience that is transformative in this sense. In section 1, I discuss an argument that holds that the experience of pregnancy and parenthood is transformative in this sense, and that concludes that it is impossible to make a rational decision about whether to become a mother in the way that society encourages women to make such decisions. I argue that pregnancy and parenthood are not transformative in this sense, but that even if a woman is unable to imagine what pregnancy and parenthood would be like for her, she can still make a rational decision about whether to have those experiences by relying on the testimony of others about the *value* of these experiences. This is reliance on a certain kind of moral testimony.

In section 2, I discuss a worry we might have about reliance on this kind of testimony. When a person reports on the comparative value of the

Res Philosophica, *Vol. 92, No. 2, April 2015, pp. 323–339*
http://dx.doi.org/10.11612/resphil.2015.92.2.8

life she has actually had versus an alternative life, such as a life without parenthood, she may focus on the question of whether she *is glad* to have become a parent or whether she wishes she had had the different, childless life she would have had if she had not become a parent. I argue that facts about whether one is glad that something happened are not always good guides to the comparative value of the two possibilities being compared—the possibility in which it happened and the alternative in which it did not—even when one is perfectly reasonable in being glad that the thing happened. A person may reasonably be glad to have a child she loves, even if her life would have gone better if she had not had this child. The phenomenon extends to other life events, besides becoming a parent. A person may reasonably be glad to have become the person she is; the fact that she does not identify with the person she would have been in the alternative may be sufficient to make her glad to have her actual life rather than the alternative; but this may be so even if the alternative would have been better for her. For example, a person who grew up deaf may reasonably be glad to be the person she is, shaped by her deafness as she has been, and may not wish that her deafness had been cured when she was a child (if that had been possible); this may be true because she does not identify with who she would have been in the alternative, in which case her being glad to be deaf does not support the claim that her life would not have been better if her deafness had been cured.

I draw several lessons in section 2. While there is a way that a person's judgments about the comparative value of her life (versus an alternative life) can be distorted by a reasonable attachment to the actual, that does not mean that all such judgments are distorted in this way. Discussion with a person about the basis of her value judgment can shed light on whether this kind of distortion is occurring. So my claim in section 1 survives this worry: we indeed can reasonably rely on others' testimony about the value of their lives. I also argue that a certain argument against curing deafness in babies can be seen to fall prey to a mistaken move between what a person is glad happened and what would have been best for her.

1 Parenthood as Mysterious and Unknown

In this section, I will discuss the following argument:

> Consider a woman deciding whether to become a parent; she is deciding whether to get pregnant, carry the pregnancy, and raise the created child. Society urges a woman making this decision to think about *what it would be like* to go through with it, and to base her decision on what it would be like. But being pregnant and raising a child are transformative experiences. One cannot know what it is like to have these experiences before having them, just

> as someone who has never seen color before cannot know what it is like to see red until she sees red. So a woman deciding whether to become a parent cannot make a decision based on what it would be like to be a parent, because she cannot know what it would be like. If she cannot know what it would be like to be a parent, she cannot know how valuable that experience would be for her, so she cannot compare it to the value for her of not becoming a parent. Therefore, it is not possible to make a rational decision whether to conceive and raise a child by contemplating what it would be like to do so.[1]

I will argue that this argument fails.

Let's set aside an important worry about the argument above. What about moral considerations? Self-interested considerations are not the only considerations relevant to whether one should conceive a child. Suppose one knows that if one conceives a child now, then the child will have a very short life full of suffering with no joy and no meaningful experiences. This is sufficient to settle that one should not conceive now. Suppose one knows that if one conceives a child now, then some of one's relatives, for whom one is a caretaker, including siblings and/or elderly relatives, would end up without enough food to eat and their health would be in serious danger. This is sufficient to settle that one should not conceive now. These cases provide two examples of cases in which one can make a *rational* decision against conceiving and becoming a parent. As I understand the argument above, it is not meant to apply to such cases. We are to imagine a woman who knows about her situation that it would be *morally permissible* for her to conceive and become a parent; there is no consideration present in her situation that would make it morally wrong to conceive.[2] But it is also morally permissible to refrain from conceiving. So she is in the realm of moral permissibility, and self-interested considerations regarding what it would be like for her to become a parent are certainly *relevant* to making

[1] L. A. Paul (2015) makes this argument. I present her argument in my own words. Paul is explicit that she wants to focus on the choice to conceive, carry a pregnancy, and raise the child. I think her argument doesn't turn on the distinction between becoming a parent in this way and becoming a parent in another way, such as by adopting, but at one point she suggests that the biology of pregnancy may be important to her argument.

[2] In supposing that a woman can find herself in such a situation, knowing that it is morally permissible to conceive, we are rejecting David Benatar's view that it is always morally wrong to procreate. Benatar holds that the harms people will suffer in their lives count against procreating, while the benefits they will experience do not count in favor of procreating nor can they justify procreating. Seana Shiffrin also argues that procreation is morally problematic, because we cannot get the consent of the child before creating her. (See Benatar 2006 and Shiffrin 1993.)

this decision; perhaps society even holds that they settle what she should do.[3]

I will object to two claims the argument makes.

1.1 First Objection

The argument makes this claim:

> Being pregnant and raising a child are transformative experiences. One cannot know what it is like to have these experiences before having them, just as someone who has never seen color before cannot know what it is like to see red until she sees red.

This claim is simply false. In my own case, I have a sister who is seven years younger than me. My feelings for my sister are sisterly but also, at times, parental in their nature. One of my closest friends had her first baby before I became a parent. This gave me two kinds of evidence about what it would be like to be a parent. First, I witnessed up close the relationship that she had with her baby, how it affected her daily life, and I learned a lot about what it was like for her by watching her and listening to her talk about her life; it helped of course that I knew her well. Second, I experienced my own love for her baby, which was unlike any feelings I had ever had (as an adult) for a baby. I also witnessed my own parents throughout my life, and I saw up close what their experience of parenthood was; I saw ways that it was hard for them, as well as ways that it provides good experiences to them. When I did choose to get pregnant and become a parent, I already had a lot of information about what it would be like. There was no huge revelatory experience that prompted in me the thought, "Oh, so *this* is what it is like to have a baby, to be a parent; I could not have pictured that it would feel *like this* before having the experience." On the contrary, it's been nice to experience on the inside what I had previously only witnessed from the outside. But loving my child feels much as I imagined it would. There is definitely a kind of joy I had never experienced until now. But I knew there would be.

What is the importance of the fact that in my case there was no radically new experience of a type I had not imagined? I am just one person. My own case is important because the argument makes the strong claim that it is *impossible* to know in advance what it would be like to be a parent. I did know. But I suspect that my case is really not that uncommon. Those who had their socks knocked off by the surprise of their parental feelings should

[3] Moral considerations are still relevant to a decision between two morally permissible options. The argument seems to assume that society does not recognize the fact that moral considerations can be relevant even within the realm of the morally permissible. For discussion of how moral considerations can be relevant to a decision between two morally permissible options, see Harman Forthcoming-a and Harman Forthcoming-b.

be careful not to overgeneralize and think that everyone's experiences are like theirs.

Let's also clarify something about this part of the argument. It includes this claim:

> One cannot know what it is like to be pregnant and to be a parent before having these experiences.

This might be ambiguous between the following two claims:

(i) One cannot know ahead of time which of several possible experiences of pregnancy and parenthood one will have.
(ii) One cannot know ahead of time, regarding any specific possible experience of pregnancy and parenthood that one may have, what *that experience* would be like.

The argument we are discussing makes claim (ii). Claim (i) is true, of course. A person who is considering pregnancy and parenthood knows that she might have an easy pregnancy or a difficult pregnancy, that her baby might be a good sleeper or a bad sleeper, and that she might get postpartum depression or not. She also knows that her child might get very sick. There are many uncertainties. Simply because there are uncertainties, it follows that a woman making this decision does not *know* ahead of time what it will all be like. But claim (i) is compatible with its furthermore being true that she knows, for each possibility I have mentioned, what it would be like to go through it. The argument we are discussing makes claim (ii), that for each possible path through pregnancy and parenthood, one simply cannot know what it would be like to experience that path before experiencing it.[4]

Claim (ii) is not true. My close friend had an easy pregnancy and a happy time while her child was a baby. As I've argued above, I went into my own pregnancy knowing a lot about what it would be like if my experience followed hers. Now consider the question of whether a person can know in advance what it would be like to suffer from postpartum depression. Many people, who have experienced depression already, can know a great deal about what this would be like. Others may know a lot about what it is like by witnessing postpartum depression in other people.

Given that there are different ways that pregnancy and parenthood can go, is one left unable to rationally decide what to do simply because one does not know *which outcome* will obtain? No. One can know a lot about the likelihoods of these possibilities in one's own case. Some people have risk factors for postpartum depression; and there is a certain likelihood of postpartum depression even for those without risk factors, so one can take that into account. A woman who is contemplating conceiving may have many friends who have become parents. She may see how their experiences of parenthood differ, partly due to their personalities. Reflecting on her

[4] In section 6.2, Paul clarifies that she is making this claim.

own personality, she may be able to make some predictions about what parenthood is likely to be like for her.

To summarize so far: My first objection to the argument is that, when it comes to conceiving and raising a child, it is actually possible to know a lot about what it will be like. While one should divide one's credence among different possible paths these experiences may take, one can know a great deal about what each of these paths would be like.

1.2 Second Objection

Let's continue to discuss the argument while setting my first objection aside. Suppose that the experience of conceiving and parenting really is unknowable in advance; suppose that it is like seeing red for the first time if one has never seen any colors before.

The argument makes the following assumption:

> If one cannot know what an experience would be like, then one cannot know what the value of that experience would be.

This claim is not true. One can receive *testimonial evidence* regarding the value of an experience, which can give one *knowledge* of the value of that experience. Here are two cases to illustrate:

> **Wonderful Mysterious Box:** There is a box that many people have opened. Each reports to me that the box gave them an amazing new experience, unlike anything they'd ever anticipated. Everyone is glad to have opened the box. Furthermore, they do not seem to be deluded. Each person seems to me to be living a more fulfilling life after opening the box, though I don't know what brought that on.
>
> **Horrible Mysterious Box:** There is a box that many people have opened. Each reports to me that the box gave them a horrible new experience, unlike anything they'd ever anticipated. Each seems to me to be less happy, in an enduring way, after having opened the box.

In these cases, I can know about the value of having a certain experience, though I do not know anything specific about what that experience would be like. In Wonderful Mysterious Box, if I am given the option of opening the box, and I choose to do so, I make my choice based on an epistemically justified belief that opening the box will provide an experience worth having. If the experience in fact is wonderful (and if this is not in some way a Gettier case), then I indeed *knew* in advance that it would be wonderful. My choice to open the box is a good choice and it is a rational choice. Even if it turns out that I have a horrible experience, I still made a rational choice to open the box; it just turns out that my evidence was misleading about what opening the box would be like for me. (Sometimes we make

good choices that are also rational choices but that nevertheless have bad outcomes.)

Similarly, in Horrible Mysterious Box, I have an epistemically justified belief that it would be a bad experience to open the box. I can make a good decision, and a rational decision, to refrain from opening the box.

These cases show that one can come to know whether an experience will be a good or a bad experience *without* knowing anything specific about what the experience will be like.

Now let's consider someone who has trouble imagining a lot of the basic aspects of parenthood; let's call her Cecile. It's hard for Cecile to imagine feeling so devoted to another living being as she sees parents feel. It's hard for her to imagine the kind of unconditional love that parents report experiencing. It's hard for her to imagine the brain-mushifying sleep deprivation that some parents experience. It's hard for her to imagine what it's like to have "so little time," as parents keep reporting they have. When parents ask her what she does with all her free time, she is puzzled. For these reasons and others, let's suppose, Cecile finds herself deeply unable to imagine what parenting would be like. Can she get any information about how valuable it would be for her to become a parent?

(Note that the question is whether, for a particular way that pregnancy and parenting could go for her, she can know what the value would be for her of things going that way.)

She can. Her friends and family members who are parents can tell her about what they find valuable in their lives, and they can reflect on what good and bad things parenthood has brought into their lives. Her friends and family members who have lived long lives without becoming parents can reflect on what good and bad things their lives have involved. They can give her a lot of information about whether parenthood is valuable, and what about it is valuable, even if she remains unable to imagine what it would be like to be a parent from the inside.

Of course, if Cecile tries to base a decision whether to have a child on this kind of testimony, there are some challenges she faces.

There are worries about whether people will be honest with Cecile and about whether she will receive all the information that is out there. For too long, the truth about rates of postpartum depression was not publicly discussed. It is still not common enough for women to talk openly about the ways that breast-feeding can be very difficult and burdensome; it is difficult for many reasons: it can be hard to get it to work properly, there is a great deal of societal pressure (in some communities) to do it and to have it go well, it is very taxing even if it is going well, and there is inadequate support for nursing mothers in many workplaces. Rates of Autism-spectrum disorder are on the rise, but Cecile may not know how common it is, and she may not learn about what particular parenting challenges it brings. The impact that climate change will have on the lives of children conceived now is inadequately discussed and publicized. These

are just some examples; there is a great deal of relevant information that is out there, but that many women contemplating pregnancy and parenthood may not receive.

These worries concern the quality of the information that Cecile will receive. But this does not actually cast any doubt on the claim that Cecile can make a *rational decision* about whether to conceive and become a parent. Whether a decision is *rational* depends on what evidence the agent has and whether her decision makes sense given her evidence. If Cecile's information is impaired, that is a shame, and it makes her decision potentially worse *objectively*, but it does not raise a problem for her capacity to make a rational decision.[5]

Besides testimony from particular parents and non-parents she knows, our imagined Cecile might also receive other purely normative information. A trusted advisor might tell her, "For the vast majority of people, parenthood is a valuable experience worth having. And if we restrict our attention to people with your personality type and in your situation, the vast majority of these people have richer lives as a result of becoming parents." If she has good reason to trust this advisor, and if what the advisor says is true, then (unless this is somehow a Gettier case) this can be a way that Cecile can come to know what she is told. And such knowledge can be a good basis for a rational decision to go ahead and conceive a child.[6]

I have argued that the argument I introduced at the outset of section 1 fails. I have objected in two ways. First, I argued that one can know a great deal about what pregnancy and parenthood will be like before having those experiences. (In particular, for particular ways these experiences can go, one can know about what it would be like for it to go in that way, even

[5] Paul mentions happiness studies that show that parents are often less happy than non-parents. Cecile may be unaware of these studies. As I've just said, this kind of consideration can affect the objective quality of Cecile's decision, but cannot affect the rationality of her decision. (Information of which she is unaware does not make her irrational.) But I also want to sound a note of caution about the significance of these studies. As I understand parenthood, it provides something deeply valuable. What makes it valuable is not primarily anything to do with ordinary happiness. It is, in Mill's sense, a higher-order pleasure, an experience with deep meaning. If it reduces a person's ordinary everyday happiness, it may still be overall more valuable than the alternative for that person. (Kauppinen [2015] also stresses the importance of considerations of meaningfulness, in discussing Paul's argument.)

[6] We might worry that it raises a red flag about Cecile's suitability to become a parent *that she cannot imagine it*. The fact that she can't picture herself as a parent may suggest that she's not suited to become a parent. After all, if she doesn't find herself longing for a future she imagines in which she is a parent, surely she's not the right type to be a parent. This worry can be addressed in two ways. First, it may be that Cecile does *desire* to be a parent though she cannot *picture* what it would be like; these may come apart. Second, I think it is actually pretty common for people to choose to become parents, because they are told by others that ultimately they will find it to be a valuable part of life, worth the sacrifices, although they cannot picture it yet, and they do not find themselves desiring it. The stereotypical person who has this kind of experience is a man; I would even say it is somewhat common for men to decide on parenthood in this way. But it could happen to a woman as well, and I'm sure it does. Many people who decide in this way end up loving parenthood.

if one doesn't know for sure which experience one will actually have if one goes ahead.) Second, I argued that one can receive testimonial evidence about the value of certain experiences, and make a rational decision on that basis, even if one doesn't know anything specific about what the experiences would be like.[7]

One might try to defend the argument by saying that my Second Objection is off-target. The argument simply concluded:

> It is not possible to make a rational decision whether to conceive and raise a child by contemplating what it would be like to do so.

My suggestion—that someone can gain testimonial evidence about the value of becoming a parent—may not be a way of deciding "by contemplating what it would be like" to conceive and raise a child.

Here we can go one of two ways. We can interpret the original argument narrowly, so that my first objection directly touches it but my second objection does not. In that case, the conclusion of the argument is more narrow and less interesting than it at first appears: it is compatible with the claim that one *can* decide rationally by discussing the choice with others, if those conversations include specifically value-laden information. Alternatively, we can interpret the original argument as drawing the more interesting conclusion that simply contemplating the choice to procreate, and talking to one's friends about it, does not put one in a position to make a rational decision whether to procreate. Understood in this way, the argument is prey to both my objections.[8]

2 Testimony about the Value of a Life Path

In section 1, my second objection relied upon the following idea:

> (*) We can gain knowledge about the comparable value of two different life paths by talking to people who have taken these life paths, and in particular by hearing their testimony about the *value* of their own lives as compared to alternative life paths.

There are a number of different worries that might be raised about reliance on such testimony. Each person has only lived her own life, so when she compares her life to an alternative, her own information is asymmetric: she knows only one of those life paths from the inside. There are well-known psychological phenomena that we might worry distort people's judgments, such as the sour grapes effect (which would lead people to downgrade good things in the alternative) and the grass is always greener effect (which

[7] See Dougherty et al. 2015 for a similar objection to the argument.

[8] I interpret Paul as making the more interesting version of the argument.

would lead people to upgrade good things in the alternative). Other worries might be raised as well.

In section 2, I will focus on a different worry that we might raise. The worry arises from one route that a person might use to arrive at a judgment about the comparative value of her actual life path versus an alternative life path. She might ask herself whether she is glad to have lived the life she did, or whether she wishes she had lived the alternative life. This way of arriving at a judgment about the comparative value of the two life paths is unreliable, I will argue. There is a systematic distortion due to our *reasonable attachments to the actual*. I will develop this idea by first discussing a kind of *decision-making* that can go awry in a systematic way.[9]

At the end of section 2.2, I will argue that this worry does not undermine (*). Despite this worry, it is indeed possible to gain knowledge about the comparable value of alternative life paths by listening to people's testimony about this issue; and it is reasonable to rely on such testimony in making life choices.

2.1 "I'll Be Glad I Did It" Reasoning

It is very natural, in trying to decide whether to do something, to project oneself forward into the future in which one has done it, and to consider how one will feel about that decision. If one would regret the decision, it seems that one should not make it. If one would be glad one did it, it seems that one should do it. This is often an excellent way to reason. "If I study for my exam, tomorrow I'll be glad I did it. So I should study for my exam" is good reasoning.

This kind of reasoning is natural and appealing when dealing with big life decisions too, not just with small questions such as what to do this evening. When deciding whether to marry someone, which career to pursue, and which college to attend, it is natural to try to imagine oneself in the future having made a certain choice, and to try to figure out how one would later feel about having made the choice.

But, I will argue, this kind of reasoning can be bad reasoning, when it is used regarding choices that are *transformative* in a particular kind of way.

Consider a fourteen-year-old girl in the United States who is deciding whether to try to conceive a child. Some of her friends have children, and though having become mothers at such a young age is clearly very hard for them, she can see that they deeply love their children. She might reason as following:

> "If I conceive a child now, I will raise this child and love him or her very dearly. I will not wish I had not conceived, because then I would not have had *this child* whom I will

[9] Section 2 of this paper builds on some ideas from Harman 2009.

> love so much. I will be glad I conceived when I did. So, I should conceive now."

This is clearly bad reasoning. But what distinguishes it from the reasoning regarding studying for an exam?

The difference, I claim, is that some experiences transform what it is reasonable for us to prefer. *Loving someone* makes it reasonable to prefer that she exists over an alternative in which she does not exist. Once a person has created and started raising a child, her love for the child makes it reasonable to prefer that the child exist. But before the child has been created, the person does not already love the child, and so *love* is not available as a basis to make a preference to procreate now reasonable. Other factors matter, such as what would be best for the person's own life: procreating now, waiting and procreating later, or never procreating.[10]

In the exam case, there is no transformative experience that alters what the agent is reasonable in caring about. Both later and now, she wants to do well on her test and she wants to have enjoyable experiences like going to a movie. If she studies, she will be glad she did so, partly because she will deem doing well on the test to be more important to her than going to the movie. For the same reason, she should prefer to study now. Both tonight and in the morning, it is more important to her to do well on the test than to go to the movie.

In the absence of a transformative experience that affects what it is reasonable for an agent to care about, predicting that one will be glad to have done something is a good reason to believe it would be reasonable to now prefer to do that thing. But when doing something would bring with it a transformative experience—and in particular, a transformative experience that brings with it a *reasonable attachment*—then the prediction that one would be glad to have done it is not evidence that it is reasonable to prefer to do it now, before one has that attachment.

This is a second sense in which an experience can be *transformative*. Experiences are transformative in the sense I discussed in section 1 if it is impossible for a person to accurately imagine what it is like to have them before having them. Experiences are transformative in the sense I discuss here in section 2 if they transform what basic preferences it is reasonable for a person to have.[11]

[10] One might claim that the teenager's reasoning is bad reasoning because if, in the future, she will indeed be glad she did it, that will be *unreasonable*: because it is a bad idea to conceive now, in the future it would be unreasonable to be glad to have conceived. This claim makes a serious mistake about the nature of reasonable preference. It is indeed reasonable to prefer the people we love, and the lives we have actually had. There is nothing unreasonable about these preferences, though it would be unreasonable to make certain inferences on the basis of these preferences, such as the inference that things would not have been better for us in the alternative.

[11] Any experience that simply provides information might transform what preferences it is reasonable to have. I might reasonably prefer to drink the liquid in front of me; once I learn that it is bleach rather than water, what it is reasonable for me to prefer changes. But there is

Let's consider two more instances of "I'll be glad I did it" reasoning, and ask whether they exhibit good or bad reasoning.

There is much controversy in the deaf community regarding cochlear implants, which can enable some deaf people to process some sounds and to function in non-signing communities, schools, and work environments. Cochlear implants are more effective the earlier in life they are implanted, and so there is a debate about whether they should be given to babies.

There are a number of different arguments against cochlear implants. Some arguments focus on the fact that surgery is required to do the implantation, that successful integration into hearing communities is more likely in some situations than others, and that cochlear implants fall well short of giving deaf children the same kinds of auditory experiences that those who are born hearing have. These arguments focus on the specific shortcomings and limitations of cochlear implants as they exist today.

But some arguments against cochlear implants provide in principle opposition to implants; these are arguments against curing deafness, even if a safe and full cure were possible. These arguments rely crucially on the following claims, made by deaf adults (who grew up deaf): "We are glad to have grown up deaf. We do not wish that we had been cured of deafness as children, if that had been possible." Let's consider how the parents of a deaf baby might reason, if they were considering whether to choose a safe and full cure for deafness for their baby:

> "If we do not cure our baby of deafness, then she will grow into a deaf adult whose life has been shaped, in part, by her experience of growing up deaf. We will love her as the person she is, and we will not wish she had been an utterly different person. We will be glad that we did not cure her deafness. So, we should not cure our baby of deafness."

This reasoning is bad reasoning, I claim. I hope it is clear that it is bad reasoning. While it is reasonable as an adult to be glad to be the person one in fact is, and it is reasonable to love one's child as the person she is, it does not follow that it would have been reasonable to prefer *this life path* for one's child at an earlier choice-point, before she had already been shaped by the path. It may be that another life path would be better for one's child, although it would be so different that one would later not wish for it if one did not take it.

Finally, let's consider an argument regarding abortion. Some advocates against abortion make moving personal speeches. "I was almost aborted," they say. They appeal to the reasonable preference that anyone might have for the people who actually do exist. Whether we know her or not, we are glad she exists; we are glad she was not aborted. Does this mean that her

---

no change in my *basic preferences* in a case like this. I still prefer to drink water and prefer not to drink bleach.

pregnant mother should have had the same preference, back when she was pregnant? It does not.

Consider the following reasoning that a pregnant woman might engage in:

> "It would be very difficult for me if I continue this pregnancy and raise a child at this time in my life. But if I do, I will love my child dearly and be glad to have him or her. I will not wish I had aborted, because then I would not have had my child. Therefore, I should continue my pregnancy."

Again, this is bad reasoning. The fact that the woman would have a reasonable attachment if she continues the pregnancy does not mean that she should decide in favor of the child *now*, before she loves the child and is attached to him or her.

Now that we've seen that "I'll be glad I did it" reasoning is often bad reasoning, should we conclude that it is never good reasoning? We should not. The exam reasoning is good reasoning. In the exam case, the considerations that in fact make it the case that one will be glad one did it are also the considerations that in fact make it the case that one should do it. (Note that I am not saying that *the fact that one will be glad one did it* is what makes it the case that one should do it.) That is, one is glad one did it because it really is more important to pass the exam than to see that movie at that time, and these are the same considerations that make it the case that one should study.

So how should we understand the cases in which "I'll be glad I did it" reasoning is bad reasoning? We can see these as cases in which the agents *are in a position to realize* that their predicted future attitude of being glad to have done it *would be due to a reasonable attachment* and would not be sensitive to the considerations that are relevant now to how they should make these choices. They do not already have these attachments: the teenager does not already have a child she loves; the parents of the deaf baby do not already know and love their child as a person who has grow up deaf; the pregnant woman does not already know and love her child. So, the attachments are not available to ground reasonable preference now for outcomes that may actually be worse, and may be the result of bad choices.

What I am claiming is that sometimes, after one makes a bad choice, or after something bad happens to one, one may nevertheless have a reasonable attitude of *being glad that one did it*, or *being glad that the thing happened*, because one may be reasonably attached to how things actually are. Reasonable attachments can lead us astray, as in the bad "I'll be glad I did it" reasoning above. But reasonable attachments can also be understood in a clear-eyed manner. A woman who became a parent as a teen might say, truly, "I should not have had a child as a teen. But I love my son and I'm so glad I did, because otherwise I wouldn't have had him. That I love him and am glad to have had him—that I would not wish to change

anything for myself—in no way makes me think that teen parenthood is a good choice for anyone to make."

2.2 General Moral Arguments and Moral Testimony

There are moral arguments that are sometimes made that are analogues of the "I'll be glad I did it" reasoning discussed above. A deaf adult might argue that curing deafness is morally wrong as follows:

> "I would have been a completely different person if I had been cured of deafness as a baby, if that had been possible. But I do not wish I had been cured of deafness, and no one should wish that I had been cured of deafness. So, no one should wish to cure deafness in babies today. So it would be morally wrong to cure deafness in babies today."

An anti-abortion activist might argue:

> "I was almost aborted, but I am glad I was not. Everyone should be glad I was not. So everyone should prefer not to abort babies now. So it would be morally wrong to abort babies now."

Like "I'll be glad I did it" reasoning, these arguments move from a claim that a certain predicted future *backward-looking* preference is reasonable (indeed, is a preference we should have) to a claim that this preference should be had now *prospectively*, before a choice and its effects have happened. Because the predicted future preferences in question are reasonable and appropriate only due to the way we are and should be *attached* to actual people who have actually turned out a certain way, that they are reasonable does not mean it would be reasonable (or that it is required) to have the same preferences now.

As I have discussed these arguments so far, they simply make a mistake about the nature of *reasonable preference*: what would make a preference reasonable in the future does not make the preference reasonable now. But we can also understand the deafness argument as implicitly relying on moral testimony; it implicitly relies on testimony about the value of certain life paths. We can see the argument as implicitly including the following line of thought:

> "I, a deaf adult who grew up deaf, am glad to have grown up deaf; I do not wish I had been cured of deafness as a baby, if that had been possible. *Therefore*, my life having grown up deaf is no worse than a life having grown up hearing."

This implicit reasoning is flawed. One cannot generally move from the claim that one reasonably is glad that something happened to the conclusion that it would not have been better if that thing had not happened—not even

that it would not have been better *for oneself*. People can be reasonable in being attached to being the particular people they are. One might prefer one's own life to a radically different life one might have had, even if the other life would have been better, if one cannot relate to who one would have been if one had had that life.

Thus, we can see that we must be cautious in accepting people's testimony about the value of their own lives. A person might mistakenly move from the claim that she does not *wish* her life had been different in a certain way—which may be perfectly reasonable—to the claim that her life would not have been better in that alternative. This inference is unwarranted when the experience at issue is a *transformative experience*, one that transforms what it is reasonable to prefer. Many transformative experiences make it reasonable to prefer worse outcomes, and to be glad after the fact to have made bad choices.

While there is reason to be cautious, these considerations do not imply that we should in general be reluctant to believe people's testimony about the value of their own lives. Rather, there is a particular pitfall of reasoning of which we should be aware. One thing this means is that when a person offers testimony about the value of her life (compared to an alternative), now that we are aware of this potential mistake, it makes sense to probe a bit into the reasons that she makes this judgment about the value of her life. If her main reason seems to be that if things had been different, then she would not have had something to which she is deeply *attached* (such as a child, or her child's particular personality, or her own particular personality), then we have reason to doubt her claims. But if her judgment seems to involve genuine reflection on what is good and bad about her actual life versus what would have been good and bad about the alternative, then the worry I have raised does not give us reason to doubt.[12]

A person's judgment of how the value of her life compares to an alternative may well be based on a genuine recognition of the truth about the value comparison between the two life paths. If such a person offers testimony about the comparative value of her life, and someone believes what she says on the basis of her testimony, then this is a way for the listener to gain some knowledge about the comparative value of these two life paths. The following claim is indeed true:

> (*) We can gain knowledge about the comparable value of two different life paths by talking to people who have taken these life paths, and in particular by hearing their

[12] It is an interesting question whether the experience of living life as a deaf person is a transformative experience in the first sense (discussed in section 1); is it impossible for someone who has not had such a life to know what it is like to have such a life? I am doubtful that living as a deaf person is transformative in the first sense, not because I think it is easy to understand what it is like, but because I think that people have wonderful capacities for story-telling and description of experiences, so that people in telling their own stories can communicate a great deal about what their lives have been like.

> testimony about the *value* of their own lives as compared to alternative life paths.

What should we conclude specifically about the kind of testimonial evidence that Cecile, our imagined prospective mother from section 1, receives?

Because parents love their children, it is reasonable for them to be glad to have had their children, and to not wish to have remained childless. A parent might mistakenly conclude that her life is better for having had her child simply because she prefers things as they actually are. This might happen for a parent whose life is actually much worse than it would have been if she had not become a parent, and even for someone who is in a position to realize that. If such a person tells Cecile that having a child made her life better, then Cecile is receiving some misleading information. But this doesn't mean she can't make a rational decision.

We might think that Cecile should be able to diagnose that parents are making this mistake, which would make her irrational to rely on their judgments. This seems right. *If* it's true that Cecile should be able to see that a particular parent's saying "my life is very valuable as a result of having my child, more valuable than it would have been if I had remained childless" is a result of the parent's simply being attached to her child, and is insensitive to any comparison of value between the two outcomes, *then* she is making an epistemic mistake if she takes this value judgment on board and relies on it. But this does not show that in general we cannot rely on the testimony of parents.

Parents and non-parents who make judgments about the comparative value of their lives need not be making the mistake I have outlined in section 2. As I have already mentioned, it is perfectly possible to look back clear-eyed at a bad choice to procreate and think, "I should not have made that choice. But I love my child, and don't wish I'd chosen differently." But similarly, it is possible to look back clear-eyed at a good choice to procreate, and think, "I'm glad I made my choice, because I love my child. The fact that I'm glad doesn't tell me whether I made a good choice. But when I think about what I gained by having a child, what it cost me, and what my alternative life would have been like, I can see that my life is richer and deeper for having had a child. This choice was for the best, for me." That speech could be true, and it could be an expression of the speaker's *knowledge* about her situation. Even people who have not articulated all of that to themselves may be expressing genuine knowledge when they offer testimony about the comparative value of their lives.

3 Conclusion

It is an interesting question how inaccessible some experiences are to our understanding if we have not had them. I discussed an argument according to which pregnancy and childbirth are transformative experiences in that it

is only by having them that a person can know what they are like. I argued that pregnancy and childbirth are not inaccessible to us in this way, but that even if they are, we can still make rational decisions whether to have these experiences by relying on the moral testimony of others about the value of their experiences.

I then went on to argue that there is a certain kind of mistake that people can make in forming judgments about the value of their own lives versus alternative lives. A person might reasonably be glad to have had the life she has actually led—she might have a reasonable attachment to the actual—even though the alternative would have been better for her. I argued that although this kind of mistake is possible, we can still reasonably rely on others' testimony about the comparative value of their lives versus alternative lives they could have led.

Elizabeth Harman
E-mail : eharman@princeton.edu

References:

Benatar, David. 2006. *Better Never to Have Been*. Oxford: Oxford University Press.

Dougherty, Tom, Sophie Horwitz, and Paulina Sliwa. 2015. "Expecting the Unexpected." *Res Philosophica* 92 (2): 301–321. http://dx.doi.org/10.11612/resphil.2015.92.2.5.

Harman, Elizabeth. 2009. "'I'll Be Glad I Did It' Reasoning and the Significance of Future Desires." *Philosophical Perspectives* 23: 177–199. http://dx.doi.org/10.1111/j.1520-8583.2009.00166.x.

Harman, Elizabeth. Forthcoming-a. "Morality Within the Realm of the Morally Permissible." *Oxford Studies in Normative Ethics*.

Harman, Elizabeth. Forthcoming-b. "Morally Permissible Moral Mistakes." *Ethics*.

Kauppinen, Antti. 2015. "What's So Great about Experience?" *Res Philosophica* 92 (2): 371–388. http://dx.doi.org/10.11612/resphil.2015.92.2.10.

Paul, L. A. 2015. "What You Can't Expect When You're Expecting." *Res Philosophica* 92 (2): 149–170. http://dx.doi.org/10.11612/resphil.2015.92.2.1.

Shiffrin, Seana. 1993. *Consent and the Morality of Procreation*. Ph.D. thesis, Oxford University.

Transforming Others: On the Limits of "You'll Be Glad I Did It" Reasoning

Dana Sarah Howard

Abstract: We often find ourselves in situations in which it is up to us to make decisions on behalf of others. How can we determine whether such decisions are morally justified, especially if those decisions may change who it is these others end up becoming? In this paper, I will evaluate one plausible kind of justification that may tempt us: we may want to justify our decision by appealing to the likelihood that the other person will be glad we made that specific choice down the line. Although it is tempting, I ultimately argue that we should reject this sort of appeal as a plausible justification for the moral permissibility of our vicarious decisions. This is because the decisions that we make on behalf of another may affect the interests and values that that person will hold in the future. As I will show, this complicates the justificatory relationship between present decisions and future attitudes, since the latter can depend on the former.

In the course of our lives, we undergo certain experiences that have the power to transform us. Such experiences radically change how we perceive the world and may alter what of the world we most value. Moreover, the way in which we will be changed cannot be fully understood or imagined prior to undergoing the experiences themselves. Recently philosophers have considered how the potential to undergo these sorts of personally transformative experiences should influence our decision-making. Elizabeth Harman (2009), for example, argues that the fact that we (or someone we love) will undergo certain transformative experiences—such as getting cochlear implants, joining the army, or having a child—limits what we can learn from our predicted future preferences that result from these changes.[1] L. A. Paul (2015) asks whether imagining the phenomenal character of undergoing such experiences can ever provide us with a reason to actually

[1] Although Harman (2009) does not use the terminology of "transformative experiences," she does talk about how deafness is "transformative of one's character" (192) and how people can be reasonable to prefer "transformative traits" (197) even if it is worse for them to have such traits. Harman (2015) makes it explicit that undergoing such experiences can be transformative.

Res Philosophica, *Vol. 92, No. 2, April 2015, pp. 341–370*
http://dx.doi.org/10.11612/resphil.2015.92.2.9

undergo them. These investigations have made significant strides in figuring out how to approach life choices that are concerned in part with whether to subject ourselves to such experiences in the first place. However, we should keep in mind that not all our transformative experiences are up to us. While we may have some choice in the matter whether to have a child or go to graduate school, many transformative experiences will happen to us without our choosing, either because of luck or because of the choosing done by others on our behalf.

Consider cochlear implant surgery.[2] This seems to be a paradigmatic case of a transformative experience, yet more and more often it is one that people do not choose for themselves, rather it is chosen for them by someone else—namely, their parents or guardians. This is because cochlear implants are generally taken to achieve the most success if they are implanted during early stages of the child's language and speech development (i.e., before the age of 3) (Connor et al. 2006).[3] The decision concerning cochlear implants can be a difficult one for parents to make. They are not merely deciding whether to change some physical attribute of the child; their decision also involves choosing the texture of the world that this child will inhabit and the sort of person the child can become. The situation is further complicated by the fact that 9 out of 10 deaf children are born to hearing parents, making it hard for parents to imagine the possible future perspectives of the child and for the child to imagine the perspective of her hearing parents.[4] Finally, there seems to be reasonable disagreement about whether cochlear implant surgery at such a young age is in the best interest of the child. While proponents of the surgery champion cochlear implants as giving deaf children greater access to the hearing world, opponents worry that implant proponents both misunderstand deafness and overpromise what the implants can deliver. Given these challenges, how ought parents go about deciding whether pursuing or forgoing cochlear implants for their child is a morally justified course of action? And, more generally, what sorts of considerations can play a justifying role in the decisions we make

[2] Cochlear implant technology uses electric stimulation of the auditory nerve to help users who are profoundly deaf to perceive sound, and in particular, to perceive speech.

[3] I should note that what counts as "success" for cochlear implants is a matter of contention. Whereas medical researchers often measure success primarily in terms of oral language fluency, deaf advocates have argued that cochlear implant success should be determined by whether the child has access to language acquisition more broadly (through oral communication, through signing, or through a combination of both). If cochlear implant surgery at a young age restricts the child's access to sign language, this may have deleterious results for the child's overall language and emotional development even as it presents the most acute improvements in terms of oral language fluency. On such occasions, surgery done later in the child's development may turn out to be more successful for the total wellness of the child even if his/her speaking fluency is diminished. For further information see the position offered by the National Association of the Deaf on cochlear implants: http://nad.org/issues/technology/assistive-listening/cochlear-implants. I thank an anonymous referee for pushing me to clarify this point of contention.

[4] See http://www.nidcd.nih.gov/health/statistics/pages/quick.aspx, last accessed Oct. 28, 2014.

on behalf of others, especially if those decisions may change who it is these others end up becoming?

To begin answering this broad question, I will evaluate one kind of justification that may tempt us when we are charged with the role of making decisions on behalf of others, which I call 'Predictive Glad.' A more formal account will be on offer in section 2, but for now, one could say that when we employ Predictive Glad, we rely on the prediction of people's future pro-attitudes to justify a present action. Here is the general phenomenon in which Predictive Glad may tempt us in our practical thinking: Sometimes we are thinking of doing something to someone else or of making a decision on someone else's behalf and we worry about whether it is morally permissible to treat this person that way (e.g., should we throw John a surprise fortieth birthday party? Should we make little Madison take SAT prep classes?). We wonder whether there is any decisive moral complaint against the action *on that particular person's behalf*. Given these concerns, we might reason as follows: "May I treat John this way? If I do, he'll be glad I did it. So it is okay to treat John this way." Or, "Should I treat Madison this way? I know that she will complain now about the classes, but when she gets into Fancy Pants College, she'll be glad I made her suffer through them. So it's okay to treat Madison this way." Predictive Glad thus arrives at the conclusion that an action is a morally permissible way to treat *a particular person*. It does not arrive at a conclusion that we *should* act in a particular way, nor at the conclusion that a particular way of acting is *all things considered morally permissible*. After all, other people may be involved.[5] While John may be glad we threw him a surprise party, Paul - John's husband—has been planning to cook John a nice meal for his birthday and our party will overshadow Paul's efforts. While Madison will be glad for the academic leg up, she already has been unduly advantaged by her fancy private school education and my financial resources would be more justly spent elsewhere.[6]

Although Predictive Glad can plausibly justify all sorts of decision-making, my focus in this paper will be on parents making decisions on behalf of their children. This focus allows us to bracket certain important questions about when our decisions on behalf of others are morally permissible, like whether or not we have the authority to make such decisions. But focusing on parents also highlights a feature of deciding on behalf of others that I think is present in many such cases but often goes unnoticed. That

[5] I am indebted to Elizabeth Harman for suggesting this phrasing as the normative significance of Predictive Glad and for helping me clarify how "You'll be glad I did it" reasoning can serve a role in our practical thinking that is distinctive from how we engage in "I'll be glad I did it" reasoning.

[6] Note that if we are not in the appropriate relationship with these particular others, it will be difficult to predict whether an altruistic seeming action on our part will generate the requisite future pro-attitudes on the part of the other person (e.g., Will John be glad to have a surprise party thrown by his Fed Ex delivery guy? Will Madison be glad her friend's parents forced her to take the SAT classes?).

is, the interests and values of these others are to a greater or lesser extent not fixed or predetermined. So the decisions that we make on their behalf may affect the interests and values that they will hold in the future. As I will show, this complicates the justificatory relationship between present decisions and future attitudes, since the latter can depend on the former. Ultimately, I will argue that we should reject Predictive Glad as a plausible justification for making a decision on someone else's behalf. Although I do think that the predicted future attitudes of others *can* play a significant role in justifying our decisions on their behalf, they can play this role only when we consider the future attitudes of all relevant possible futures.

1 I'll Be Glad and You'll Be Glad

Let's start by considering a case concerning cochlear implants as it is put forward by Harman (2009). Imagine a hearing mother is deciding whether or not to give her baby, Stevie, cochlear implants to counteract his deafness. Other things being equal, she would prefer not to have little Stevie undergo invasive surgery unnecessarily, so forgoing the implants is her default choice. She then considers whether this choice is justified. She predicts that if she does not give Stevie the implants, he will grow up to lead a fulfilling life that will no doubt be greatly influenced by his deafness. Through her interactions with adult friends who are deaf, she has come to realize that they greatly value their membership in the Deaf community. Her friends take their participation in Deaf culture to play a significant role in their life and in how they have forged their identity; they find features of this culture to have no clear counterparts in the hearing world and so they are glad they are deaf. 'Glad' here does not merely connote an emotional state; for her friends to be glad is for them to have *preferred*that things turned out as they did rather than turning out some other way. If they could do it all over again, they would still have preferred for their parents to forgo the implants.

Not only does the mother recognize that her friends are glad they are deaf, she also recognizes that *she* is glad that her friends are deaf since she acknowledges that their character has been shaped by their deafness. Likewise, she predicts that when Stevie will grow up, she will love Stevie for the person that he will have become and that person will be significantly shaped by his physical condition. She will be glad that she chose to forgo the implants and would not wish to have chosen otherwise. The following argument using what Harman calls, "I'll be glad I did it" reasoning underlies her justification for not giving Stevie the cochlear implants:

Deafness Argument:

(1) If I do not cure my baby of deafness, I'll be glad I made that choice.
(2) Therefore, I should not cure my baby of deafness. (2009, 178)[7]

[7] I am using Harman's language of cochlear implants as presenting a "cure" for deafness. Many members of the Deaf community argue that such terminology is misguided both because

Harman thinks this is a bad argument and so do I. However, we diverge on where and how this argument goes wrong. First, Harman doesn't think that it leads to the right practical conclusion. Since Harman contends that it is worse to be deaf than it is to be hearing, she infers that the mother must be making *some* mistake in her practical reasoning if she ends up concluding that forgoing the surgery is justified (2009, 189). This is the case for Harman even though she acknowledges that the mother can reasonably predict that, were she to forgo the surgery, she would be glad down the road that she chose as she did and that this future gladness would itself be reasonable.

I do not share Harman's intuition that it is worse to be deaf than it is to be hearing—especially when we think about the lives of particular people. I don't think we have sufficient reason to believe that *Stevie's* life on the whole will go better for him were he to undergo the cochlear implant surgery as a young child. While being deaf and implant-free may limit one's possibilities in important ways (or alternatively, may make certain possibilities more difficult to attain), this does not imply that the possibilities left open or those that open up are inherently worse than those that have been closed off. Harman claims that whatever unique experiences one is provided by forgoing the cochlear implants, these "do not outweigh what is lost" (2009, 189). Much more needs to be said to make good on this sort of comparative claim and it is one that many members of the Deaf community will not concede. In my argument, I will try to steer clear of these types of comparisons of what sort of life is worse or better for a person and what sorts of benefits outweigh which burdens. So in contrast to Harman's analysis, I think that the mother may actually be arriving at the *right* decision, but by engaging in some faulty reasoning.

In this paper I show how Deafness Argument goes wrong though I do not maintain the controversial claim that being hearing is better than being deaf.[8] To highlight this, I should add that the position I will defend holds the converse argument to be bad as well (and not because I think that it is worse to be hearing than it is to be deaf):

Cochlear Implant Argument:

(1) If I give my baby cochlear implants, I'll be glad I made that choice.
(2) Therefore, I should give my baby cochlear implants.

---

deafness is not the sort of condition that needs a cure and also because cochlear implants do not actually cure deafness, since they cannot make someone hearing. Rather the implants, if successful, allow profoundly deaf people to approximate hearing, which is most helpful in oral communication in the hearing world.

[8] I should note that Harman is careful not to rely on this assumption in order to argue that Deafness Argument is an instance of bad reasoning. Her intuition that it is worse to be deaf than to be hearing serves as evidence that Deafness Argument goes wrong somehow, though the intuition does not explain what makes Deafness Argument go wrong. Rather she argues that the argument goes wrong because it relies on the appealing but fallacious "Reflection for Desires" principle (2009, 182–184).

While I think both of these arguments are problematic, I don't think either obviously leads to the wrong practical conclusion. This is because, as I mentioned in the introduction, the decision regarding whether to choose for one's young child to undergo cochlear implant surgery is a hard one to make. But just because it is a hard decision on a controversial matter doesn't mean that there is no right thing or wrong thing to do in the situation. Now there are some arguments that are structured very much like Deafness Argument and Cochlear Implant Argument, which I do think *more obviously* lead to the wrong practical conclusion. Consider the following supposedly true story by A. J. Liebling:[9]

> **Happy Old Clown:** "One of the last of the Fratellini family of clowns, an old man, made a television address in Paris a few years ago in which he [offered an explanation] for the dearth of good young circus clowns. 'When I was a child, my father, bless him, broke my legs, so that I would walk comically, as a clown should,' the old man said. . . . 'Now there are people who would take a poor view of that sort of thing.'" (1962, 149)

It is not hard to imagine how the old Fratellini clown would have turned out differently—with a different set of aims and values—if not for the actions that his father took long ago. And this old clown now seems to fully endorse the measure his father took; a measure which, no doubt other people consider unconscionable. Imagine the Fratellini father invoking Predictive Glad reasoning: "Being a clown is the best profession in the world. I love doing what I do and I can predict that my son will love it too. But the only way for my son to be a proper clown like me and like his grandfather and his great-grandfather is to break his legs now so that he walks comically as any clown should. I know that this is a big burden for my son to bear now but I predict that someday my son will be glad that I did it. After all, I am glad that my father broke my legs when I was a child."

We can reframe the father's reasoning so that it shares a similar structure with Deafness Argument:

> **Happy Old Clown Argument:**
> (1) If I break my son's legs, he will be glad I did it.
> (2) Therefore, I should break my son's legs.

If the father could have reasonably predicted *his son's* acceptance of the measure, was he justified at the time in breaking his son's legs? My intuitive answer to this question is, obviously no. So whereas Deafness Argument

[9] It is noteworthy that Liebling (1962) offers us this snippet as he laments the changing child labor laws and compulsory education laws in France that prohibited the training of children as apprentices in restaurants from a young age. He suggests that since chefs are no longer raised in the kitchen, the quality of the top Parisian restaurants in general has been on the decline.

doesn't strike me as an obvious challenge to Predictive Glad justifications, I do think that Happy Old Clown makes a strong case that there is something fishy about these kinds of justifications.

I will argue that Cochlear Implant Argument, Deafness Argument, and Happy Old Clown Argument all suffer from the same structural problem: they all appeal to the future attitudes that result from a single course of action, while ignoring the likely future attitudes that would result from the competing courses of action. I will then offer an approach for how to look at the children's future attitudes in light of these alternatives, and in doing so come to different conclusions in the cases of the Fratellini father and of Stevie's mother. So the future attitudes of others *can* play a justificatory role in our practical reasoning, but not in the way that these arguments suggest.

There is a second way in which I think Deafness Argument goes wrong: as Harman presents it, the argument is grounded in the mother's, and not in little Stevie's, future attitudes. While it may be the case that much of what would make the mother glad about her choice depends on her son's future attitudes, Deafness Argument does not require this to be the case. However, this doesn't strike me as how parents usually make decisions on behalf of their children (at least important life altering decisions such as this one). In all likelihood, parents will allot a central role in their deliberation to *their child's* future attitudes about the decision.[10] Attending to Stevie's future attitudes rather than merely her own seems like a clear-cut way to improve the mother's reasoning in this case. Thus whereas Harman focuses on the merits of "I'll be glad I did it" reasoning, I will consider the corollary, "You'll be glad I did it" reasoning.

There is a sense in which this shift from a first-person to a vicarious outlook is a friendly amendment to Harman's approach. While ostensibly an easy fix, I think the shift does reveal a fundamental difference between the type of reasoning that Harman calls, "I'll be glad I did it" reasoning and the type of "You'll be glad I did it" reasoning we engage in when making decisions involving others. "I'll be glad I did it" reasoning is concerned with figuring out what one ought to do all things considered. It relies on the prediction that one will be glad one did something in the future to draw the conclusion that one *should* do that thing at present. According to Harman, this prediction can provide for the agent an epistemic justification for believing that she should do something rather than actually justifying doing it. Look at the following argument in which Harman believes "I'll be glad I did it" reasoning is employed to good effect:

Paper Argument:

(1) If I work on my paper, I'll be glad I did it.

(2) Therefore, I should work on my paper. (2009, 177)

[10] In this very limited way, Happy Old Clown Argument can be seen as making a better argument than Deafness Argument.

Harman asks us to imagine her in the fairly routine scenario of deciding what to do one evening: should she continue working on a paper or go out to a movie? She reasons that if she continues to work on the paper, she would be glad she did it the next day. This realization is enough to convince her that she *should* continue working on the paper. However we should be clear that her future preferences don't in themselves justify her continuing to work on the paper. What justifies continuing to work on this paper is that she has a deadline to meet or that the paper will significantly improve with the extra attention or that she will experience a sense of accomplishment from her work. Her predicted future preference is only indicative of the fact that working on the paper is the justified thing to do on this occasion. There are many cases like this one. So Harman concludes that the prediction that we'll be glad we did something is a good but defeasible reason to believe that we should do it.

"You'll be glad I did it" reasoning, on the other hand, is employed to determine whether actions that we take on behalf of others are morally permissible. When a parent says to her child, "You may not like piano lessons now, but you'll be glad I made you take them," she is not articulating an all things considered reason she has to make her child take the piano lessons. Candidates for an all things considered reason would be more like: "Piano lessons will make you musically literate which is a valuable disposition and you have nothing better to do with your Wednesday afternoons and there is nothing better for me to do with my money" or alternatively, "I promised your late grandmother that if I had the means, I would send you to piano lessons; she always cared about giving you opportunities that she did not have." So the question that I ask in this paper is whether the prediction that someone else will be glad we did something is a good reason for that person to believe that we are morally permitted to do it on that person's behalf. If "you'll be glad I did it" reasoning can offer for others a reason to believe that our action is a morally permissible way to treat them, then it can play a legitimate justificatory role in our vicarious decision-making. Here is Deafness Argument in its vicarious form:

Vicarious Deafness Argument:

(1) If I do not cure my baby of deafness, he'll be glad I made that choice.

(2) Therefore, it is okay to not cure my baby of deafness.

Given that Harman is investigating how our future attitudes can inform us about what at present we ought to do all things considered, she examines when it is the case that these future attitudes are indicative that the decision will lead to things turning out for the best (in all the ways we should care about). On the other hand, I am looking at how future attitudes can play a role in determining the moral permissibility of certain decisions on behalf of others, and future attitudes can play this justificatory role even if we don't think they offer any such indication. When we start thinking about how to

justify the things we do to others as morally permissible, it is easier to see how some justifications are legitimate even when we cannot demonstrate to the other person that the action will result in the best outcome for them.

So while "I'll be glad I did it" reasoning and "You'll be glad I did it" reasoning play quite different roles in our practical thinking, I do think that it is useful to compare these two types of reasoning to determine whether we can learn anything about the limits of using future preferences of another to justify our actions on the other's behalf. For instance, thinking about these two different types of reasoning leads Harman and me to different conclusions about how the prospect of transformative experiences should figure into our practical thinking. It follows from Harman's view that if we know that a future attitude will result from a transformative experience, then that attitude can no longer play its typical role in recommending certain actions. This is not the case on my account. I will argue that future preferences *can* play their typical role in justifying present actions as morally permissible even if they are likely to be the result of a transformative experience. This is because the typical role that I think our future attitudes play in justification is more complicated than how it is described in Predictive Glad/"I'll be glad I did it" reasoning. My approach thus offers a way to think about how our future preferences can inform our practical thinking in many situations for which Harman's account falls silent. Before presenting the details of my view, however, I want to explore what it is about Predictive Glad that appeals to us when we must decide for others (especially, for our children); and also what it is about such reasoning that should cause some suspicion. I turn to these questions next.

2 The Promise and Peril of Future-Oriented Consent

When thinking about deciding on behalf of children, some have defended versions of "You'll be glad I did it" reasoning. Most notably, Gerald Dworkin provides a clear articulation of such a method in thinking:

> There is . . . an important moral limitation on the exercise of such parental power that is provided by the notion of children eventually coming to see the correctness of the parent's intervention. Parental paternalism may be thought of as a wager by parents on children's subsequent recognition of the wisdom of the restrictions. There is an emphasis on what could be called future-oriented consent—on what children will come to welcome, rather than on what they do welcome. (Dworkin 1983, 28)

In other words, Dworkin argues that a parent's decision in the present is justified if it is reasonable to believe that the child will one day come to accept it. This general method of justification can be formalized as follows:

> **Predictive Glad**: If I can predict that you will be glad I φ-ed in the future, then my φ-ing now is a morally permissible way to treat you.[11]

Now clearly, as it stands Predictive Glad is unacceptable as a justifying principle. This is because it ignores the possibility of defective future pro-attitudes. We may be able to predict that the other person's future gladness will be based on misleading evidence, or due to a paucity of evidence, or will be blatantly irrational. If the vicarious decision-maker is in a situation to predict any one of these defects, then the future pro-attitudes cannot be a legitimate source of justification. So we should amend Predictive Glad in the following manner:

> **Full Predictive Glad**: If I can predict that you will—reasonably and with adequate knowledge about your situation—be glad I φ-ed in the future, then my φ-ing now is a morally permissible way to treat you.[12]

Full Predictive Glad seems to be the underlying principle that, if true, would make "You'll be glad I did it" reasoning an instance of good reasoning. It is a general formulation of a class of arguments that appeal to some future pro-attitude to justify a current measure; this pro-attitude could be future consent, future acceptance, future endorsement, future satisfaction, or some other future retrospective preference. So as Harman mentions, gladness need not be taken to be a mere emotional state, rather the attitude of being 'glad' is best understood as a sort of placeholder for whichever of these future pro-attitudes one may think is relevant. Different pro-attitudes may make more or less stringent limitations on what sorts of vicarious decisions are morally permissible. Dworkin argues that the relevant future attitude that can justify parental decisions is consent but there is reason to worry about whether consent can be retrospective (Husak 2010, 114).[13] I will therefore use endorsement rather than consent as the target future pro-attitude. The idea of future endorsement adheres to the spirit of Dworkin's claim that the child should one day come to see the "wisdom" of the earlier decision.

Endorsement is stronger than mere acceptance. One can accept what one takes to be an unjust state of affairs if one can imagine no better alternative or if the costs of rejecting it are too high. When one endorses a previous decision, on the other hand, one not only finds the consequences of the decision to be tolerable but also sees that decision itself as justified. Imagine a parent who decides to risk his child's modest college fund in order to try his luck at slots. The risk pays off and the parent now has quadrupled the

[11] The formulation of Predictive Glad is influenced by Harman's principle "Reflection for Desires" (2009).

[12] For clarity, in the rest of the paper, I will use "reasonably" to connote "reasonably and with adequate knowledge."

[13] For a defense of the conceptual coherence of subsequent consent, see Chwang 2009.

fund. The child may later come to accept the windfall without necessarily endorsing the parent's decision. He may be happy about the outcome of the decision without seeing the decision itself as having been justified. Filling in Full Predictive Glad inspired by Dworkin's account would look like this:

> **Future Endorsement:** Though you may not or cannot endorse my φ-ing on your behalf now, if I can predict that, in the future, you will reasonably endorse my decision to φ, then my φ-ing on your behalf now is a morally permissible way to treat you.

There are reasons to be optimistic about Future Endorsement as a means to justify one's decisions on behalf of others who do not have the capacity to make decisions on their own behalf. First, as already noted, it appeals to our commonplace practices. Parents sign their children up for piano lessons with an eye toward the child's future appreciation of his own musical ability. They introduce unfamiliar foods into their children's diets with the hopes that someday pad thai will be as appetizing as pizza.

Second, the idea of Future Endorsement requires the parent to consider the child's own attitudes (albeit future ones) in deciding what one may do. Rather than justify her action by demonstrating some objective benefit that the child stands to gain, the parent must regulate her actions with an eye toward the child's actual preferences and attitudes. This feature is especially important when it is controversial what would constitute the best interest of the child. The child's interests, projects, and values may just be taking shape; and there may be no general consensus about which path of action is of most value. While people generally agree that braces are a worthwhile burden for children to bear, there can be reasonable disagreement whether home school or public school would be a justifiable form of education for a particular child. Future Endorsement seems to sidestep some of these issues since it appeals to the future subjective states of the child rather than some objective standard.

Third, relatedly, the child's future endorsement can be a powerful target when it comes to parents making decisions that have the potential to be transformative of the child's values, interests, and perspective of the world. On the one hand, the parents may themselves be on the other side of a kind of transformative experience that they are now thinking about choosing for their child. They may know the joy of musical literacy or the delight of papaya salad with the perfect amount of fish sauce, but such considerations may not yet be salient to their child without the proper training or habituation. Defending their decision by saying "you'll be glad I did it" needn't be flippant in such situations. Instead it can be a compelling justification to a child who does not yet have the experiences necessary to fully endorse her parents' decision on its own merit but who has reason to trust her parents' instincts on the sorts of things she will come to value.

On the other hand, parents may be required to make a decision on behalf of a child that involves phenomenological experiences that are foreign to them and for which they have no direct access. Hearing parents must make decisions on behalf of deaf children, parents who are cisgender must make decisions on behalf of trans* children, parents must make decisions on behalf of children of a different race. In such situations, parents should be open to the possibility that they do not know what it is like or what it will be like for their child to deal with the consequences of their vicarious decisions. However, they may come to accurately predict their child's future preferences by taking seriously the testimony of others who have been through the relevant phenomenological experiences.[14] In such cases, parents may claim, "Although I do not fully understand why in the future you will endorse the decision I am making now, I can reasonably predict that you will, and so I am justified in making it."

Fourth, the child's future endorsement is the right sort of pro-attitude to appeal to when thinking about which vicarious decisions can be justified rather than which actions are the best actions to take. Parents have endless choices regarding how to raise their children, but these vicarious decisions must be made alongside other important decisions they face. While it is reasonable for parents to want to do what is best for their child, it is also reasonable for them to want to pursue a rewarding career path, to want to maintain healthy relationships with other adults, or to want to live up to their civic duty. Some of these other aims that parents are reasonable in pursuing may lead them to act in ways that they recognize may *not* be the best for their child. For instance, after much research, a deaf single mother may decide that while her deaf son may benefit most from receiving cochlear implants, this is not the best decision *for her*. Medical experts suggest that a child with cochlear implants should be raised using oral communication at home and in school (Ouellette 2009, 1248). Since the deaf mother may have both cultural and economic reasons to raise her child at home using only sign language, it at least appears reasonable to raise her son without the implants. The idea of securing future endorsement makes room for the moral permissibility of many actions that parents may choose which would benefit the child but may not necessarily secure the *most* benefit for the child given the alternatives.

So Future Endorsement has a lot going for it. Despite these encouraging features, I will argue that justifying vicarious decisions by appealing to Future Endorsement is insufficient. This is because Full Predictive Glad, taken on its own, is an incomplete justification, regardless of the future pro-attitude one wishes to insert in the schema. The tricky part about making decisions for children is that guardians may act not only in ways that they

[14] See Sharadin 2015 and Dougherty et al. 2015 about how testimony of others can be informative in first personal practical decision making. See McKinnon 2015 for a powerful illustration of how one can effectively know the probable disutility associated with forgoing gender transitioning, even if one recognizes that there is great variation in trans* experience.

think are in the best interest of the child, but the vicarious decisions that they end up making also shape what the child himself takes to be in his best interest in the long run. Therefore, if a child's future attitudes are determined by certain decisions we take at present, we should be wary about justifying these decisions solely by invoking the future pro-attitudes.

This skepticism is bolstered by John Rawls's assertion that the fact of future acceptability is not enough for the legitimacy of paternalistic power. Rawls considers the case of an involuntary conversion (1999, 220). He asks us to imagine a psychiatrist who is deciding whether to administer some treatment, such as shock therapy, that will cause the patient to abandon a presently-held philosophical belief for a different one in the future. The fact that the patient may one day conscientiously endorse both her new belief as well as the course of treatment seems irrelevant, says Rawls, to whether or not the psychiatrist's intervention is justified.

Rawls's argument relies on the view that the sole determining factor for the patient's pro-attitude toward the new belief is that she underwent the treatment. The event itself is the cause the subsequent pro-attitude about that event. This can lead to what seems like a bad form of bootstrapping. Were we to ask the psychiatrist, "Why is it permissible for you administer the treatment?" he might respond, "If I administer the treatment, she'll be glad I did it." Our natural follow up, "But *why* will she be glad?" It is unsatisfying for the psychiatrist to respond: "Because I administered the treatment." In an attempt to justify the moral permissibility of the treatment, the psychiatrist is appealing to the predicted future pro-attitudes of his patient and when pushed on what would justify these future pro-attitudes, the psychiatrist is giving a *causal* story about how these attitudes would come about. But causal stories are not justifying. Administering the treatment is not a morally permissible way to treat another person just in virtue of the fact that in the future the treatment will have been administered. Similarly, one cannot justify an event by direct appeal to a pro-attitude that is exclusively, causally dependent on that event having taken place.

However Rawls's rejection moves too quickly. It all depends on *how* the treatment causes the patient to conscientiously accept the new belief and also the treatment. If the acceptance itself has just been implanted, then yes, the future acceptance doesn't add much in terms of justification. But the patient may have other reasons for being glad that she underwent the treatment. Conscientious acceptance may come from the patient now holding certain views and values that have developed over time as a result of the treatment by which she now endorses it. Certainly the treatment caused her to have these values. But her reasons for acceptance are grounded in the values that she now conscientiously holds and for which she can offer independent justification. And after all, *all* of our held values have some causal story. Our life circumstances, our formative relationships, our bodily capacities—these are the causes of the values we hold. What is so special

about a subjectively held value being causally traced to some treatment chosen by another person on one's behalf? Of course one feature that seems distinctive about this hypothetical case that Rawls puts forth is that the treatment is involuntary. However this is not a relevant factor in parental decision-making. Parents manipulate the circumstances of their children in an involuntary way that influences their interests and values all the time. It is uncontroversial that education can and should influence the interests and values of children. So even involuntary procedures can and should be justified.

Rawls is right to highlight what it is about justifications based on future preferences that should make us wary, but he is wrong to reject the entire structure. It seems that in some cases, a parent's vicarious decisions are perfectly justifiable *even when* the child's future pro-attitude is causally dependent on those decisions. Recall piano lessons, dreaded at the time but remembered fondly in the future. Dworkin is right that these actions taken by parents are indeed wagers that the child will one day see the wisdom of their efforts. While some wagers about the child's future preferences may be optimistic, they do not necessarily manifest bad reasoning. So the lesson we should draw from Dworkin is that the child's future attitude can play some role in justifying our vicarious decisions; the lesson we should draw from Rawls is that predicting that the child will endorse the decision in the future cannot on its own justify that decision. We thus need to investigate further the justificatory structure of Full Predictive Glad to see when such wagers are acceptable ways of making decisions on behalf of a child and when they are problematic.

3 Justification and Optimality

Harman has examined closely the structure of arguments that employ something akin to Full Predictive Glad in its first-personal form. So it will be useful to see where she draws the line between acceptable and problematic cases of "I'll be glad I did it" reasoning and why. Doing so will help to clarify where I think Full Predictive Glad goes wrong as a justification for the moral permissibility of an action.

Although Harman does not think that "I'll be glad I did it" reasoning is deductively sound, she does take it to be generally good reasoning. Such reasoning may not always provide conclusive reasons to do something, but it does generate some good reasons to think that I should do it. And so it is the sort of commonsense reasoning we should continue employing in our everyday practical thinking. Harman writes that, "*typically*, the fact that I will be glad I did it is genuinely indicative that I should do the thing in question" (2009, 194). This, according to Harman, is because *typically* the fact that we will be glad we φ-ed is genuinely indicative of that fact φ-ing would be best (in that it would bring about the best state of affairs in all the ways we should care about). And furthermore, we often should

do what would be best in this way (2009, 188). Harman then offers some defeaters that identify when the situation is no longer typical and when our future preferences cannot justify our present ones. She writes, "if there are facts that would defeat this reason, and the facts are sufficiently salient to an agent, then 'I'll be glad I did it' reasoning is bad reasoning" (2009, 194). Here is a partial list of defeaters:

"I'll be glad I did it" is bad reasoning if I believe that . . .

(1) "I'll be glad I did it" will be unreasonable, or
(2) "I'll be glad I did it" will be due to misleading evidence, or
(3) "I'll be glad I did it" will arise out of love for and attachment to someone, or
(4) "I'll be glad I did it" will arise out of my inability to identify with the person I would have been in the alternative state of affairs.

So what, according to Harman, makes Deafness Argument an instance of bad reasoning when the structurally similar Paper Argument manifests perfectly good reasoning? What defeater should be sufficiently salient to Stevie's mother?

First, we should be clear on what Harman thinks is *not* a defeater in this case. She does not think the problem with Deafness Argument is that the mother's future preferences will be unreasonable. Consider how that view would go: "Since it is worse to be deaf than to be hearing, it follows that it is unreasonable to prefer being deaf over being hearing. Furthermore, when it comes to our preferences concerning our loved ones, it is unreasonable to prefer their being deaf over their being hearing. So it should be salient to the mother that her predicted future preferences will be unreasonable (Defeater 1)." Harman rightly rejects this assessment of what goes wrong in Deafness Argument. While she maintains that it is worse to be deaf than it is to be hearing, she does not think that this licenses any conclusions about what is unreasonable for Stevie or his mother to end up preferring. As she compellingly argues, it is reasonable for our future preferences to be sensitive to how we and how the people we love will have actually turned out. We can call these preferences for how people turn out "Person-Affecting Preferences."[15] Deafness Argument involves Person-Affecting Preferences for how Stevie turned out rather than inherently unreasonable ones.

It is this involvement of Person-Affecting Preferences that is the salient feature that makes Deafness Argument problematic on Harman's account. Although Harman believes that our predicted Person-Affecting Preferences can be reasonable in the future, she doesn't think that they can give us reason to believe the we should do one action or another at present. Harman reminds us that just because "a preference is reasonable given that a person has a certain character [this] does not imply that the preference

[15] Both defeater 3 and defeater 4 on the list describe a Person-Affecting Preference. This title is adopted from Barnes 2009a.

is reasonable before the person has come to have this character" (2009, 191). Once Stevie's character has been shaped by his deafness, it can be reasonable for him and for his mother to be glad that he is deaf. His deafness has become a significant and invaluable part of his life and his identity. But at infancy, his deafness is not yet a significant and invaluable part of his life. It is just a physical condition. Our attitudes about a single event can change over time, so we can reasonably come to be glad that some event took place even though it would be unreasonable to prefer it coming about or to make an effort to bring it about (Harman 2009, 188).[16]

All this leads Harman to what she takes to be the basic problem with the reasoning employed in Deafness Argument: often, we should do what brings about the best outcomes and, *typically*, our future preferences are indications of what will turn out best. But sometimes they fail to be indications of this. This is one such occasion. It should be salient to the mother that her future preferences will grow out of her love for her son and how he will have actually turned out (Defeater 3). Such Person-Affecting Preferences track how people actually turn out rather than what would have been best. So the mother's future gladness about her decision in Deafness Argument is not indicative of the fact that the outcome will be best in all the ways that she cares about. In cases like Deafness, the mother's future preferences cannot offer her a reason to believe that she should forgo the cochlear implants.

I don't want to dispute the important lessons that Harman draws about how our future preferences can be reasonable even when things don't turn out for the best. But her assessment of Deafness Argument is problematic for our purposes in two other important ways. First, Harman thinks that whenever they aren't Person-Affecting Preferences, our future preferences can be straightforwardly indicative of optimality (i.e., they can be indicative that some option will bring about the best outcome in all the ways we should care about). In contrast, I think that when taken on their own, *all* predicted preferences, not just the person-affecting ones, are insufficient indications of optimality. Even in Paper Argument, one has to compare alternative courses of action and the resultant preferences to discover that working on the paper is the optimal choice. The second problematic feature results from the fact that Harman and I are investigating different types of practical reasoning and so the role that future preferences ought to play in each type of reasoning differs as well. Harman asks whether we can employ "I'll be glad I did it" reasoning to figure out what she should do all things considered. She argues that the only way our future preferences can be informative for this purpose is if they are indicative of what will be best. I on the other hand, care about the moral permissibility of a vicarious action and not necessarily which action one ought to do all things considered. Pointing out the optimality of some course of action is not the only way

[16] See also Heathwood 2008 and Hare 2011.

to justify it as morally permissible. Given that there is a diversity of ways to justify our actions, future preferences can play a role in some of these justifications, even if we predict that they will be person-affecting. I will develop both of these arguments in turn.

First, let us examine why Person-Affecting Preferences are not indicative of optimality. Once a personally transformative experience has happened and has shaped you, you may come to reasonably endorse that experience; but if the personally transformative experience did not happen and its not-happening shaped you, you may reasonably have endorsed *not* experiencing it. Either way, your future Person-Affecting Preference for how you turned out is not indicative that you turned out for the best. But even in the case of normal non-person-affecting preferences, our future gladness is not always a mark of optimality. The intuitive nature of Paper Argument relies on the presumed disparity between the goodness of the two options—built into one's preference to continue working on the paper is the implicit knowledge that one would have regretted going to the movie instead. However, consider if writing the paper is pitted against some other good option that is harder to compare: you can either continue working on the paper or you can catch up with an old friend. Neither of these options is going to be transformative or otherwise person-affecting, and yet the fact that you will be glad in either case does not indicate that it is the option that will be the optimal one. You choose to work on the paper and you will be glad you did it—your work will have progressed and your friend isn't going anywhere. You choose to catch up with an old friend and you will be glad you did it—it is always interesting to hear what she's up to and the conversation will feel like a well-deserved a break from work.[17] So regardless of whether our future preferences are Person-Affecting or not, we still have to pay attention to how our predicted preference compares to the future preferences that will result from other possible actions to know whether it is genuinely indicative of the optimal choice.

Second, insofar as we are in the business of justifying actions as morally permissible rather than figuring out what we should do all things considered, then our future preferences may play certain justificatory roles even when they are not indicative that things will have turned out best. Justifications for an action needn't explicitly demonstrate how the action is for the best; all they need to do is offer the other person (or the agent herself) a reason to believe that the agent is morally permitted to φ. Figuring out whether a vicarious action is justifiable to another and figuring out which vicarious

[17] I should note that I am committed to the view defended by Dietrich and List (2011) that our preferences can shift without us necessarily learning any new information about our situation. This is because experiences can shift what sorts of considerations are motivationally salient for a particular person rather than provide for the person new information. Working on the paper or spending time with a friend can have this affect on what you will find to be a motivationally salient consideration. I'm disinclined to call these shifts personally transformative experiences, though I recognize that they alter our priorities and self-conception slightly.

action will bring about the best outcome in all the ways that the other person should care about are conceptually distinct mental activities. Of course, when we have authority to make decisions on behalf of another, one important way to justify the morally permissibility of the particular decision we are thinking of making is by demonstrating how that decision to φ would be best for that person. But sometimes we cannot make a good prediction about what would be best in advance and yet we still need to figure out which actions on behalf of this person are permissible ways to treat that person. On such occasions, we may try to determine whether all the alternatives to φ-ing are impermissible courses of action (even as we recognize that the deontic status of φ-ing is still up for dispute).[18] Other times, we may have a view about what would be best in all the ways that the person should care about but recognize that *that person* may view things quite differently. When we find ourselves in such a situation, the morally permissible thing to do may be to track that person's actual cares and interests rather than the ones we think the person should have. If this is the morally permissible thing to do in such a situation, then the vicarious decision cannot be justified by appealing to our belief that the decision will best in all the ways that *the other person should* care about.

These are just a few ways in which justifications about the moral permissibility of an action do not require appealing to what would be best for the other person in all the ways she should care about. Given that there is a diverse range of methods of justification, Person-Affecting Preferences may have some role to play in justifying our decisions even when they are not indicative that the decision would be for the best. Moreover, if the method of justification we are employing is appealing to optimality, then there should be nothing especially misleading about our predicted future preferences which are person-affecting. The typical way that our future preferences are genuinely indicative of the fact that some course of action is optimal is if we consider the predicted future attitudes of doing that action along with (at least) the hypothetical attitudes of not doing it.

Given these two arguments about optimality and justification, I am led to different conclusions than Harman. Whereas she thinks that "I'll be glad I did it" reasoning is typically good but defeasible reasoning, I think that, taken on its own, the fact that one will be glad one did it *never* offers up a conclusive reason to believe that one should do it. We should always be suspicious of such a justification if we cannot find a further feature of the situation that supports it. The mere fact that we (or someone we love) will be glad we did something is not a defeasible reason to believe we should do it, nor for that matter is it a reason to believe that we are permitted to do it. Such reasons too easily lead us astray or are employed for pernicious ends. This is especially clear when we look at the case of

[18] For discussion about the normative implications of making decisions in situations of moral ignorance and moral uncertainty, see Harman 2011 and Guerrero 2007.

justifying our vicarious actions. Rather than think that "You'll be glad I did it" is generally good but defeasible justification, and put the burden on the vicarious decision-maker to find some salient feature that would defeat it, our practical reasoning will go better if we regard "you'll be glad I did it" reasoning as dubious when taken on its own. The burden is then on the decision-maker to argue why 'you'll be glad I did it' is evidence of some further feature that would indeed make the reasonableness of the future preference justify some action in the present as morally permissible. I think that we often *can* meet this burden even if it is likely that the future preference results from a transformative experience or is otherwise person-affecting. I will turn to some ways in which considering our future preferences can play a role in justifying our action in the final two sections of the paper.

4 Person-Affecting Preferences versus Adaptive Preferences

Harman says that "I'll be glad I did it" reasoning is typically good but defeasible reasoning. But once we start paying attention to the list of defeaters that she compiles, such reasoning ends up being unsuitable for the many predicaments in which we thought it could be of distinctive help in figuring out what we should do. In particular, if Harman is correct about the fact that Person-Affecting Preferences in the future cannot play a role in informing us about whether we ought to perform a present action, then it seems like "I'll be glad I did it" and "You'll be glad I did it" style reasoning lose a lot of their intuitive deliberative power. Insofar as some decision has the capacity to transform the child in ways that will resonate with her identity and affect her deeply held values, it can no longer be justified by "You'll be glad I did it" reasoning based on Harman's account. So all sorts of parental decisions we thought could be justified by appeal to the child's future preferences can no longer be justified in this manner: piano lessons, camping, throwing away a kid's tattered security blanket when they have outgrown it, etc. "I'll be glad I did it" may turn out to be informative for working on a paper and "You'll be glad I did it" may justify getting your kid to go to bed, and other sorts of mundane things that are obviously good courses of action, but that's about it.

This deflationary conception of the uses of "I'll be glad I did it" reasoning may be a salutary upshot to Harman's view since part of what she is doing is trying to uncover some common mistakes people make when they think that their current preferences for how things actually turned out should universalize to what other people's preferences should have been prior to things turning out their preferred way. But as I articulated in section 2, I think that predicting our children's future preferences can be of immense use for our vicarious decision-making, particularly for clarifying our thinking in the face of reasonable disagreement about what would be best for the child. So rather than dismissing people's future attitudes in all these different

situations, I think we should dismiss Predictive Glad as the principle that should guide us in employing such reasoning. The problem is not Person-Affecting Preferences; the problem is appealing to the predicted preferences of only one course of action.

In the next section, I will present an alternative approach to Predictive Glad. I think that my approach allows a justificatory role for people's future attitudes in a way that guards against some of the problems that Harman discusses while at the same time retaining some of its intuitive uses. But more importantly, my approach will reveal how our future attitudes can be informative in cases where previously they seemed to be misleading at best. To see this, I want to take some time to discuss the differences I take there to be between Vicarious Deafness Argument and Happy Old Clown Argument.[19] Recall how these two arguments proceed:

Vicarious Deafness Argument:

(1) If I do not cure my baby of deafness, he'll be glad I made that choice.

(2) Therefore, it is okay to not cure my baby of deafness.

(Modified) Happy Old Clown Argument:

(1) If I break my son's legs, he will be glad I did it.

(2) Therefore, it is okay to break my son's legs.

Given the defeaters that Harman provides, how would we assess these two cases of practical reasoning? In regard to Vicarious Deafness Argument, we can claim that the mother should be able to predict that Stevie's future preference are potentially person-affecting in a problematic way. Namely, his preferences for remaining cochlear-implant-free may arise out of his inability to identify with the person he would have become had he undergone the surgery (Defeater 4). Notice that deciding to go forward with cochlear implants based on Stevie's future pro-attitudes would also fall prey to Defeater 4. Were Stevie to get cochlear implants, he would be glad that he got them, but these future preferences may also arise out of an inability to identify with the person he would have become had he remained fully deaf. It follows that Vicarious Deafness Argument as well as a vicarious form of Cochlear Implant Argument would be cases of bad reasoning. Similarly, the Fratillini father should be in a position to be wary of the possibility that his son's future preference for having his legs broken

[19] A number of differences between the two cases that I do not discuss: (1) The father is physically hurting his son and the mother is not. Although, obviously true and important, I don't think that this is the most interesting difference between the two cases. So if one wishes, one could alter the Happy Old Clown Case so that the father agrees to some medical procedure that will alter his son's body for the sake of comical walking. (2) The Fratellini father is deciding whether to cause his son's impairment and Stevie's mother's is deciding whether to not alter her son's impairment. I have a hard time seeing how this distinction between causing vs. not causing is of normative significance in the case of parents making decisions on behalf of children, though I could be wrong. For an interesting discussion of this distinction see Barnes 2014.

would arise from an inability to identify with the person he would have become had he not been forced to walk comically. So if we are going by the list of defeaters, Happy Old Clown Argument is a case of bad reasoning in exactly the same way as Deafness Argument, they both cannot rule out the possibility of Person-Affecting Preferences.

However, not all Person-Affecting Preferences are on par. There is something about the way that some of our future Person-Affecting Preferences may be formed that make them distinctively unreliable grounds for justification. Whereas I think that Vicarious Deafness Argument and Happy Old Clown Argument do face structurally similar problems in employing Predictive Glad—I do think that we can look directly at the future attitudes underlying Happy Old Clown Argument and dismiss certain courses of action as impermissible in a way that I don't think is possible for Stevie's mother in Deafness Argument. This is because not only does Happy Old Clown Argument rely on Person-Affecting Preferences, it also cannot rule out the possibility of adaptive preferences. When we are in a position to see that some preferences are potentially adaptive and compare those to the preferences that result from the alternative course of action, we should be able to draw stronger conclusions than the conclusion that they don't justify our actions.

There are many different ways of understanding how to determine whether some agent's preferences are problematically adaptive. Some views understand adaptive preferences as those preferences formed in oppressive circumstances (Superson 2005),[20] others view adaptive preferences as those preferences formed in response to diminished options but only if people end up preferring *suboptimal options* (Nussbaum 2001). But in order to have something to challenge the justificatory structure of the Fratellini father's reasoning, I am going to follow Jon Elster's original formulation of adaptive preferences as being unreliable purely because of some formal features about the way that a person comes to have them rather than some normative view about the badness of the circumstances or the badness of the preferences that develop. Doing this allows me to show what is bad about the father's reasoning without making comparative claims about whether it is worse to walk comically than it is to walk plainly. Thus my account can end up presenting reasons to think that it is impermissible to break one's son's legs that can be made salient to people like the old Fratellini Clown who deeply value their clownish gait. If such reasons are sufficiently salient to the father than we can charge him with engaging in bad reasoning rather then merely reasoning from tragically false assumptions.

For Elster, the problem with adaptive preferences is that they are subconsciously formed in response to a person's diminished set of feasible options. As the person's set of options is diminished, that person's preferences change to the point where the person prefers something that is within the feasible

[20] Superson (2005) calls these sorts of adaptive preferences, 'deformed desires.'

set of options rather than preferring some option from the larger set of conceivable alternatives—some of which may no longer be within reach (Elster 1983, 114). When this happens, the person's future preferences become indistinguishable from accepting a suboptimal situation.

Let us consider how the possibility of adaptive preferences would work for the case of Happy Old Clown. While the father may not be in a position to think that his son's future pro-attitudes are unreasonable by looking at their content, he must concede regardless of his values, that the decision to break his son's legs will significantly diminish his son's set of feasible life options. The position that the father could consistently hold in light of this fact is that such a diminishment is justified by the overwhelming value of his son being able to masterfully carry on the Fratellini family tradition of clowning. It is a necessary tradeoff, the Fratellini father could say, between the diminishment of his son's options and the exclusive focus on a path to clowning excellence. If this is the case, the Fratellini father would have to concede that his son's future preferences may be unreliable markers for the moral permissibility of his actions—since they are indistinguishable from accepting a suboptimal situation. Although the son may value his physical impairment for the intrinsic value of clowning, he may also only value it in response to his constrained circumstances. In an effort to cope with his diminished set of options, the Fratellini son may manage to convince himself that he not only accepts his condition but that he does not regret his father's decision to break his legs.[21] Such convincing may help the son get by, and may in some sense be an understandable response to his sub-optimal situation, but it should still be seen as an instance of unreliable adaptive preferences. The following argument should thus be salient to the father given his values and deliberative position:

Possibly Adaptive Happy Old Clown Argument:

(1) If I break my son's legs, he will be glad that I did it (and possibly, reasonably so.)
(2) My son will be glad in spite of a diminished set of feasible life options.
(3) Given my present deliberative position, I have some reason to think that my son's future preferences may result from adaptive preference formation and so I cannot distinguish these predicted preferences from mere acceptance of a sub-optimal state.
(4) However, my son's future acceptance would not justify my present action as a morally permissible way to treat him.
(5) Therefore, my son's future gladness cannot justify my breaking his legs as a morally permissible way to treat him.

[21] Regret here just is the opposite of being glad. It is to prefer a state of affairs in which things would have turned out differently rather than the state of affairs in which things turned out as they did. I thank an anonymous referee for asking to clarify this point.

We should note that the possibility that his son's future attitudes will be adaptive may be salient to the Fratellini father even if he endorses his own broken legs and believes his own gladness to be warranted.

If adaptive preferences are a problem for the Fratellini father, shouldn't they also be a problem for Stevie's mother? One could argue along similar lines as Possibly Adaptive Happy Old Clown that growing up deaf in today's society may constrain Stevie's options to such a significant extent that it may affect the reliability of his attitudes about his condition. In an effort to cope with his situation, he may manage to convince himself that he not only accepts his deafness but that he does not regret his mother's decision to forgo the surgery. Such convincing may help Stevie get by, and may in some sense be an understandable response to his sub-optimal situation, but it should still be seen as an instance of unreliable adaptive preferences. This is the story one could tell in order to argue that Stevie's future attitudes may be the result of adaptive preference formation.

But I don't think that the mother has to concede this story in the same way that the Fratellini father should concede Possibly Adaptive Happy Old Clown Argument.[22] Stevie's mother and the Fratellini father are at present in different deliberative positions. The mother can reasonably maintain that forgoing cochlear implants and instead learning American Sign Language as one's first language does not diminish one's feasibility set. Disability rights advocates and philosophers have forwarded the view that at least some disabilities are mere differences from standard physicality rather than inherently sub-optimal (Aas Unpublished; Barnes 2009b; Silvers and Francis 2005; Thomson 1996). While it is undoubtedly the case that the way the world is set up, being deaf presents one with certain hardships that being hearing does not, these hardships don't necessarily diminish one's feasibility set, though they do make some goals more challenging to reach. Moreover, when the hardships that are presented to a deaf person become diminishments of that person's feasibility set, this is not necessarily a result of their impairment but rather the result of the way their impairment is accommodated by society and the way that others, including their loved ones, relate to them in light of their deafness.[23] So while the mother must accept that being deaf in this society presents one with certain hardships, she may have practical and political reasons not to accept the view that just because one is presented with some hardship, this inherently represents a diminishment of one's feasibility set. Accepting this assumption means accepting unjust conditions of society as fixed features of the condition of the disability. In particular, her personal acceptance of such a view could limit Stevie's life prospects in unwarranted ways. Notice that this is different from the Fratellini father case. The whole point of breaking his

[22] For a more comprehensive argument against regarding the preferences of people who are disabled as adaptive, see Barnes 2009a.

[23] For a historical example of a society in which deafness was not a disability look at Martha's Vineyard from the seventeenth century to the early twentieth century, see Groce 1985.

son's legs is to diminish the feasibility set in exchange for what the father presumes is a worthy outcome. The Fratellini boy is to have his legs broken to become a clown, not to become whatever his heart desires. The mother on the other hand can be committed to the view that little Stevie can remain profoundly deaf and still pursue the same variety of worthwhile life plans as someone with cochlear implants.

This is obviously too simple a gloss and the view that disability rights advocates defend is much richer and more nuanced. But this gloss highlights (and perhaps exaggerates) a fairly weak claim that I wish to defend here. Namely, that the mother would be reasonable in refusing to see forgoing cochlear implants as diminishing Stevie's set of life options in any significant way and in refusing to view the genuinely held gladness of her deaf friends as potentially adaptive. The mother has reason to believe that if she decides to forgo the surgery, her son will be reasonably glad for the decision and this gladness should not be mistaken as a mere coping mechanism. So the following augmented argument should still hold:

Reasonable Deafness Argument:

(1) If I do not cure my baby of deafness, he'll be glad I made that choice and reasonably so.
(2) Given my present deliberative position, I have no specific reason to think that his future gladness will be the result of adaptive preference formation.
(3) Therefore, my baby's future gladness offers me a reason to believe that not curing him of deafness is a morally permissible way to treat him.

From looking more carefully at their sons' predicted future attitudes and the circumstances in which those attitudes have been formed, Stevie's mother and the Fratellini father should reach different practical conclusions. These conclusions are fairly minimal, however. The father learns that his son's predicted attitude cannot justify breaking his legs and the mother learns that her son's predicted attitude offers a reason to believe that forgoing the implants is morally permissible. But is this reason decisive? In the next section, I will show how once we compare different options that are available to these parents they may be led to more decisive conclusions about which decisions are morally permissible.

5 Entertaining Competing Options

When we are charged with making decisions on behalf of others, we cannot justify our decisions as morally permissible solely based on the prediction that the others will glad we made that choice. I have argued that, when taken on their own, people's future pro-attitudes about some action can never justify that action at present. This does not mean that people's future pro-attitudes have no role to play in our thinking. Nor does it mean that we

should always be suspicious of future attitudes if they come about because of a transformative experience. As we have seen already in the case of Happy Old Clown, thinking about the future attitudes of the child *can* inform our decisions about how to act on their behalf, at least in a negative way. Let us look at more schematized version of Possibly Adaptive Happy Old Clown:

Schematized Possibly Adaptive Happy Old Clown Argument:

(1) If I break my son's legs, he will be glad I did it. But I have reason to believe that this gladness may be the product of adaptive preference formation.

(2) Therefore, my son's future gladness does not justify my breaking his legs as a morally permissible way to treat him.

One thing to notice about this argument is that it does not yet justify any action on the part of the father. The practical conclusion that this argument offers is a negative one: the father is *not* justified in breaking his son's legs by appeal to his son's future attitudes. In appealing to the future attitudes of the son, can we learn anything about what the father would be justified in doing?

When we are trying to figure out how to act, either on our own behalf or on behalf of another, we are often faced with two conflicting courses of action that can be taken. So if it turns out that one's future pro-attitude does not justify one course of action, it seems reasonable to assume that *the negation* of the action would be justified. But this does not follow. This is because the opposite action may not be justified *by* one's consequent future attitudes either.

Consider Milo at the beginning of the children's book *The Phantom Tollbooth*. He is a sad specimen of a young man, "when he was in school he longed to be out and when he was out he longed to be in" (Juster 1961, 3). Milo is presented with two tedious seeming options: he can either go to school or not go to school. When faced with such a decision, he pictures each action he can take and then imagines how he would feel about the outcome of that action. If he pictures himself going to school, then he can predict his attitude would be to wish he had stayed home; and if he pictures himself staying home, then he can predict his attitude would be to wish he had gone to school. Either way, his predicted attitudes do not justify the opposite action. While there is something indeed sad about Milo's state, it doesn't seem to be characterized by practical irrationality. It may just be the case that Milo is saddled with two bad options. Each predicted attitude could be a reasonable response to the choice that Milo would make. So his predicted future attitudes justify neither going to school nor staying home. There may, of course, be *other* sorts of justifications such as going to school is good for Milo or it gets Milo out of his parent's hair for a few hours. But considerations about his future attitudes end up being silent on what Milo should do.

However, the old Fratellini father is not in the same camp as Milo. While his son's predicted pro-attitudes cannot justify breaking his legs, we have yet to explore what attitude his son would have were his father to refrain from breaking his legs. Let us imagine that in deliberating about what to do, the father pictures *both* possibilities of action and tries to determine his son's consequent attitude in response to either path:

Expanded Happy Old Clown Argument:

(1) If I break my son's legs, he will be glad I did it. But I have reason to believe that this gladness may be the product of adaptive preference formation.
(2) If I don't break my son's legs, he'll be glad I didn't do it and reasonably so.
(3) My son's predicted future attitudes cannot justify breaking his legs but they can justify refraining from doing so.
(4) Therefore, considering my son's future attitudes, deciding to refrain from breaking his legs is a morally permissible way to treat him.

This finally looks like the kind of practical reasoning that can offer up a justification for why the father should refrain from breaking his son's legs. The father thinks about both paths he could take and imagines whether his son would approve of each path. Whereas Milo's deliberation about the possible paths he can take turns out to be unsettled, the Fratellini father deliberation can—perhaps unsurprisingly—lead to persuasive results. Notice that the father is not comparing the son's future attitudes against each other and seeing whether the son would be *more glad* about one course of action or another. This strategy would be problematic because the father's actions transform what the values and aims of the son turn out to be and hence what attitudes he would come to hold. Rather, what the father is testing out is whether the son would approve or disapprove of each particular course of action and *if* the son approves, whether that attitude could reliably be predicted to be reasonable and not maladaptive. *Ceteris paribus*, if the father is faced with a choice between two conflicting courses of action and as a result of one of the choices the son's pro-attitudes can be viewed as reasonable and as a result of the other course of action the son's pro-attitudes cannot be assured to be reasonable, then the father is justified in pursuing the course of action that would lead to his son's *reasonable* pro-attitudes.

Although unwieldy, this way of appealing to our future pro-attitudes seems like the appropriate way to proceed in our practical deliberations. We cannot simply appeal to the future pro-attitude of the principal to justify some specific action. Moreover, we cannot appeal to the reasonableness of the principal's future pro-attitude alone to justify that action. Instead, we must determine the predicted future pro-attitudes of both courses of action to see if there are any lessons we can draw. This holistic method could even be of use in first personal deliberation cases such as Paper Argument:

Expanded Paper Argument:

(1) If I work on my paper, I'll be glad I did it and reasonably so.
(2) If I don't work on my paper, I will *not* be glad that I didn't do it and reasonably so.
(3) Therefore, I should work on my paper.

Expanded Paper Argument demonstrates that my action can be justified by appeal to my future pro-attitudes if I can claim that (a) I would be glad if I worked on the paper; (b) I would be not glad if I didn't work on the paper; and (c) both these future preferences are reasonable responses to the competing possibilities of action. Again, the argument does not rely on the view that I would be more glad if I worked on the paper than if I put it off. While it may be true that in one situation I would be more glad than the other, the important point of comparison is what course of action I would be glad about and what course of action I would regret. *Ceteris paribus*, if I would regret the course of action and would be glad about the opposite course of action then I am justified in doing what I would not regret.

It is possible then for future attitudes to justify one's current actions, but not in the way that was suggested by "I'll be glad I did it" reasoning. Given these two expanded arguments, we can see how Predictive Glad is an inadequate schema for the purposes of justifying our vicarious decisions. Predictive Glad only focused on *one* possible line of action and determines what attitudes would be reasonable in response to that line. So in its place we may want to offer the following justificatory schema:

> **Predictive Glad/Conjectured Regret:** For any person, *A*, making a decision on behalf of person, *B*, the reasonable prediction that *B* will, reasonably be glad *A* φ-ed along with the reasonable conjecture that *B* would reasonably have regretted *A* not φ-ing, justifies *A*'s φ-ing now.[24]

In the case of vicarious decision-making, *A* and *B* represent two different people; in the case of first personal decision-making, *A* and *B* represent the same person.[25] Where does Predictive Glad/Conjectured Regret leave us with Deafness Argument? Let me expand on the competing courses of action put before the mother:

[24] Harman (2009, 192) considers and ultimately rejects a principle that is very similar to entertaining competing options. She argues that just as "reasonable attachments" (i.e., person-affecting preferences) act as defeaters, so too should "reasonable aversions" act as defeaters. Since I don't think that there is anything particularly problematic with person-affecting preferences justifying our actions when understood in the right way, I also don't think that there would be anything problematic about reasonable aversions. I thank an anonymous referee for pushing me on this point.

[25] Predictive Glad/Conjectured Regret is suitable for Paper Argument but not for Expanded Old Clown Argument since the father doesn't predict that his son will regret either choice. In both cases, what is important is that in entertaining competing options and their consequent attitudes the agents are able to come to some persuasive practical conclusions about which courses of actions are justified.

Expanded Deafness Argument:

(1) If I forgo cochlear implants on Stevie's behalf, he'll be glad I made that choice and reasonably so.
(2) If I agree to cochlear implants on Stevie's behalf, he'll be glad I made *that* choice and reasonably so.
(3) Predicting reasonable gladness does not help me to adjudicate between these two options as morally permissible.
(4) Stevie's future pro-attitudes are insufficient to justify either action in this case as morally permissible.

As opposed to the Expanded Paper Argument and Expanded Happy Clown Argument, we are left with a negative conclusion when we expand the Deafness Argument. Like Milo, neither course of action is fully justified *if we appeal solely* to the principal's pro-attitude. So in some situations—I think in many situations actually—when we compare the different possible lines of action, we are left with inconclusive results. However, this result should not be surprising. After all, as we saw with Milo, our predicted attitudes are responding to the different courses of action that we may take. These courses of action are themselves mutually exclusive, so the manner in which it would be reasonable to respond to each action needn't correspond to the manner in which it would be reasonable to respond to the other action.

One may be tempted to say that in the case of Stevie, the mother just can't go wrong. Regardless of how she decides, she can reasonably predict that Stevie will be glad for it and reasonably so. However, I do not think that this conclusion is warranted either. There are people in the Deaf community and who work as determined advocates of deaf infants who seem to think that the mother would be mistaken were she to agree to cochlear implants for Stevie at such a young age. Like Harman, these advocates recognize that if everything goes well enough, Stevie would be completely reasonable in preferring whichever option his mother ended up choosing. Nonetheless, they argue that cochlear implants do a genuine disservice to the child's welfare and that the practice of providing cochlear implants as the default medical position is disrespectful to the Deaf community at large. I take this to be an open question. The point I want to emphasize is that when considering future attitudes leads to an inconclusive result, that doesn't mean both options are equally good; it just means that we need to continue the deliberation on other grounds.

So here we have seen different ways in which future attitudes *can* play some role in justifying a present action, even if those attitudes are person-affecting and even if those attitudes result from a transformative experience. Importantly, in none of these cases does the future attitude play the sole justificatory role. Moreover, there are many cases in which even when we entertain competing options, understanding our future pro-attitudes about these options is just not going to be sufficient in figuring out what we are

justified in doing. This is not to say that our future attitudes aren't actual sources of reasons for action and are always merely epiphenomenal of other reasons that we may have. Sometimes it is perfectly reasonable to invoke, "I'll be glad I did it and I'll regret it if I don't" reasoning as the justification for our actions. When one thinks about whether or not to get up to do a song during Karaoke night, the fact that one will be glad one did it and regret not doing it is a good enough reason to go up there.

When it comes to parents making decisions on behalf of children, however, the predicted attitudes do seem to be indicative of some other underlying reason. This leads to one final conclusion we can draw from this discussion. For parents choosing a potentially transformative experience for their child, they can make a justified decision using the predicted attitudes of the child without necessarily understanding *what makes it the case* that they ought to make that decision. In considering their child's future attitudes, parents needn't imagine what it is like to undergo the transformative experience first hand, which we have good reason to believe they would do poorly.[26] Rather, they just need to reasonably predict the child's attitudes and preferences that result from the experience. This can be done without understanding fully what it will be like for their child to be in those circumstances and what it is about their child's experience that will be of distinctive value. Instead parents can reasonably predict the child's future preferences by taking into consideration the testimony of others whose experiences more closely relate to those that the child is likely to undergo. In this way, my account offers a way for the predicted attitudes of others to serve a distinctive role in justifying our decisions on their behalf.

Dana Sarah Howard
E-mail : howard.1146@osu.edu

References:

Aas, Sean. Unpublished. "Disabled—Therefore, Unhealthy?"

Barnes, Elizabeth. 2009a. "Disability and Adaptive Preference." *Philosophical Perspectives* 23: 1–22. http://dx.doi.org/10.1111/j.1520-8583.2009.00159.x.

Barnes, Elizabeth. 2009b. "Disability, Minority and Difference." *Journal of Applied Philosophy* 26 (4): 337–355. http://dx.doi.org/10.1111/j.1468-5930.2009.00443.x.

Barnes, Elizabeth. 2014. "Valuing Disability, Causing Disability." *Ethics* 125 (1). http://dx.doi.org/10.1086/677021.

[26] At least hearing parents and deaf parents who have not undergone cochlear implant surgery are likely to be unable to fully appreciate the phenomenological character of living with the implants. See Paul (2015).

Acknowledgements I would like to thank the following people for helpful feedback on these issues: Benjamin McKean, Asha Bhandary, Caspar Hare, Julia Markovitz, David Estlund, Charles Larmore, Sharon Krause, Genevieve Rousseliere, Sean Aas, Derek Bowman, Alex King, Michael Conboy, Timothy Syme, Nicolas Bommarito, Inés Valdez, David Wasserman, Joshua Schechter, Christopher Hill. Special gratitude is given to Elizabeth Harman for her helpful and clarifying feedback on an earlier draft of this paper as well as two anonymous referees.

Chwang, Eric. 2009. "A Defense of Subsequent Consent." *Journal of Social Philosophy* 40 (1): 117–131. http://dx.doi.org/10.1111/j.1467-9833.2009.01441.x.

Connor, Carol, Holly Craig, Stephen Raudenbush, Krista Heavner, and Teresa Zwolan. 2006. "The Age at Which Young Deaf Children Receive Cochlear Implants and Their Vocabulary and Speech-Production Growth: Is There an Added Value for Early Implantation?" *Ear & Hearing* 27 (6): 628–644. http://dx.doi.org/10.1097/01.aud.0000240640.59205.42.

Dietrich, Franz and Christian List. 2011. "A Model of Non-Informational Preference Change." *The Journal of Theoretical Politics* 23 (2): 145–164.

Dougherty, Tom, Sophie Horwitz, and Paulina Sliwa. 2015. "Expecting the Unexpected." *Res Philosophica* 92 (2): 301–321. http://dx.doi.org/10.11612/resphil.2015.92.2.5.

Dworkin, Gerald. 1983. "Paternalism." In *Paternalism*, edited by Rolf Sartorious, 19–34. Minneapolis, MN: University of Minnesota Press.

Elster, Jon. 1983. *Sour Grapes: Studies in the Subversion of Rationality*. Cambridge: Cambridge University Press.

Groce, Nora. 1985. *Everyone Here Spoke Sign Language*. Cambridge, MA: Harvard University Press.

Guerrero, Alexander. 2007. "Don't Know, Don't Kill: Moral Ignorance, Culpability and Caution." *Philosophical Studies* 136: 59–97. http://dx.doi.org/10.1007/s11098-007-9143-7.

Hare, Caspar. 2011. "Obligation and Regret When There is No Fact of the Matter About What Would Have Happened if You Had not Done What You Did." *Noûs* 45 (1): 190–206.

Harman, Elizabeth. 2009. "'I'll Be Glad I Did It' Reasoning and the Significance of Future Desires." *Philosophical Perspectives* 23: 177–199. http://dx.doi.org/10.1111/j.1520-8583.2009.00166.x.

Harman, Elizabeth. 2011. "Does Moral Ignorance Exculpate?" *Ratio* 29 (4): 433–468.

Harman, Elizabeth. 2015. "Transformative Experience and Reliance on Moral Testimony." *Res Philosophica* 92 (2): 323–339. http://dx.doi.org/10.11612/resphil.2015.92.2.8.

Heathwood, Chris. 2008. "Fitting Attitudes and Welfare." *Oxford Studies in Metaethics* 3: 47–73.

Husak, Doug. 2010. "Paternalism and Consent." In *The Ethics of Consent: Theory and Practice*, edited by Franklin Miller and Alan Wertheimer, 107–131. Oxford: Oxford University Press.

Juster, Norton. 1961. *The Phantom Tollbooth*. New York, NY: Random House.

Liebling, A. J. 1962. *Between Meals*. New York, NY: Simon and Schuster.

McKinnon, Rachel. 2015. "Trans*formative Experiences." *Res Philosophica* 92 (2): 419–440. http://dx.doi.org/10.11612/resphil.2015.92.2.12.

Nussbaum, Martha. 2001. "Symposium on Amartya Sen's philosophy: Five Adaptive Preferences and Women's Options." *Economics and Philosophy* 17: 67–88. http://dx.doi.org/10.1017/S0266267101000153.

Ouellette, Alicia. 2009. "Hearing the Deaf: Cochlear Implants, the Deaf Community and Bioethical Analysis." *Valparaiso University Law Review* 45 (3): 1247–1270.

Paul, L. A. 2015. "What You Can't Expect When You're Expecting." *Res Philosophica* 92 (2): 149–170. http://dx.doi.org/10.11612/resphil.2015.92.2.1.

Rawls, John. 1999. *A Theory of Justice*. Revised edn. Cambridge, MA: Harvard University Press.

Sharadin, Nathaniel. 2015. "How You Can Reasonably Form Expectations When You're Expecting." *Res Philosophica* 92 (2): 441–452. http://dx.doi.org/10.11612/resphil.2015.92.2.2.

Silvers, Anita and Leslie Francis. 2005. "Justice through Trust: Disability and the "Outlier Problem" in Social Contract Theory." *Ethics* 116 (1): 44–77. http://dx.doi.org/10.1086/454368.

Superson, Anita. 2005. "Deformed Desires and Informed Desires Tests." *Hypatia* 20: 109–126. http://dx.doi.org/10.1111/j.1527-2001.2005.tb00539.x.

Thomson, Rosemarie. 1996. *Extraordinary Bodies: Figuring Physical Disability in American Culture and Literature*. New York, NY: Columbia University Press.

What's So Great about Experience?

Antti Kauppinen

Abstract: Suppose that our life choices result in unpredictable experiences, as L. A. Paul has recently argued. What does this mean for the possibility of rational prudential choice? Not as much as Paul thinks. First, what's valuable about experience is its broadly hedonic quality, and empirical studies suggest we tend to significantly overestimate the impact of our choices in this respect. Second, contrary to what Paul suggests, the value of finding out what an outcome is like for us does not suffice to rationalize life choices, because much more important values are at stake. Third, because these other prudential goods, such as achievement, personal relationships, and meaningfulness, are typically more important than the quality of our experience (which is in any case unlikely to be bad when we realize non-experiential goods), life choices should be made on what I call a story-regarding rather than experience-regarding basis.

On the standard picture of rational choice, we should choose the option that has the highest expected value. Expected value, in turn, is the sum of the values of possible outcomes of the option multiplied by their probability. The value of many possible outcomes, like eating some delicious chocolate, is largely a matter of *what it is like for us* to experience them. As I will say, the value they have is mainly *experiential value*. If we don't know what it's like to experience them, we won't be able to form well-grounded beliefs about their value, nor consequently make normatively significant rational choices regarding them.

In her novel and exciting *Transformative Experience*, L. A. Paul (2014; see also Paul 2015) argues that especially when it comes to important life choices, such as choosing where to live or whether to have a child, the possible outcomes involve experiences that are *epistemically transformative* in the sense that we cannot know what they are like for us until we have actually experienced them, and hence cannot form rational estimates of their experiential value, or *personally transformative* in the sense that they change our preferences in unpredictable ways. This calls into question the very possibility of making rational choices about such matters. Ultimately, however, Paul is not a skeptic. She believes that there is another kind of

Res Philosophica, *Vol. 92, No. 2, April 2015, pp. 371–388*
http://dx.doi.org/10.11612/resphil.2015.92.2.10

value of possible outcomes that can at least in some cases serve as the basis for rational and authentic choice, namely *revelatory value*: when we choose a transformative option, we choose to *find out what it is like* for us to experience the outcome, or find out what *we* will be like after the experience. Sometimes, then, it is rational to choose to come to learn what an experience is like or how we will change, she maintains.

In this paper, I will examine and reject three theses about the value of experience that feature in Paul's argument:

> **Non-Hedonism**: The intrinsic value of an experience is not determined by its hedonic quality or contribution to happiness.
>
> **Value of Veridicality**: Veridical experiences are more intrinsically valuable than non-veridical experiences.
>
> **Sufficiency of Revelatory Value**: The value of coming to know what it is like for us to have epistemically transformative experiences or how our preferences change as a result of personally transformative experiences suffices to ground rational choice in (at least some) major life decisions.

(Note that Value of Veridicality entails Non-Hedonism, but not vice versa.) Against the first thesis, I argue that the intrinsic prudential value of experiences *is* exclusively hedonic (when understood broadly to encompass contribution to happiness). I provide an error theory for why other features of experience, such as variety or richness or particular phenomenal character, may seem intrinsically valuable. Against the second thesis, I argue that experiential value supervenes exclusively on the phenomenal character of experience, which is identical between veridical and non-veridical experiences. Finally, I reject the sufficiency of revelatory value for important rational choices.[1] While it is indeed good for us to find out something about experiential value and ourselves, the good involved in such knowledge isn't great enough to justify choosing an option that may be very bad for us. It may be good to come to know the hard way that one hates being a parent, but the positive value of coming to know the unpleasant truth is radically outweighed by the risk of realizing the negative value of the unpleasant truth itself. In brief, no one should have a child in order to find out whether one hates or loves having a child, when much more important values are at stake.

If we can't anticipate what our experience will be like and revelatory value doesn't suffice for rational choice, must we make our life choices without a normatively significant rational basis? Only if there are no other significant prudential values that could do the job. Fortunately, there are. Many other things besides the quality of our experience are

[1] While Paul at least suggests this thesis, as we'll see below, she has indicated in correspondence that she does not fully endorse it herself.

intrinsically good for us, and that is reflected in common preferences. For the purposes of my argument here, it doesn't matter precisely what these non-experiential goods are. Popular candidates include achievement, friendship, developing and exercising our rational capacities, and meaning in life. I have argued elsewhere (Kauppinen 2012, Kauppinen Forthcoming-b) that there is a notion of a prudentially good life story that nicely unifies these non-experiential values, and will employ the terminology in my sketch here, but one need not accept my particular account to see that there is room for cautious optimism about the possibility of rational life choices on the basis of non-experiential values, even if experience is transformative in the way Paul argues.

This optimism is further buttressed by two considerations regarding experiential value. First, precisely in the case of life choices, experiential value is relatively insignificant in comparison to non-experiential values like being in a valuable relationship or scientific achievement. That is the truth in Mill's (1863, 14) notion that it's better to be Socrates dissatisfied than a fool satisfied. So it may be rational to choose an option we anticipate to have high non-experiential value, even if it means risking bad experiences. Second, since what intrinsically matters about experience is its contribution to our happiness, we can look to the science of happiness to draw some conclusions about the likely effect of our choices. As it turns out, the most important result from the science of affective forecasting is that when it comes to major life events, we radically overestimate the difference they make. In fact, when it comes to experiential value, the result is likely to be a wash in the case of hard choices: in the long run, our average level of happiness is likely to be roughly the same whether we, say, have a child or not have a child.

In brief, our life choices, in particular, should be *story-regarding* rather than *experience-regarding* in order to be rational in the normatively significant sense. We shouldn't be concerned with how they will affect our experience, but rather, roughly, with what they mean for the successful pursuit of something objectively valuable that builds on our past efforts and experiences, and is consistent with our commitments. This is the rule followed, for example, by those who choose to have a child because they see it as the next stage in their evolving relationship with someone they love, and those who choose not to have a child because they dedicate themselves to some cause they believe in.

1 Experiential Value Is Broadly Hedonic

Here are some experiences I'm confident it is good for at least some of us to have (or so my own experience suggests):

- Tasting Gino's raspberry ice cream
- Performing music with friends in front of an excited audience

- Hearing your son come up with a delightful new word
- Feeling the medication take effect

Note that we often individuate experiences by reference to their content, as I do in the list—I am talking about the experience of performing music, for example, not about the action that results in the experience. (One might, theoretically, have the experience without the action.) Sometimes we can also talk about experiencing the *taste* of ice cream, say. Other experiences, such as the feeling when the medication starts to work, don't have such content. Further, it is not trivial to provide a criterion for identifying *an* experience—we say that some people experienced World War II, which, if true, is very different from experiencing what it's like to eat a bowl of ice cream. For my purposes, these distinctions are unimportant, and I will continue to refer to experiences in various ways.

Here are some experiences I'm confident it is bad for at least some of us to have (or so my own experience suggests):

- Eating Hershey's chocolate
- Getting tongue-tied and flustered in front of an audience of respected colleagues
- Losing a parent
- Placing a hand on top of burning hot steam rising from a sauna stove

What is good or bad about these experiences? One obvious candidate is that some are pleasant and others are unpleasant. It is not, in my view, felicitous to say, in general, that the experiences *cause* pleasure or displeasure. This way of speaking suggests that pleasure and pain are sensations that are distinct from the experience. It is better to say that pleasure and displeasure are aspects of the experiences: it is part of what it is to like to have the experience that it is pleasant or unpleasant. The hedonic quality of an experience is part of its phenomenal character. If one person's experience of eating Hershey's chocolate is unpleasant and another person's experience of it is pleasant, they don't have the same experience of eating chocolate. This way of thinking about pleasure does not commit us to think that there is any single common experiential quality common to all pleasant experiences: as far as it goes, they may be many distinct ways of being pleasant. (For the purposes of this paper, I am not going to take a stand on the nature of pleasure.)

So here is a hypothesis about the value of experience:

> (**Prudential**) **Hedonism about Experiential Value**: What is intrinsically valuable about experience for someone is its hedonic quality: when it comes to experience considered merely as such, the more pleasant it is like to have the experience, the better it is for the subject to have the experience.

Note that (Prudential) Hedonism about Experiential Value (HEV) is distinct from hedonism *sans phrase*, the thesis that pleasure (and the absence of

pain) is the only intrinsic good. The former is only a claim about the value of *experience*, and allows that there may be other intrinsically good things that have nothing to do with experience, such as achievement or meaningfulness. Also, since HEV is a thesis about intrinsic value, it doesn't deny that it can be good for us to have unpleasant experiences. Sometimes such experiences teach us something about the world, or indeed about ourselves, and such knowledge may be good for us. But this is merely instrumental value. Individual unpleasant experiences may also result in more pleasant experiences in the future. This is a different way they can be of instrumental value. But considered on their own, they are bad for us. Only pleasant experiences are good *as* experiences, regardless of what follows from them.

HEV easily accounts for the value of the experiences I listed above. Yet L. A. Paul explicitly denies HEV:

> I take these values of experiences to be values that do not reduce to anything else: they are primitive and they are not merely values of pleasure and pain. Instead, the values are widely variable, intrinsic, complex, and grounded by cognitive phenomenology. So such values, as I shall understand them, are values that can be grounded by more than merely qualitative or sensory characters, as they may also include arise from nonsensory phenomenological features of experiences, especially rich, developed experiences that embed a range of mental states, including beliefs, emotions, and desires. (2014, 12)

So Paul maintains that some experiences, especially "rich, developed" ones, are intrinsically good for us beyond their hedonic quality. Perhaps the claim is that the experience of performing music with friends, for example, is a valuable experience to have just in virtue of its distinctive "cognitive phenomenology," which is different from any other experience. This, however, is a highly dubious claim. After all, it amounts to claiming that valuable experiences have nothing in common *qua* experiences apart from being valuable. Or does Paul claim that it is intrinsically good for us to have *rich and developed* experiences? To test this theory against HEV, we need to look at cases of rich and developed experiences that are not pleasant in any way. If they are *intrinsically* good for us to have, Paul's theory has an advantage over HEV.

Alas, it is hard to think of such experiences. Suppose you go a performance of Macbeth, which you follow attentively. Plausibly, your experience is rich and developed. Is this a good experience for you to have? Well, it's not a stretch to assume that it is *instrumentally* good for you—perhaps it yields some insight into Shakespeare or even the human condition that you wouldn't have otherwise had. To that extent, it's a means to something intrinsically good, perhaps even a necessary means. But is it good in itself?

If it is not in any way an *enjoyable* experience, the answer seems to be negative. After all, if you neither got anything further from the experience nor enjoyed it, what would be in it for you? Or consider the experience of being exquisitely tortured. It might be rich and developed. But unless its being rich and developed was to some extent your liking, this wouldn't make it any better. It might be less boring, and hence in one respect involve less suffering, and so hedonistically preferable, if other things were equal. But if it wasn't, being rich and developed wouldn't in itself be a redeeming feature.

In other places, Paul suggests other non-hedonic criteria for valuable experience. Consider the following passage:

> Our experiences, especially new ones, are valuable, that is, we value having them, and we especially care about having experiences of different sorts. As such, experiences have values that carry weight in our decision-making. (2014, 11)

Here, it seems, Paul suggests that good-making features of experience include *novelty* and *variety*. This is an appealing thought. But it seems to me that the appeal is illusory. The reason is that novelty and variety are unquestionably *instrumentally* valuable features of experience, so while they are good, they are not *intrinsically* such. It's a well-established fact that for many kinds of experience, repetition reduces the hedonic quality of the experience. Watching *Groundhog Day* is a positive experience for most of us. Watching *Groundhog Day* again, and again, and again less so. We'd rather have a new kind of experience. But that's because novelty is often pleasant, and so is variety. (Although there are people, like Elvis, who prefer the predictable experience of eating a cheeseburger for lunch every day.) So while we do indeed care about having experiences of different sorts, it's because we don't want to get bored and lose in terms of pleasure. It is no doubt good, because delightful, for Frank Jackson's (1982) Mary to see red for the first time. But would it really be good for her to be introduced to a new, slightly different shade of red (or some other color) every day for the rest of her life? Hardly—because it would hardly be an enjoyable life, although it would involve a constant stream of new and different experiences.

Paul has one more argument against HEV. She maintains that *veridical* experiences are more valuable than non-veridical ones:

> I will assume that an experience has this sort of value only when it correctly represents what's in the world or it is produced in the right way. So these values are values for lived experience, where such experience is "real" or veridical. (2014, 11)

If this were the case, HEV would be false, since the hedonic quality of experience is independent of how it is produced or its veridicality. However, there is good reason to believe this is not the case. The argument is simple:

(1) If the value of tokens of *X* depends solely on what they are like in respect *R*, then tokens *A* and *B* of *X* can differ in value only if *A* and *B* are not alike in respect *R*.
(2) The intrinsic value of an experience depends solely on what it's experientially like.
(3) Veridical and non-veridical experiences are experientially alike.
(4) Hence, veridical and non-veridical experiences cannot differ in intrinsic value.

The first premise is a kind of supervenience thesis, which can hardly be denied (it may even be a conceptual truth). The second premise says that the intrinsic value of experiences depends only on their quality as experiences, on the what-it-is-like to have them. It doesn't deny that veridical and non-veridical experiences have different *instrumental* value. Only veridical experiences tell us something about the world, and may yield knowledge. In that sense veridical experiences are better than non-veridical ones, assuming knowledge is intrinsically or instrumentally good. But *as experiences*, considered apart from their consequences, their value depends solely on their experiential quality.[2] The third premise simply points out that there is no difference between veridical and non-veridical experiences in this respect. Their intrinsic qualities are identical. So it is no surprise that their intrinsic (prudential) value must be identical.

So I don't find Paul's arguments against HEV convincing. Nevertheless, I do believe it needs to be modified for the sort of reasons that Dan Haybron has pointed out. Haybron (2001) notes that some pleasures leave us cold or fail to touch us, and thus fail to contribute to our happiness. It is plausible to me that such peripheral pleasures are not good for us, or are only marginally good. Equally importantly, Haybron argues that moods and positive emotional states, such as being calm, relaxed, or in the 'flow,' contribute to our happiness over and above their hedonic quality (Haybron 2008). Some, such as Paul Dolan (2014), might add that sense of purpose or reward is an independent element of happiness or positive experience. It is not necessary here to go into detail of emotional condition theories of happiness (see Kauppinen Forthcoming-a), but assuming that it is apt

[2] It is true that Fred Feldman considers the possibility of what he calls truth-adjusted hedonism, according to which pleasure taken in a truth is more valuable than pleasure taken in a falsehood, even if there is no experiential difference (Feldman 2004, 111–114). But first, while this view denies premise 2, it is still a form of hedonism, as Feldman emphasizes: only the hedonic aspect of experiences matters to their intrinsic value, although the degree to which it matters hangs in part on veridicality. Second, and more importantly, the idea of adjusting the value of pleasure for truth seems *ad hoc*—the only motivation for doing so is avoiding counterexamples to hedonism that appeal to 'false pleasures,' such as pleasure taken in the mistaken belief that one is loved and respected by others.

to label the aspects of experience that contribute to positive emotion as broadly hedonic, something close to it may well be true:

> **Broad (Prudential) Hedonism about Experiential Value:** What is intrinsically valuable about experience is its broadly hedonic quality: when it comes to experience considered merely as such, the more it directly contributes to happiness (i.e., the higher the degree to which it is happiness-constituting), the better it is for the subject to have the experience.

If Broad (Prudential) Hedonism about Experiential Value is correct, our epistemic burden is reduced when it comes to making rational choices on the basis of the experiential quality of the outcomes. We don't need to know exactly what the possible experiences are like, since the only aspect that matters for their value is their broadly hedonic quality. I will return to the implications of this in the final section.

2 The Relative Unimportance of Revelation

It is useful to divide Paul's argument in *Transformative Experience* into two parts. The first part is skeptical, and the second constructive. The skeptical argument begins with the claim that in order to make rational choices—or, as she puts it, to meet the normative standard for choice—we must assign both the probabilities and values of possible outcomes of our options on the basis of evidence. As Paul says,

> If we are to meet the normative standard when we make our choices, we must be rationally justified in our assignments of values and credences to the outcomes and states of our decision problem. That is, we must assign our values and credences based on sufficient evidence. (2014, 22)

This is a substantive and potentially controversial thesis, since it involves rejecting the strictly subjectivist view that any preferences meeting the axioms of decision theory are a possible basis for rational choice. I am not going to question this part of Paul's argument. I am happy enough to grant that if you prefer back-breaking labor in a coal mine to a happy life of luxury and leisure, other things being equal (as far possible), it is in one sense rational for you to choose it. But I will say that *normatively significant* rational choice requires that our preferences are not arbitrary but are based on evidence about the value of the possible outcomes. That is, if a choice's being rational is going to have a bearing on what you *should* do, the preferences that underlie it must be based on evidence about what is actually valuable. In this way, the theory of normatively significant rational choice connects with value theory.

Given the assumption that normatively significant rational choice requires not only evidence about probability but also evidence about value,

the involvement of transformative experiences calls the possibility of rational choice into question, when the value of an outcome is importantly experiential. And Paul argues that when it comes to some of the most important life choices we make, the experiential (or as she puts it, "subjective") value swamps other values. For example, she says that "Major life decisions determine our personal futures, and centrally concern what it will be like for us to experience the futures we make for ourselves and those we care about" (2014, 23). For example, in deciding whether to have a child or not, we (educated middle- or upper-class Westerners) naturally and rightly put aside other people's expectations, and consider what it would be like for us to be a parent. It would be *inauthentic* to make such choices on the basis of what others think. But insofar as becoming a parent is a transformative experience, we simply do not have sufficient evidence regarding what it's like to be a parent, and hence cannot make a rational choice on (what Paul regards as) the usual basis.

So goes the skeptical argument in outline. In the next section, I'll say a little bit about how it might be countered. But first, I want to examine Paul's own non-skeptical argument. For she doesn't think that rational life choices are impossible. That's because there's another value that experiences can have. Think of tasting a new kind of fruit. Beforehand, you are not in a position to know what the experience will be like. But you do know something: once you've tasted it, you *will know* what it's like. And that may be valuable knowledge. Here the epistemically transformative experience has *revelatory* value: without the experience, you would never have come to know what it is like for you to eat that kind of fruit. In this vein, Paul appears to argue that revelatory value is a possible rational basis for making transformative life choice. For example, she says that "I'll argue that the best response to this situation is to choose based on whether we want to discover who we'll become" (2014, 4), and later that "the proposed solution is that, if you are going to meet the normative rational standard in cases of transformative choice, you must choose to have or to avoid transformative experiences based largely on revelation: you decide whether you want to discover how your life will unfold given the new type of experience" (2014, 120). Or in more detail:

> When we choose to have a transformative experience, we choose to discover its intrinsic experiential nature, whether that discovery involves joy, fear, peacefulness, happiness, fulfillment, sadness, anxiety, suffering, or pleasure, or some complex mixture thereof. If we choose to have the transformative experience, we also choose to create and discover new preferences, that is, to experience the way our preferences will evolve, and often, in the process, to create and discover a new self. On the other hand, if we reject revelation, we choose the status quo, affirming our current

> life and lived experience. A life lived rationally and authentically, then, as each big decision is encountered, involves deciding whether or how to make a discovery about who you will become. (2014, 178)

It is undeniable that transformative experiences have revelatory value, as Paul defines it. The only question is whether such value suffices to make one's choice rational in the normatively significant sense. Take, once again, the choice of whether to become a parent. If I don't have a child, my life will go on much as before, although I can't be quite sure what it is like to be childless when I'm older—let's say that in my case, the utility of this choice is between 20 and 60. This means also that I'll never find out what it would have been like to have a child, which may have some disutility for me (although I will find out what it is to live childless into old age[3]). If I have a child, I will find out what it is like to be that particular child's parent (and as Paul argues, there's just no other way to find out). This is valuable information—let us stipulate that it gives me 10 utils. This has to be balanced against the disutility, of say −5, of never finding out what it is like to remain childless for the rest of my life—especially in life choices, we must bear in mind the opportunity cost of learning what it is like to choose one way.

Of course the discovery of what it is like is not the only outcome of having a child. It also means, among other things, that I will have the experience of being that particular child's parent, which, Paul assumes, may be fantastic or terrible, and make a huge difference for how the rest of my life goes. (I will later call this assumption regarding experience into question, but since becoming a parent will also have consequences of non-experiential value, it is nevertheless true that it can make a vast difference to how good my future will be.) Let us say that it will have a value or disvalue somewhere between 100 and −100 utils for me—the problem being precisely that I don't know where the experience falls on that scale.

Here is a decision matrix, ignoring other outcomes. In line with standard decision theory, the expected value of an option is the sum of the values of possible outcomes once they've been multiplied by their probability:

[3] As a referee pointed out, the fact that life choice situations are typically symmetrical in this way—if you marry, you'll discover what it's like to be married, if you don't marry, you'll discover what it's like to remain unmarried (which is not going to be the same as having been unmarried until now)—means that revelatory value is not going to rationalize choice in either direction. Paul tends to write as if not choosing a new thing means that things will go on as before, so that there's no revelation to be had. Here's a representative passage: "In either case, when choosing to have a child or choosing to remain childless, if you choose rationally, you choose on the basis of whether you want to discover new experiences and preferences or whether you want to forgo such a discovery. You choose whether you want revelation, or whether you don't" (2014, 120). But as I've said, you'll get unpredictable new experiences either way, so this can't be the right description of the choice situation: you'll get revelation whether you want it or not!

Option	Outcomes	Values of Outcomes	Probabilities of Outcomes	Expected Value of Option
Having a child	Finding out what it is like to be the particular child's parent, not finding out what it's like to keep living without child	10u − 5u = 5u	1	5u+??
	Having the experience of being the particular child's parent	−100u to 100u	??	
Not having a child	Not finding out what it is like to be a parent, finding out what it's like not to be a parent	−10u + 5u = −5u	1	−5u+??
	Leading a life that is unchanged in this respect	20u to 60u	??	

So, taking into account the revelatory value of having a child, can I now make a rational choice about whether to have a child? No! There are too many question marks in the matrix. I still don't know whether my life will be miserable or glorious with a child, nor for that matter what it will be like if I never have one. While I may want to know what it is like to have a child, there are things I want even more, such as leading a happy life and avoiding spending the rest of my life in worry and misery. While it is rational for me to value coming to know what it is like to have a particular child, it is not rational for me to value this knowledge more than my future happiness or other prudential goods. As the matrix shows, even if I give a rather large value to coming to know what it is like to have a child, the value of revelation dwarfs in comparison to my future quality of experience, not to mention other prudential goods. Insofar as I genuinely can't give a rational estimate to what the broadly hedonic (and other prudentially significant) consequences of a choice are, it is deeply irrational for me to make the choice on the basis of the relatively minor value of coming to know what an experience is like or how my preferences will change.

What might Paul say in response? When she writes about making a choice on the basis of revelatory value, she talks about "reframing" or "reconfiguring" our choices in terms of coming to know what it's like, leaving aside the experiential value (which she calls "subjective value" or "subjective well-being"). Here are two typical passages:

> To configure this decision to make it rational, we need to keep in mind, again, that the values of these outcomes are *not* determined by whether the experience involved is good or bad, but solely by the subjective value of the discovery of the nature of the experience, whatever it is like. (2014, 114)

> Similarly, the decision to have a child could be understood as a decision to discover a radically new way of living with correspondingly new preferences, whether your subjective well-being increases or not. (2014, 119)

I agree that one could, *de facto*, make life decisions on such a basis. The problem is that doing so would not be rational *in the normatively significant sense*. Imagine someone making the choice of whether to become a vampire on the basis of wanting to stay out all night and sleep during the day. That would be a possible basis for making the choice, and it would be possible to opt for becoming a vampire in a kind of rational manner this way. But it would hardly be rational in the normatively significant sense to simply ignore the most important things that are at stake in the choice when making it (such as what it is like to be immortal, to live off people's blood, and so on)—the kind of things that decisively matter for one's subjective (and objective) well-being. In general, we can't be rational in the normatively significant sense if we *ignore* the values of some of the outcomes of possible choices. (If I'm thinking about which restaurant to go to, I can't rationally ignore the price and simply make the choice on the basis of which one serves the best food.) To be "rationally justified in our assignments of values" to options, we must take all the values of possible outcomes of the option into account, in particular those that significantly affect our future well-being. Thus, when reframing or reconfiguring a choice means leaving significant values out of the calculation (whether they are experiential or non-experiential), it results in a choice that is not rational in the normatively significant sense that Paul herself is interested in.

So, in short, while Paul is right in that transformative experiences have revelatory value, such value is not sufficient to rationalize life choices, if their effects on the agent's subjective and objective well-being are unknown and unknowable. Unless there is some other basis for rationally assigning values to outcomes, the skeptical argument carries the day.

3 Beyond Experiential Value

In the previous section, I endorsed Paul's requirement for normatively significant rational choice: we must have justified beliefs about the value of possible outcomes as well as about their probability. I haven't called into question her claim that in the case of transformative experience, it is not

possible for us to form justified beliefs about what the possible outcomes are like for us, but I have rejected her own proposed solution for how to make rational life choices on the basis of revelatory value. Should we then be skeptics about the possibility of rational life choices?

No, we shouldn't, although we shouldn't expect such choices to be easy either. In this section, I will sketch an argument that gives us some reason for optimism about the possibility of rational life choices in spite of everything. The argument hangs on two main assumptions. First, there are other kinds of prudential value that are arguably more significant than experiential value. Insofar as we can reliably enough predict what our choices mean for the realization of these non-experiential prudential values, we can after all rationally assign values to outcomes even if they involve transformative experience. Second, while in the case of transformative choice, we can't predict exactly what our experiences will be like, it turns out not to matter so much. This is because precisely when it is hard to know what our life will be like, it is likely that there is no dramatic difference in experiential value between the possible outcomes in the long term. This strengthens the case for making the choice on the basis of non-experiential values.

First, then, I will offer a brief sketch of why experiential value is relatively unimportant (I give a fuller account elsewhere). There are *non-experiential* prudential goods—things that are good for me regardless of the quality of my experience. I take it that this is an overwhelmingly plausible assumption on the face of it. The most famous argument for it is, of course, Robert Nozick's (1974) Experience Machine thought experiment. There are many ways to construe it, but for my purposes, the essential point is that a person who is only concerned with her own good would be better off actually leading the life of her dreams—such as being a Nobel Prize-winning rock star and Wimbledon champion—than having a perfect machine-generated illusion of leading the same life. The thought experiment is silent on just why this axiological fact obtains. Nozick's own suggestion—that being in touch with reality matters for its own sake (1974, 42)—isn't particularly plausible. There are, after all, many things that are absent in the machine scenario as a result of not being in touch with reality. For example, there are no significant achievements and no significant relationships with actual other individuals, and little autonomy or knowledge. Consequently, life inside the experience machine has very little meaning (Kauppinen 2012, Metz 2013).

All these things are candidates for non-experiential intrinsic prudential goods. When it comes to non-experiential value, Objective List theorists mention things like achievement, friendship, and self-respect as things that are intrinsically valuable for us to have (Fletcher 2013). Perfectionists talk about the development and exercise of human capacities, such as practical and theoretical reason, and emotional and physical skills (Kraut 2007). I have recently argued that a *narrativist* account of non-experiential prudential value captures the truth in both of these accounts, since prudentially

good life histories involve successful pursuit of objectively valuable goals in a way that makes intelligent use of our capacities and builds on our past (Kauppinen 2012, Kauppinen Forthcoming-b). For my purposes here, any of these answers would do. What matters is that outcomes of our choices have value for us that is independent of our possibly unpredictable experience of them.

Suppose, for example, that other things being equal, it is better for me to create something of higher aesthetic value than something of lower aesthetic value—that artistic achievement is intrinsically good for me. The prudential value of producing great art isn't reducible to my own *experience* of doing so. Maybe I can't know what it's like to create a truly great painting before I've done so. Maybe it doesn't feel that great. But it may nevertheless be good for me to succeed in such a project. I will not have wasted my time, but will have drawn on my unique history and abilities to create something that no one else could have. That this is a valuable outcome is something I could have known beforehand. And indeed, people do. Presumably Gauguin didn't know what it would be like for him to leave his family and move to Tahiti. Nor could he have known that he would succeed in producing art of great value. But he was in a position to know that it is better for him to become a great artist than to remain a mediocre one, and perhaps in a position to form a rational estimate that he was more likely to become a great artist if he left his family than if he stayed in France. In any case, the decision problem wasn't about which outcome is better and which worse for him. It was about which action is more likely to bring about which outcome.

Factual uncertainty, obviously, is always going to be a problem with life choices. A theory that implied it is *de facto* easy to make rational life choices would be implausible. I can't know for sure what happens when I marry Gary. I won't be completely in the dark, if I know him (and myself): I've got evidence to support forming credences regarding how our relationship might develop and what commitment would mean for my other projects. I'm in a better position to assign values to possible outcomes. It will be good for me to stick with someone who has seen me at my worst and stuck with me. It will be good for me to commit to a relationship that benefits from what I've been through in the past. It is good if I'm in a relationship that nourishes projects that do some real good beyond the confines of my own life: for example, I should be with someone who supports me in becoming a better teacher and researcher, and inspires me to do right by strangers who need my assistance. I will say that when my choice is explicitly or implicitly guided by this kind of consideration (in addition, obviously, to assignments of choice-dependent probabilities to outcomes), it is a *story-regarding* one. It should be clear that story-regarding choices are *authentic* in the sense that Paul deploys—they involve thinking about "who you really are and what you really want from life" and taking "charge of your own destiny" (2014, 105) rather than letting the preferences, values,

or even needs of others determine what we do. So they can offer the kind of basis for rational life choices that Paul herself accepts, and not some ersatz substitute.

But, Paul might object, if we make story-regarding choices, aren't we guilty of irrationally ignoring what matters most about our life choices, namely what it will be like for us to lead a particular kind of life? (This is parallel to my own complaint against making choices on the basis of revelatory value.) I think there are two reasons why this objection is weak. First, when it comes to determining the overall prudential value of an option, especially in the case of life choices, non-experiential values are typically weightier than experiential values. I admit that this is not easy to show, in part because values of, say, achievement and pleasure are plausibly incommensurable. But it is something that is manifest in people's actual choices. Faced with having to choose between integrity, commitment, friendship, meaning, or achievement, on the one hand, and happiness on the other, we frequently go with the former option. Not everyone and not always, to be sure. But this brings us back to Mill's Socrates and the swine. Mill himself, problematically, frames the distinction in terms of higher and lower pleasures (1863, 11–17). But the basic point he's making—that those who have experience of, say, artistic achievement or the use of "higher faculties," prefer a life that involves such goods to a life that lacks them, even if the latter holds more happiness for them—still holds. Of course, we'd rather have good experiences along with non-experiential goods—and indeed, experience suggests that we're more likely than not to feel good when we enjoy a thriving friendship or succeed in an academic endeavor, for example. But we can rationally take the risk of bad experiences, if we thereby gain in some significant non-experiential goods. That's what happens when we make story-regarding choices in ignorance of what the outcomes will be like for us. I thus deny Paul's claim, already quoted above, that major life decisions "centrally concern what it will be like for us to experience the futures we make for ourselves and those we care about" (2014, 23). The quality of our future experience is just one consideration, and frequently not the most important one.

The second reason why we shouldn't worry too much about our ignorance of future experience is that in the long run, the choices we make are unlikely to matter too much to the quality of our experience, at least when the effect is genuinely unpredictable. This claim is supported by empirical psychology. I argued in the first section in favor of Broad Hedonism About Experiential Value—roughly, experiences are good for us *qua* experiences insofar as they directly contribute to our happiness (insofar as they are happiness-constituting). I also observed that this simplifies our epistemic situation: in order to form rational estimates of experiential value, we don't need to know exactly what an outcome is like for us, but just its broadly hedonic quality. This task is arguably easier—even if I have only the remotest idea of what it's like to eat durian fruit, I do know that it

won't be as horrible as having a tooth pulled out, nor as enjoyable as winning a Nobel Prize. Still, it is difficult. Psychological research on what is known as 'affective forecasting' suggests that people are quite bad at predicting what, how intense, and how long-lasting their affective responses are in various possible contingencies (Wilson and Gilbert 2003). Even in the case of non-transformative experience, we misconstrue future events, frame them misleadingly, have poor recall of past experiences, rely on bad but culturally prevalent theories, allow our current experience to bias our expectations, focus narrowly on just one aspect of the event, are ignorant of our psychological defense mechanisms, and so on. Clearly, we're not great judges of broadly hedonic value.

However, according to Timothy Wilson and Daniel Gilbert, the "most prevalent error" in affective forecasting is *impact bias*, whereby "people overestimate the impact of future events on their emotional reactions" (Wilson and Gilbert 2003, 353). Study after study has shown that the impact of future events and changes in our life on our affective condition is much smaller and more short-lived than we think. People expect that they'll be unhappy if they fail to get a job, break up with their partner, fail to get tenure, lose a limb, or, perhaps most pertinently for our purposes, have a child with Down syndrome. But in fact, after a period of adjustment that is much shorter than most people expect, their affective state typically returns to its ordinary level, or close to it.

To be sure, there are some circumstances people don't tend to adjust to. For example, it is, unsurprisingly, tough to be the primary caregiver to a severely disabled child, in particular without family and community support (Cummins 2001). But this is not a problem for the present argument, since it is not unpredictable that such outcomes are low in experiential value (even if we can't know exactly what it is like to take care of a severely autistic child, say, before we've done so). If we know all the facts about living with an abusive spouse, say, apart from what we can only learn by actually leading such a life, we already know enough to know that it's bad for us. Transformative experiences are not a barrier for rationally estimating the value of such outcomes. These outcomes only pose the traditional challenge to any rational decision-making: it can be hard to form justified beliefs about their likelihood—it can be hard to find out whether a child will turn out to be severely disabled or a partner turn out to be abusive.

Here, then, is a brief argument in favor of thinking that we can make rational life choices, even if we accept that they involve transformative experiences, and deny that revelatory value suffices to rationalize choice in the normatively significant sense:

(1) Rational choice in the normatively significant sense requires justified belief about the relative values of outcomes and their probability.

(2) We can (often/at least sometimes) form justified beliefs about the narrative value of outcomes, regardless of whether they involve transformative experiences.
(3) We can (often/at least sometimes) form justified beliefs about the probability of possible narrative outcomes, given our choices.
(4) So, we can (often/at least sometimes) make life choices that are rational in the normatively significant sense insofar as they are story-regarding. (1, 2, 3)
(5) The narrative value of possible futures typically trumps experiential value in the case of life choices, especially since life choices are unlikely to make a lasting difference to experiential value (except in exceptional and predictable circumstances).
(6) So, it is typically or at least sometimes prudentially rational in the normatively significant sense to make life choices that are story-regarding rather than experience-regarding. (4, 5)

I don't want to pretend that the conclusion is stronger than it is. We can't always reliably estimate narrative value, or what kind of turns our life history will take, given a choice. And in atypical circumstances, life choices may have both dramatic and unpredictable lasting impact on our experience. In such rare cases, the skeptical part of Paul's argument remains unanswerable.

4 Conclusion

Life choices are difficult. In part, they are difficult for us because we are unable to estimate the difference they make to our future experience. But the quality of our experience is not the only thing that is at stake, nor is it the most important consideration, even if we restrict ourselves to self-interested choices. So when we decide which job to take or what kind of family to have, if any, it is rational for us to focus on the *non-experiential* consequences of our choices. One relatively minor consequence is that we will discover what it is like for us to live in a certain way (while never finding out what it would have been like, had we chosen the other way). But there are far more important values at stake. Which option will put us in a better position to achieve something genuinely valuable? Which choice involves more intelligent use of our abilities? What do the options mean for our existing commitments? Which outcome would better build on our past efforts or redeem failures? When we make the decision on the basis of solid evidence regarding the likely consequences of our choice to such non-experiential sources of value, it has a good chance of being both authentic and rational in the normatively significant sense, especially since the odds are that our choice won't have a dramatic effect on the overall quality of our experiences. Indeed, it seems likely that insofar as there are lasting effects on experience, they roughly track the trajectory

of non-experiential value—when we succeed at finding meaningful work, building a good personal relationship, or creating a work of art, realizing the non-experientially valuable outcome is likely to have a positive effect on experience as well. So while positive experiences are genuinely valuable for us, we are better off focusing on non-experiential values, especially when it comes to life choices like deciding whether to have a child.

Antti Kauppinen
E-mail : a.kauppinen@gmail.com

References:

Cummins, Robert A. 2001. "The Subjective Well-Being of People Caring for a Family Member with a Severe Disability at Home: A Review." *Journal of Intellectual and Developmental Disability* 26 (1): 83–100. http://dx.doi.org/10.1080/13668250020032787.

Dolan, Paul. 2014. *Happiness by Design: Finding Pleasure and Purpose in Everyday Life*. London: Allen Lane.

Feldman, Fred. 2004. *Pleasure and the Good Life: Concerning the Nature, Varieties, and Plausibility of Hedonism*. New York, NY: Oxford University Press.

Fletcher, Guy. 2013. "A Fresh Start for an Objective List Account of Value." *Utilitas* 25 (2): 206–220. http://dx.doi.org/10.1017/S0953820812000453.

Haybron, Dan. 2001. "Happiness and Pleasure." *Philosophy and Phenomenological Research* 62 (3): 501–528. http://dx.doi.org/10.1111/j.1933-1592.2001.tb00072.x.

Haybron, Dan. 2008. *The Pursuit of Unhappiness*. New York, NY: Oxford University Press.

Jackson, Frank. 1982. "Epiphenomenal Qualia." *The Philosophical Quarterly* 32: 127–136. http://dx.doi.org/10.2307/2960077.

Kauppinen, Antti. 2012. "Meaningfulness and Time." *Philosophy and Phenomenological Research* 84 (2): 347–377. http://dx.doi.org/10.1111/j.1933-1592.2010.00490.x.

Kauppinen, Antti. Forthcoming-a. "Meaning and Happiness." *Philosophical Topics* 41 (1): 161–185. http://dx.doi.org/10.5840/philtopics20134118.

Kauppinen, Antti. Forthcoming-b. "The Narrative Calculus." In *Oxford Studies in Normative Ethics*.

Kraut, Richard. 2007. *What Is Good and Why*. Cambridge, MA: Harvard University Press.

Metz, Thaddeus. 2013. *Meaning in Life: An Analytic Study*. Oxford: Oxford University Press.

Mill, John Stuart. 1863. *Utilitarianism*. London: Parker, Son, and Bourn.

Nozick, Robert. 1974. *Anarchy, State, and Utopia*. New York, NY: Basic Books.

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

Paul, L. A. 2015. "What You Can't Expect When You're Expecting." *Res Philosophica* 92 (2): 149–170. http://dx.doi.org/10.11612/resphil.2015.92.2.1.

Wilson, Timothy D. and Daniel T. Gilbert. 2003. "Affective Forecasting." *Advances in Experimental Social Psychology* 35: 345–411. http://dx.doi.org/10.1016/S0065-2601(03)01006-2.

Acknowledgements I want to thank Daniel Star, Laurie Paul, and two anonymous referees for this journal for very useful comments and criticisms of earlier drafts of this paper.

The Self-Transformation Puzzle: On the Possibility of Radical Self-Transformation

Ryan Kemp

Abstract: In this paper, I argue that cases of radical self-transformation (cases in which an agent willfully changes a foundational element of their motivational structure) constitute an important philosophical puzzle. Though our inclination to hold people responsible for such changes suggests that we regard radical transformation as (in some sense) self-determined, it is difficult to conceive how a transformation that extends to the heart of an agent's practical life can be attributed to the agent at all. While I contend that the best way to solve this puzzle is to deny that radical transformations are in fact self-determined, many maintain the opposite. The defense of my thesis involves showing how the conditions that must be met in order to coherently attribute transformation to an agent are not satisfied in cases of radical transformation. Radical transformation is, thus, something that happens *to* an agent, not something that is done *by* her.

> *Perhaps the only answer was that by the time we understand the pattern we are in, the definition we are making for ourselves, it is too late to break out of the box. . . . To break out of it, we must make a new self. But how can the self make a new self when the selfness which it is, is the only substance from which the new self can be made?*
>
> Robert Penn Warren, *All the King's Men* (2001, 490)

1 Introduction

Among the many remarkable features of *The Brothers Karamazov* is Dostoevsky's depiction of the moral transformation of the Elder Zosima. Reflecting on his early life, Zosima characterizes it as one of "drunkenness, debauchery, and bravado," "a life of pleasure, with all the impetuousness

Res Philosophica, *Vol. 92, No. 2, April 2015, pp. 389–417*
http://dx.doi.org/10.11612/resphil.2015.92.2.11

of youth" (Dostoevsky 2002, 296). In a typical display of bravado, the young Zosima challenges an innocent man to a duel and returns home—in a "ferocious and ugly" mood—to beat his servant, something he "had had occasion to [do] before." The next morning, however, Zosima awakens to a strange feeling, "something, as it were, mean and shameful in [his] soul" (297). This general feeling of unease soon blossoms into an all-consuming sense of guilt. "Indeed," he reflects, "I am perhaps the most guilty of all, and the worst of all men in the world." This moral epiphany affects Zosima so deeply that he decides to abandon his military post and enter a monastery where he spends the rest of his life in service to the conviction that he is "guilty before everyone and for everyone" (298).

There are a few details that make Zosima's case especially interesting at a philosophical level. First, his transformation purports to be *radical*—it extends to the root of his practical life. When Zosima abandons his former values, there is a real sense in which his post-transformation self is a "*new* self." Secondly, we regard Zosima as responsible for his transformation. Unlike cases of indoctrination where there is little sense in attributing agency to the person who undergoes the change, in Zosima's case we think that his transformation is a product of a *willful choice*. Finally, and perhaps most significantly, we regard Zosima's case as *realistic*. Though personal transformation isn't an everyday affair, we acknowledge that it is a part of a recognizably human life. Additionally, we regard self-transformation as central to our moral experience. If it isn't possible for people like Zosima to transform themselves, it is difficult to understand why we fault them for their moral failures.[1]

Despite our intuition that cases like Zosima's genuinely exist, the radicalness of the purported transformation raises questions. In cases of radical change, a person doesn't simply modify some lower-order preference (like a penchant for after dinner brandy); they shed a value that so deeply anchors their practical life that its loss amounts to a kind of normative suicide. Cast in these terms, one might begin to wonder why we praise the person who purports to make such a transformation. For, if this basic sketch is correct, radical self-transformation has a lot in common with self-*betrayal*, both involving the abandonment of some central personal value.[2] Furthermore, if we think that a person's values constitute her will, then we might wonder how one could even undertake such a radical venture. Can a person really will to undermine such a central element of their volitional structure?

[1] I grant that this interpretation of Zosima's transformation is up for debate. In fact, if my thesis is correct, there are no transformation cases that meet these three criteria. I take it, however, that the kind of radical transformation that I hypothesize in Zosima's case is affirmed by both folk and philosophical intuitions as, at the very least, within the scope of possibility. The argument of the paper is noteworthy insofar as it cuts against these intuitions.

[2] This is clearer in cases where we think that a radical transformation involves a turn for the worse, for instance, the person who goes from being a dutiful father to an abusive drunk.

Not only do such cases pose a problem for our first-order intuitions, they also stand at the center of a thriving research project in contemporary ethics. Insofar as morality purports to be universal in scope, transparent in its demands, and suited to even the most "common human reason,"[3] it seems it should be possible (certainly in principle) to justify morality to any rational agent, even to characters for whom morality has no acknowledged grip. These misfits, sometimes referred to as "egoists" or cast as "Mafiosi," constitute a test case for the contemporary moral theorist who is tasked with showing how the transformation from egoism to ethics is rational.[4] To fail at this task, it seems, is to fail to establish the kind of justification that morality purports to possess; it's to admit that some agents have no reason to be moral.

In this paper, I argue that cases of *radical self-transformation* constitute an important philosophical puzzle. Though our inclination to hold a person responsible for such a transformation suggests that we regard such changes as, in some sense, self-determined, it is difficult to conceive how a transformation that extends to the heart of an agent's practical life can be attributed to the agent at all. A decision to make such a change seems to necessarily outstrip an agent's reasons, requiring her to turn against the very principles that constitute her practical outlook. While I will contend that the best way to solve this puzzle is to deny that radical transformations are *self*-determined, many philosophers maintain the opposite, that even radical transformations can be understood as actions that can be coherently attributed to the agents who undergo them. The defense of my thesis, over and against these other views, will involve showing how the conditions that must be met in order to attribute transformation to an agent are not satisfied in cases of radical transformation. In this regard, I present a challenge argument: if we cannot explain how an agent radically transforms herself, then we should conclude that she doesn't, that radical transformation is paradigmatically *heteronomous*.

In section 2, I begin by making clear what it is about cases of radical transformation that makes them problematic. In addition to explaining how the transformation puzzle falls out of a plausible view of agency, I engage with L. A. Paul's recent work on so-called "transformative experiences" (2014, 2015). Though I argue that such experiences are distinct from radical self-transformation,[5] contrasting the respective phenomena highlights what it is about the latter that is so difficult to explain. In section 3 and section 4,

[3] "Common human reason" is what Kant appeals to in order to satisfy morality's epistemic condition, that is, the requirement that moral norms be epistemically accessible to the agents they purport to apply to. By "common" [*gemein*] Kant emphasizes both the universality and ordinariness of the insight. See Kant 1996, 58.

[4] Importantly, I don't mean to say that justification depends on convincing an agent to *comply* with morality, merely that the agent can, *in principle*, be shown that morality is rational for him.

[5] I claim that radical transformation is a *species* of what Paul calls a "personally transformative experience" (2014, 16–18). I say more about this in section 2.1.

I examine two models for solving the self-transformation puzzle, what I call the "Kantian" and "Sartrean" models respectively. As I understand them, these two models mutually exhaust the explanatory options for the theorist looking to preserve agent control in cases of radical transformation. My claim is that neither model is successful in this regard. The Kantian model, represented here by Christine Korsgaard, preserves agent control at the expense of radical transformation; the Sartrean model, represented by Ruth Chang, preserves radical transformation at the expense of agent control. While I offer specific criticisms of Korsgaard's and Chang's accounts,[6] this is not my primary aim. Insofar as their more sophisticated versions of the respective models fail, the models are shown to be all the more inadequate. In section 5, I consider the further suggestion that the rationality of radical transformation is best evaluated by examining an agent's retrospective attitude toward the transformation. While this proposal moves the debate forward in certain important respects, it doesn't give a satisfying answer to why such cases should be viewed as rational to begin with. Having then argued that none of the various models satisfactorily explain how radical self-transformations take place, I conclude in section 6 that, contrary to our intuitions in such cases, radical transformations are caused by sources outside an agent's volitional structure: they are something that happens *to* a person, not something that is done *by* a person. Finally, I conclude the paper by briefly considering the all-important question: Why care that *radical* transformation can't be plausibly construed as *self*-transformation?

2 The Irrationality of Radical Self-Transformation

In his classic paper "Internal and External Reasons," Bernard Williams describes what he calls an agent's "subjective motivational set" (1981). Roughly speaking, a motivational set is the complex of desires and dispositions that constitute a person's practical point of view: things like whether a person likes chocolate, is disposed to courageous behavior, or loves the Chicago White Sox. For our purposes, I will call any change in a person's motivational set an instance of transformation. When someone goes from being a person who loves the White Sox to a person who hates them, he is transformed.

Of course, transformation understood in such general terms is not difficult to explain. I might come to hate the White Sox when I discover that they rigged the World Series, my affection for the team coming into conflict with my higher-order commitment to honesty. Alternatively, I might come to love the White Sox after undergoing a bout of brainwashing.

[6] In the case of Chang, it is especially important to note that her account is not explicitly formulated with radical transformation in mind. I apply her account of rational decision making to such cases in order to test its explanatory power in these unusually momentous cases of rational equipoise.

Cases of *radical self-transformation* are considerably more difficult to explain. These are cases where a person (1) willfully acts (2) to change a foundational element of his motivational set. An element of a person's motivational set is foundational when it cannot be justified in terms of any other members of an agent's motivational structure. In this sense, a foundational value is, for the agent who possesses it, *intrinsic*.[7] Additionally, foundational values plays a *justificatory* role vis-à-vis all the other members of an agent's motivational set. If an agent's desire (say, to misreport her taxable income) is in conflict with her foundational value (to live ethically), then that desire isn't—all-things-considered—a rational desire for the agent to have.

Just as Aristotle imagines a number of possible ends for a human life, foundational values may vary from person to person. One person might operate according to a general program of selfishness, another might be fundamentally moral, and a third might have a much more specific fundamental value, the good of their family or the prosperity of a sports team.

Harry Frankfurt offers a version of this basic framework cast in the language of higher- and lower-order desires. For Frankfurt, an action is free if and only if it is caused by a lower order desire that conforms with an agent's highest-order volition, that is, a desire for some lower-order desire to be motivationally effectual (1998). For Frankfurt, the most important higher-order volitions are an agent's "loves" and "cares." An agent is never more autonomous[8] than when he acts from his loves; he is never more heteronomous than when he acts to undermine them. While, typically, agents have many diverse loves and cares, it is compatible with Frankfurt's framework (and human psychology more generally) to suppose that, for any given agent, some one care (or perhaps a narrow set of cares[9]) has pride of

[7] Thanks to an anonymous referee for pressing me to explain foundational values in terms of their intrinsic-ness.

[8] It should be noted that Frankfurt's original goal in his early essay "Freedom of the Will and the Concept of a Person," is to establish the conditions of "free action." See Frankfurt 1998. John Martin Fischer rightly notes that the framework initially marshaled to explain freedom is later utilized (in essays like "Autonomy, Necessity, and Love"; see Frankfurt 1999) to explain autonomy. Fischer refers to this shift in focus as Frankfurt's "mission creep" (2010, 314).

[9] A person could, it seems, have a *set* of foundational values instead of a single foundational value. In such cases, the values that belong to the foundational set would be of equal normative weight, such that no individual member of the set possessed normative priority over the others. For an agent with a foundational set of values, the problem of radical self-transformation remains a puzzle. Because none of the elements of the foundational set have normative priority over any other members of the foundational set, he does not have the volitional resources to supplant any of his foundational values. I address this sort of case further in section 4.1 and section 4.2 in the context of discussing Ruth Chang's account of decision making in cases of rational equipoise.

Additionally, my argument doesn't require that all human beings have a foundation value. I take it that the puzzle I raise is philosophically interesting even if it pertains to only those agents with foundational values.

place. This care—say, love of family or duty to God or ethical uprightness—forms the volitional core of a person. It is a care that structures and guides all of an agent's practical decisions, and one that an agent betrays at the risk of abandoning her normative compass.[10]

Put in these terms, the problem of radical self-transformation comes into focus. If an agent's will is free only when it conforms with one's higher-order desires, then willing to undermine one's foundational value is paradigmatically unfree. But, of course, this is what is at stake in putative cases of radical self-transformation. In such cases, a person purportedly *chooses* to transform himself and the transformation is, we presume, creditable to more than just akratic failure.

Thus, if we accept a broadly hierarchal account of agency,[11] radical self-transformation becomes exceedingly difficult to explain. Though we pre-reflectively grant that such transformations occur, on closer inspection they appear paradoxical: they seem to be both something we do to ourselves and something that happens to us. Among contemporary philosophers few have examined this puzzle as closely as L. A. Paul, who—in her recently published book *Transformative Experience*—explores the difficulties that arise when we attempt to understand how certain "transformative experiences" can be undertaken rationally. Because Paul's interest overlaps significantly with ours, our discussion stands to benefit by considering both how she formulates the transformation puzzle and her solution to it.

[10] The view of agency recommended by Frankfurt, while not without its detractors, is not uncommon. Christine Korsgaard, J. David Velleman, and Alasdair MacIntyre all provide accounts that are compatible with this basic picture in the respect that we are interested in. For Korsgaard, a person's practical life is constituted by her "practical identities" with some identities playing a more central role than others. While a person might abandon a lower level identity (say, her identity as a chess player) at minimal normative expense, she abandons her most central identity (namely, her moral identity) on pain of "normative death." Since her identity as a moral agent is foundational, there is a distinct sense in which acts that undermine that identity also undermine her integrity as an agent. For this reason, such self-undermining acts aren't—for Korsgaard—strictly speaking actions; they do not demonstrate full agent control. See, especially, Korsgaard 2009, 159–176. For Velleman and MacIntyre it is an agent's "self-conception" and "narrative identity" respectively that limit what actions count as rational. An act that is fundamentally in conflict with one's self-conception is not a full-blooded action. See MacIntyre 1986 and Velleman 2000.

[11] What if, however, we simply reject the account of rationality that this framework is premised on? If we instead appeal to what is "objectively rational" (what, for instance, an agent with the "right kind" of values and desires should do) then, in at least some radical transformation cases, we can say that even self-undermining behavior can be rational (namely, when it leads to the formation of a disposition that is, in some relevant sense, objectively better). While this does seem to erase the major obstacle that confronts accounts of autonomy that are cashed out in terms of subjective rationality, it's not clear that an appeal to objective rationality is—all-things-considered—an improvement. Not only are such accounts contentious in their own right, it has been recently argued by philosophers like L. A. Paul that *subjective* rationality is what is properly at stake in transformation cases where we are concerned with questions of agent autonomy. See Paul 2014, 24–30 and 124–131.

2.1 L. A. Paul on Transformative Experiences

As the title of her book suggests, Paul is interested in what she calls "personally transformative experiences." An experience is personally transformative when (1) its subjective value is indeterminable prior to actually undergoing the experience, and (2) the experience stands to substantially revise a person's core preferences (2014, 16–18). Paul thinks that most people undergo a number of such experiences throughout the course of their lives and that having children for the first time is one such experience.[12]

As Paul describes them, it is not difficult to see why personally transformative experiences pose a problem for subjective rationality. In order to rationally decide between two possible outcomes it must be possible to determine, prior to the choice, the subjective value of both options, say, the value of having children versus the value of foregoing parenthood. With personally transformative experiences this is ruled out: (1) the value of the experience for one's current self is subjectively inaccessible prior to undergoing the experience and (2) the experience stands to change the agent's preference structure so thoroughly that even if one's current self knows that he will value the experience, he cannot know this about the self he stands to become.

Though Paul is clear that these cases pose a unique problem for rational decision making, she also thinks that there are contexts when deciding to undergo such experiences can be subjectively rational. If a person values undergoing transformative experiences for their own sake—or, conversely, they know that they don't value such experiences—they can rationally decide to either open themselves up to such experiences or avoid them. Paul writes: "[T]he proposed solution is that, if you are to meet the normative rational standard in cases of transformative choice, you must choose to have or to avoid transformative experiences based largely on revelation: you decide whether you want to discover how your life will unfold given the new type of experience" (2014, 120).

While clearly transformative experiences and radical self-transformation are closely related, the two notions differ in at least one important respect. When a person decides to undergo a transformational experience she is open to the possibility that such an experience will radically change who she is, as Paul puts it, to having her core preferences revised. In cases of radical self-transformation, however, things aren't so open-ended: a person explicitly decides to make a change with the precise intent of uprooting her central preferences. To put the point a bit more colorfully, transformative experiences involve a decision to *risk* normative death in order to experience something new; radical self-transformation involves a decision to *embrace* normative death at the outset. Thus, while Paul may provide

[12] The paradigmatic case is choosing to become a "vampire." Though a bit fantastic, this example has the advantage of making clear what is at stake in a personally transformative experience.

an answer to how a certain subset of transformational experiences can be undertaken rationally, her solution isn't broad enough to include radical self-transformation.

In the next two sections of the paper, I examine two accounts that model strategies one might adopt in order to establish the rationality of self-transformation. These two accounts are instructive because I take them to (1) exhaust the conceptual territory and (2) fall short of their explanatory goals. I begin with what I call the "Kantian" model.

3 Kantian Conservatism

In recent years, the question of self-transformation has received attention in mainstream ethics as part of a larger debate concerning the rationality of moral norms. Because many of the philosophers who attend to the "Why be moral?" debate utilize arguments that employ a "transcendental" method of proof, the debate itself is often associated with "neo-Kantian" ethics more generally. Figures like Alan Gewirth (1978), Thomas Nagel (1970), Christine Korsgaard (1996), and (most recently) Julia Markovits (2014) have all addressed the question in one form or another and identify themselves as broadly "Kantian." Because self-transformation plays an implicit role in this debate, I propose to revisit it and—with any luck—formulate a Kantian conception of self-transformation. Thus, my goal, in this section at least, is not to weigh-in on the moral rationalism debate, as much as to tease out the debate's implications for the transformation puzzle.[13]

While Kant famously thought that even the "most hardened scoundrel" (1996, 101) acknowledges the authority of the moral law, many philosophers since Kant have taken such characters to represent a genuine threat. If morality is rational, it must be possible to show that even egoists and scoundrels are subject to it. In order to show this, however, one must establish that such people have a reason to be moral, that there is a rational transformational path from egoism to ethics. Generally speaking, the Kantian approach to addressing this worry has been to construct a "consistency argument."[14] These arguments purport to show the egoist that her egoistic

[13] One of the paper's referees helpfully wonders: What is the relationship between radical self-transformation and the question of whether the transition from egoism to ethics is rational? In short, my claim is that a whole swath of moral philosophers—I call them "Kantians"—consider the radical self-transformation question only insofar as it touches on the latter, more traditional, question. In this section of the paper, I attempt to show that the Kantian's commitment to moral rationalism undermines her ability to account for the radicalness of a transformation. To admit that transformations can be radical is—for the Kantian—tantamount to admitting that they can be rationally discontinuous. In section 4, I examine the "Sartrean" position, a position that maintains that—to the contrary—transformations can be radical *and* rationally determinable.

[14] Alan Gewirth (1978) and Thomas Nagel (1970) are often associated with this style of argumentation. In the case of Nagel, at least, this picture may not be entirely accurate. In

desires are best satisfied when she takes the desires and reasons of others into account. If sound, such arguments show that a *rationally consistent* egoist must (upon pain of contradiction) acknowledge that she has a reason to act morally.

While critics have raised a number of worries for consistency arguments, there is one worry that is, for us, particularly salient: consistency arguments fail to show that an agent has a non-instrumental reason to be moral. On this point, consider a version of our earlier example from the introduction: Zosima's conversion from self-interestedness to selflessness. We can imagine a slightly different version of events where Zosima commits to the monastic life, not—as the story goes—after a radical change of heart, but after an earnest conversation with a concerned friend—Gregor. Deeply concerned by Zosima's moral recklessness, Gregor kindly informs Zosima that he has been going about his hedonism all wrong. The life of maximal pleasure, claims Gregor, is *actually* a life that embraces moral duty. If only Zosima devotes himself to duty and the development of a good will, all his dreams and desires will come true. He'll finally be happy! Gregor conveys all this to Zosima and, after some careful thought, Zosima concludes that this sounds nice. He decides to embark upon a life of duty.

The problem with Zosima's decision to reform is obvious: Gregor doesn't succeed in arguing Zosima over to an ethical view; he succeeds in showing him that he is currently a bad hedonist. As the argument goes, if Zosima wants to be a truly superb hedonist, one who maximizes pleasure at every step, then he needs to engage in a different set of practices—namely, ethical practices. Put in terms of our discussion, Zosima's decision to become "ethical" doesn't result in a *radical* transformation. While some of his first order desires might change, at root, he is still a person who fundamentally values his own happiness; this is, in fact, his motivation for acting ethical. Even if, through repeated play-acting, Zosima ends up becoming ethical, he will not have *chosen* to become ethical in the sense we're interested in. *Being* ethical undermines Zosima's ability to enjoy the non-moral pleasures that merely *acting* ethical provides.

Christine Korsgaard acknowledges that traditional Kantian arguments succumb to this worry (1996, 134). In response, she proposes an alternative means to the same end: egoists have a reason to be moral because the very notion of a private reason (a reason that has normative authority over only one person) is a fiction. In what remains of this section, I explore whether Korsgaard's modified framework is able to overcome the deficiency of its

The Possibility of Altruism, Nagel explicitly rejects accounts that begin by appealing to an agent's desires. Instead, Nagel thinks that the publicity of reasons follows from a certain basic conception of what it means to be a person. Though Nagel's argument does not follow the strict pattern of the "traditional consistency argument," it does function in a similar (transcendental) way. A basic account of personhood is offered, then it is made clear that one of the upshots of accepting this picture is acknowledging that one has a reason to respect the reasons of others. See, especially, Nagel 1970, 79–124.

Kantian predecessors. At stake, is the Kantian's ability to explain how personal transformation can be both radical (constitute a genuine change) and rational (self-authored).

3.1 Kantian Conservatism Updated

In lecture four of *The Sources of Normativity*,[15] Korsgaard makes a series of claims that are intended to motivate her conclusion that ethical reasons are public. She writes: "If I say to you 'Picture a yellow spot!' you will. What exactly is happening? Are you simply cooperating with me? No, because at least without a certain active resistance, you will not be able to help it," and "It is nearly impossible to hear the words of a language you know as mere noise. And this has implications for the supposed privacy of human consciousness. For it means that I can always intrude myself into your consciousness," and finally, "But why should you have to rebel against me? It is because I am a law to you. By calling out your name, I have obligated you. I have given you a reason" (1996, 139–140).

Korsgaard takes these considerations to show that other people's reasons already—and without my explicit consent—exercise authority over me; they give me "a reason to stop" (140).[16] When I hear a person speak in a language that I'm familiar with, I can't help but recognize that they are a "someone," a person whose humanity I must respect (143). Korsgaard's preferred image—a person's ability to get under one's skin—is supposed to challenge the egoist's ability to cleanly delineate himself from others. Taken in this sense, private reasons disappear because personal reasons can no longer be easily discriminated from public reasons. The reasons of others just are *my* reasons.

Korsgaard thinks that the above analysis permits an improved consistency argument to be run. It goes like this: When I encounter another person I am struck by an undeniable assurance that they are a "someone" like me. In realizing this shared humanity, I am compelled to consider how I would feel if I were treated just as I'm treating them. If I happen to be treating them unfairly, a brief consideration of their perspective makes it clear that I have an obligation to stop. If, however, I take it that my humanity merits respect (i.e., I am a law to others in virtue of it), then the humanity of others must obligate me as well. By making me consider my actions in this way, the other person forces me to acknowledge the value

[15] My analysis focuses on *The Sources of Normativity* because this earlier work more explicitly engages with the "egoism" problematic, especially as it is formulated by G. A. Cohen's now famous "Mafioso" example (which I address below). In her more recent *Self-Constitution*, Korsgaard addresses these issues in the book's penultimate chapter, "Integrity and Interaction" (2009, 196–206).

[16] This is stronger than the separate claim that other people's reasons are, in principle, publicly intelligible. While this weaker claim seems right, the stronger point about *obligation* is what Korsgaard requires for her argument to succeed.

of her humanity, and obligates me to act in a way that respects it (1996, 142–143).

What allows her argument to succeed where traditional consistency arguments fail is the purported unavoidability of taking another person's reasons as your own. Where traditional consistency arguments attempt to convince the agent that her own private reasons give her reason to acknowledge the private reasons of others, Korsgaard's argument purports to eliminate the need to appeal to an agent's private reasons by suggesting that privacy is itself a myth. Merely hearing the words of another already gives me a reason to respect the person who speaks them.[17]

Notice that Korsgaard's modified account stands to eliminate the very element that our earlier criticism of consistency arguments depended on. Insofar as traditional arguments attempt to justify the ethical life by appealing to egoistic desires, they simply promote a more subtle egoism. Korsgaard, on the other hand, refuses to play this game. Instead of descending to the egoist's level by appealing to his egoistic desires, she requires the egoist to ascend to an ethical point of view by acknowledging that he already in fact has ethical reasons.

One of the more compelling criticisms of Korsgaard's account is G. A. Cohen's now famous "Mafioso" example. In a spin-off of the traditional egoist worry, Cohen asks Korsgaard to consider an "idealized Mafioso" who "lives by a code of strength and honor that matters as much to him as some of the [moral] principles . . . he disbelieves in matter to most of us" (Cohen 1996, 183–184). If Korsgaard is right to claim that "autonomy" is the source of obligation, and that giving proper weight to one's moral identity is a necessary condition of autonomy, Cohen wonders how one explains the fact that the Mafioso leads a recognizably autonomous life with its own recognizably mafiatic obligations.

Korsgaard's response to Cohen is crucial for understanding what she takes to be at stake in cases of putative transformation. In cases like the Mafioso's, she claims:

> that there is no coherent point of view *from which* [the Mafioso's immoral form of self-identity] can be endorsed in the full light of reflection. If Cohen's Mafioso attempted to answer the question why it matters that he should be strong and in his sense hounour-bound even when he was tempted not to, he would find that its mattering depends on the value of his humanity, and . . . he would find that that commits him to the value of humanity in general, and so to giving up his role as a Mafioso. (Korsgaard 1996, 256)

[17] This analysis is meant only to capture the larger picture of Korsgaard's strategy. For a more fine grained reconstruction see Street 2012.

This response is significant because it highlights a remarkable upshot of Korsgaard's no private reasons claim: namely, that egoists and Mafiosi are *already* ethical.[18] If they reflected on their values in the full light of day, they would see that their foundational value isn't egoistic (or mafi-atic) at all. It's *moral*.

This casts the debate in a new light. Korsgaard's strategy isn't (it seems) to convert egoists, it's to show "egoists" that their central practical identity is already moral. Interestingly, this means that Korsgaard's argument (like traditional consistency arguments) doesn't aim to provoke a radical transformation. Where traditional arguments encourage egoists to dress up their egoism in ethical clothing, Korsgaard exposes the egoist by suggesting that he is involved in an elaborate masquerade. Upon removing his mask, the egoist discovers his inner ethical-self.

This has important implications for our analysis of radical transformation. If Korsgaard is right, when a person goes from being a hedonist to a person governed by ethical standards, they aren't strictly speaking undergoing a radical transformation. They are undergoing a process of *reflective equilibrium*, finally arranging their lower order desires and goals in a way that coheres with the fundamental ethical value that they have had all along. In other words, in order to preserve the ability to attribute transformation to the agent, the Kantian denies that self-transformation is radical: she either extends egoism into ethics (traditional consistency arguments) or ethics into egoism (Korsgaard).[19]

[18] One of the paper's referees rightly asks me to further defend my claim that Korsgaard takes figures like the Mafioso to be "already ethical," a claim that is potentially confused with the claim that figures like the Mafioso are *capable* of ethical action. By "already ethical" I mean that Korsgaard takes the Mafioso to already have a moral practical identity, albeit one that he denies. This moral identity is itself, according to Korsgaard, a pre-condition of having any other practical identities and is thus an identity that the Mafioso would—if he weren't self-deceived—acknowledge and respect. See Korsgaard 1996, 121.

As for the issue of ethical capacity, Korsgaard has a rather idiosyncratic view: immoral *acts* are not full-blooded (rationally endorsed) *actions*. Since an agent's foundational identity is her moral identity, immoral actions are paradigmatically self-undermining; they are instances of volitional failure. Furthermore, since Korsgaard thinks that an agent's capacity for action is contingent upon the practical identities that provide her reasons for action, an agent cannot—strictly speaking—perform a full-blooded action unless she has the accompanying identity. It's this basic agential structure that makes Korsgaard's account "conservative": because reasons only come from already present identities, no one has a reason to act "out of character." This, of course, is just another way of putting the problem of radical self-transformation. See Korsgaard 2009, 159–176.

It should also be noted that Korsgaard's account of bad action as volitional failure is not unique to her. Not only is Kant often accused of the so-called "no-bad-action problem," contemporary "constructivist" accounts seem, by their very nature, to fall victim to it. Michelle Kosch gives an excellent review of this issue as it pertains to Kant in her work *Freedom and Reason in Kant, Schelling and Kierkegaard* (2006). As for the contemporary manifestations of this problem, see Paul Katsafanas's recent book *Agency and the Foundations of Ethics: Nietzschean Constitutivism* (2013).

[19] It is important to emphasize that Korsgaard's account prevents radical transformation not just because moral reasons are available to the Mafioso prior to his conversion, but rather

The claim that all agents are *already* moral has implications for all putative transformation cases, not just those that involve a transformation from egoism to ethics. Consider, for instance, the case of a person who transforms from art devotee to devotee of physical pleasure. If Korsgaard is right, the art devotee isn't fundamentally devoted to art, she is fundamentally devoted to morality—her moral identity is, in fact, her central practical identity. This is why the "Kantian" position allows us to draw broader conclusions about all instances of radical transformation. All putative radical transformations are either just lower-level changes in an agent's motivational structure (as in the case of the art devotee) or mere reformations, that is, a return to an ethical self that has been present all along (as with the Mafioso).

3.2 Considering the Kantian Alternative

Clearly, a lot hangs on whether Korsgaard's analysis is right. If agents like Zosima are already moral, then we have an answer to the self-transformation puzzle: transformations are self-determined precisely because they aren't radical.[20] Is this right, though?

To begin, it's important to note that Korsgaard thinks that such discussions should be cashed out in roughly "internalist" terms. In both *Sources* and *Self-Constitution*, she is interested in answering what she calls the "normative question": "a first-person question that arises for the moral agent who must actually do what morality says" (1996, 16). According to Korsgaard, the "substantive moral realist"[21] fails to answer the normative question because—when confronted by the skeptic—he can do nothing more than "dig in his heals" and insist that there are, as a matter of fact, stance independent moral norms. The "constructivist" answer purports to

---

because—according to Korsgaard—the Mafioso is *fundamentally moral* prior to "conversion." Because Korsgaard thinks that a person's reasons are a product of her practical identities, if it were the case that the Mafioso's "mafiatic" identity were more central than his moral identity, then—by that very fact—he would have an all-things-considered reason to continue his life as a Mafioso. Since, however, Korsgaard wants to deny this, she must claim that, in fact, the Mafioso's fundamental identity is moral.

One of the paper's reviewers rightly remarks that even if the Mafioso's reformation merely involves attaining motivational coherence (that is, finally honoring values that he has had all along), then such a reformation is still "radical." This sense of "radical," however (meaning something like, "remarkable" or "profound"), is importantly different than the sense that I am interested in, namely "radical" as pertaining to a thing's "root."

[20] Thanks to a referee for pushing me to make this explicit.

[21] Korsgaard defines moral realism as: "the view that propositions employing moral concepts may have truth values because moral concepts describe or refer to normative entities or facts that exist independently of those concepts themselves" (2008, 302). She summarizes realism's normative failings as follows: "If someone finds that the bare fact that something is his duty does not move him to action, and asks what possible motive he has for doing it, it does not help to tell him that the fact that it is his duty just is the motive. The fact isn't motivating him just now, and therein lies the problem [of realism]" (1996, 38).

be different insofar as it gains internal leverage; it appeals to certain undeniable features of agency that commit an agent to morality. For Korsgaard, the answer to the question, "Why be moral?" is just, "Because you *value* morality!" "Valuing" is central to the constructivist account because it satisfies a motivational version of the "open question argument." If it can be shown that a person *values* some thing *x*, there is no further question concerning whether they also have a *reason* to promote *x*.[22]

If Korsgaard is right, some agents value morality without having any notion that they do. While this may be true (surely some agents are self-deceived about what they value), one wonders how far this can go before it itself becomes an instance of what Korsgaard is trying to avoid—a case of empty heel digging. What is the difference, for instance, between Korsgaard's claim that the moral realist simply insists that we have a reason to be moral and Korsgaard's own insistence that our supreme practical identity is moral? For all we know, hidden in the depths of every human being may lurk the "value" that Korsgaard describes. What seems unbelievable, however, is that this "value"—the so-called value of humanity—is central to the self-conception of the Mafioso or the Zosima of our earlier example. What would it be like to value something supremely while being oblivious to its presence and authority? Whatever this kind of "valuing" amounts to, it appears utterly foreign to discussions of agent autonomy. An inner voice that is "incapable" of making itself heard is, for all intents and purposes, no voice at all.[23]

In the end, the updated Kantian answer to the transformation puzzle depends on the same move that we criticized traditional consistency arguments for making: it assumes that there isn't a *radical* change in agency when a person goes from being egoistic to ethical (or anything else). While traditional arguments attempt to "convert" egoists by appealing to egoistic reasons, the Kantian denies that non-ethical agents (agents who do not

[22] Sharon Street's (2012) account makes this especially clear.

[23] In her 1986 essay "Skepticism about Practical Reason," Korsgaard offers a surprisingly liberal (internalist) account of the conditions in which an agent has a reason. I quote at length:

> 'Available to us' is vague, for there is a range of cases in which one might be uncertain whether or not to say that a reason was available to us. For instance there are (1) cases in which we don't know about the reason, (2) cases in which we couldn't possibly know about the reason, (3) cases in which we deceive ourselves about the reason, (4) cases in which some physical or psychological conditions makes us unable to see the reason; and (5) cases in which some physical or psychological condition makes us fail to respond to the reason, even though in some sense we look it right in the eye. . . . For toward the end of the list we will come to claim that someone is psychologically incapable of responding to the reason, and yet that it is internal: capable of motivating a rational person. I do not think there is a problem about any of these cases; for all that is necessary for the reason claim to be internal is that we can say that, if a person did know and *if nothing were interfering with her rationality*, she would respond accordingly. (1986, 13–14)

fundamentally value morality) even exist. This means that all cases of apparent radical transformation are really just cases of *radical reformation*, cases in which an agent comes to endorse the values that have formed her normative core all along. While it is possible that the Kantian is right about this, it requires her to offer an exceedingly implausible account of agents like Zosima, who—by their own lights—do not acknowledge any overriding commitment to moral norms. Even if such agents are capable of acting on moral imperatives, they don't take such imperatives to be categorical.

In the face of the Kantian model's explanatory deficiency, the paper's next section turns to consider the merits of a so-called "Sartrean" model of transformation. Unlike the Kantian, the Sartrean is intent to explain how self-transformation can be *radical*.

4 Sartrean Radicalism

In his essay "Existentialism is a Humanism," Jean-Paul Sartre defends the view that human beings are capable of radical self-transformation. In the essay's most famous example, Sartre describes the plight of a young student faced with two equally compelling choices: *either* avenging his brother's death by serving in the French Resistance *or* staying home to care for his grieving mother. What makes this case especially momentous is that in choosing, the young man will—in a very real sense—be deciding what kind of person he is. Will he be a person for whom honor and national pride take priority *or* someone for whom politics takes second seat to the personal demands of family? In this respect, his choice doesn't just say something about who he is, it *determines* who he is by establishing his value hierarchy.[24]

What makes Sartre's analysis[25] controversial is that he thinks that contexts in which one's reasons underdetermine the proper course of action are precisely the contexts in which a person exercises maximal freedom. Thus, instead of looking at the student's situation as a challenge to his autonomy—a kind of tragic dilemma—Sartre sees the situation as autonomy affirming, as a moment in which the student is capable of transcending his received identities by performing an action that is underdetermined by his reasons. It is precisely in this rational under-determination that true freedom lies.

[24] This, at least, is Sartre's recommended analysis.

[25] It should be kept in mind that by "Sartre's analysis" I just mean the analysis found in "Existentialism is a Humanism" (2007). As a point of historical fact, Sartre's earlier *Being and Nothingness* offers a much more balanced account of autonomy. In the famous section on "bad faith," Sartre acknowledges that ignoring one's facticity (i.e., one's received identities) is another instance of bad faith (1993, 86–118). The "Sartre" of this paper is thus something of a caricature, albeit one that is commonly made use of in the secondary literature. For an excellent account of Sartre's broader account of freedom, and certain problems thought to plague it, see Wilkerson 2010.

While the motivation for Sartre's position is understandable, his analysis appears incapable of explaining how radical transformation can be attributed to the person who undergoes it. How, for instance, does the mere exercise of a formal power of freedom (the freedom to transcend my given practical identities) constitute an action that is coherently attributable to me, the person of flesh and blood who has three children and roots for the White Sox? This criticism is, of course, familiar: it's a worry that Hume raises in the *Treatise* (2000, 257–264) and Frankfurt revives in his now classic essay, "Freedom of the Will and the Concept of a Person" (1998). The common concern of both Hume and Frankfurt is to show that attributing an action to someone is contingent upon being able to tell a story about how the action follows from the dispositions and desires of the agent. In Sartre's case, the paradigmatically autonomous action is precisely the one that has no such backstory; it's an action that doesn't follow from an agent's antecedent reasons. Insofar as Sartre fails to explain how acts of radical self-transformation are actually *self*-determined, he reverses the Kantian mistake: he compromises agent control for the sake of radicalness.

4.1 Updated Sartreanism

In a recent series of essays,[26] Ruth Chang defends a view ("hybrid-voluntarism") that looks to preserve elements of Sartre's account that are explanatorily important, while leaving behind elements that are explanatorily improbable. Chang is interested in cases like the young student's where a person does not have sufficient reason to choose one alternative over the other. In such cases, Chang thinks that agents have the ability to tip the normative scale by willing that some consideration in favor of one of the alternatives becomes a *reason* for one of the alternatives. Chang writes:

> This willing a consideration to be a reason is . . . akin to a *stipulation* that something be a reason in much the way that you might 'stipulate' that your newborn be called 'Winston'. But when you will something to be a reason, something beyond mere stipulation is involved: your agency is implicated. Very roughly, when you will something to be a reason, you *put yourself* behind some consideration that, as a logical matter, counts in favor of one of the alternatives. (2013, 180)

Thus, by throwing one's volitional weight behind such a consideration, Chang thinks that a decision at hand can evolve from being one of rational impasse to one that is determined by reasons.

Chang is careful to distinguish her solution from Sartre's. While Chang thinks that rolling the existential dice may sometimes be justified, she is interested in "rational determination": cases where an agent has most

[26] See, for instance, Chang 2009, 2013.

reason to choose one option over the other. She differs from Sartre in that she thinks that situations which initially admit of equipoise can, through an act of will, become situations that are determined by reasons. One needn't, Chang thinks, join Sartre in advocating a "plump" model of decision making where an agent arbitrarily throws her will behind a rationally under-determined course of action. Additionally, Chang is also clear that the ability to voluntarily create reasons is dependent on there being reasons that aren't open to volitional control. This is what makes her voluntarism "hybrid." It's only in cases where an agent's "given" reasons "give out" that an agent is capable of creating new reasons.[27]

4.2 Evaluating Chang's Solution

While initially it may seem as though cases of radical self-transformation are distinct from the cases that interest Chang, we can imagine examples that fit her analysis, for instance, cases where a person who fundamentally values *x* enters into a disagreement with an epistemic peer who fundamentally values *y*. Consider the case of Walter and Adam. In August, Walter will have been happily married to Allison for three years. This comes as a bit of a surprise to Walter's close friends because, before marrying Allison, Walter was an inveterate "aesthete"[28]: his life was entirely devoted to seeking-out one interesting experience after another. New films, new books, new restaurants, new romantic partners . . . you get the picture. One day (as the story goes), Walter meets Allison and falls head-over-heels in love. In an incredible twist of fate, he decides that he wants to get married, to commit to Allison, through sickness and health, for the rest of his life.[29] So, Walter marries and—as his friends can attest—becomes a radically different person. He settles into quiet domesticity as though it had been his life-long goal.

Well, this all catches Walter's friends off-guard. One friend, Adam—a fellow aesthete who feels especially betrayed by Walter's transformation—decides to confront Walter. He and Walter go out for a drink and Adam unleashes his frustration. "Walter, what are you doing? How can you possibly think that being married *to one person*, for your *entire life*, is the way to go? This is just your religious upbringing rearing its ugly head!" Walter, to Adam's great surprise, begins to laugh. "Adam, you have it all wrong, being married to Allison is the best thing that has ever happened to

[27] Chang writes, "Metaphysically speaking, if we have the freedom to create reasons, we have it only within the confines of the reasons we have no freedom to create, our given reasons" (2013, 179).

[28] This terms is borrowed from Kierkegaard's *Either/Or*. There, Kierkegaard calls a life that is oriented toward pleasure an "aesthetic" life. Much of this paper's debate is, in fact, indebted to various ways of interpreting Kierkegaard's classic text.

[29] Importantly, my analysis of Walter's conversion places emphasis on the act of "falling in love." This is the moment of conversion, albeit one that Walter subsequently affirms in his decision to get married.

me. I finally feel like I have something to live for. I'm completely at peace with it. You *should* get married too."

Adam comes away from the conversation a bit confused. He wasn't lying to Walter when he said that he can't imagine a better life than the one he currently lives. On the other hand, though, Walter used to say similar things, and now look at him, apparently living a very happy life. Even worse, a *happier* life than before.

I suspect that Adam could, after such a conversation, reach a place of rational equipoise vis-à-vis his current foundational value (being free of liberty-eroding personal commitments) and Walter's foundational value (the value of having and supporting a family). In spite of the fact that Adam is fundamentally *aesthetic*, he may actually have *as much* reason to abandon that value as to preserve it. This follows from the fact that his epistemic peer—Walter—reports that his new life is significantly better than the old life he shared with Adam. Though Adam isn't in a position to make a similar judgment (Walter's life even strikes him as a bit repugnant), he finds that Walter's post-transformation conviction has forced him to, at the very least, be open to the possibility that his current modus operandi is short sighted. This possibility is significant, because—as we discussed above—Chang thinks that cases of rational equipoise are rationally determinable. That is, an agent who has equal reason to prefer two distinct options might, through a rational decision making process, come to have a reason to prefer one to the other. If this is right, then we will have a case in which a radical transformation can be autonomously performed. Insofar as the transformation follows from what an agent has most reason to do, the transformation can be said to be a coherent manifestation of the agent's will. We could finally have a case of radical *self*-transformation.

So, how, according to Chang, is Adam's decision supposed to be settled *rationally*? Since Adam sincerely wants to choose between the two options and doesn't want to be a mere existential plumper, he follows Chang's advice and takes one consideration in favor of his aesthetic lifestyle (that he gets to diversify his romantic partners) and through an act of stipulation confers upon this consideration the status of a reason. Now, as Chang's story goes, Adam has an all-things-considered reason to remain aesthetic. Mission accomplished!

There are a few worrisome features to this account. First, assuming that the two original options are rationally equivalent,[30] what reason could Adam have to favor aesthetic considerations over ethical[31] ones? The only possible explanation is that Adam has some further "tie-breaking" desire, and that this is what leads him to choose aestheticism. This, however, can't be the explanation: any desire he has to be aesthetic is, in the initial

[30] Recall that we are interested in subjective rationality, thus our analysis gives preference to considerations that an agent takes to count in favor of a proposed course of action.

[31] For ease of reference, I will refer to Walter's foundational value as "ethical."

deliberation, *already* viewed as a reason in favor of an aesthetic life.[32] He precisely finds himself in this dilemma because the considerations in favor of one option (including all relevant desires) aren't obviously better than the considerations in favor of the alternative. Thus, Chang's solution—that he throw his weight behind one of the two options—seems to avoid Sartrean plumping at one level (the point where the agent decides to x) only by endorsing plumping at another level (the point where the agent decides to throw his weight behind a consideration in favor of x).

A second and similar worry arises a bit further down the deliberative stream when Adam chooses to throw his weight behind a particular pro-aesthetic consideration. Assuming that all the considerations in favor of being aesthetic have equal rational weight,[33] the choice to promote any one consideration among the several possible also begins to look like a kind of plumping. Consider Adam's case again. After deciding for no apparent reason to favor aestheticism over a more conventional ethical life, Adam goes on to favor a particular aesthetic consideration over other aesthetic considerations: for instance, throwing his weight behind the consideration that aesthetes get to meet a lot of diverse and interesting people instead of the equally worthy consideration that he will never have to change diapers. But without the means of rationally adjudicating between these two considerations, this choice doesn't appear to exclude the arbitrariness that Chang's account is designed to avoid. It is just another case of plumping.

Perhaps, though, the considerations in favor of being aesthetic are not all of equal rational weight. The fact that being aesthetic will satisfy Adam's desire to become a broad-minded citizen of the age is surely more significant (even by his own lights) than the fact that being aesthetic will satisfy his desire to score cool-points with the teenager who works the register at the local Stop-n-Go. But this, of course, raises a further and considerably deeper worry for Chang's account. When Adam stipulates that his relationship to the Stop-n-Go attendant will act as the rational tie breaker, he is doing something patently *irrational*. The reason we can't just plump for any consideration whatsoever is precisely the reason we can't will reasons to begin with: our ability to take something as a reason to x is

[32] There is a way of reading this problem as one of "double counting." Adam begins with a reason to be aesthetic (that he will have many and various romantic partners) and then, through an act of will, chooses to throw his weight behind this consideration in order to give it even further normative force. Chang is aware of this worry and denies that she is vulnerable to it. She argues that a reason is not double counted solely in virtue of the fact that it shares its content with another reason. In defense of her account, she suggests that two reasons with two separate sources can share the same content while remaining distinct. While I am not convinced that Chang's solution does avoid double counting worries, my concern here is with whether she avoids the kind of "plumping" she attributes to existential views. See Chang 2009, 257.

[33] Though this seems unlikely, it also seems necessary if an agent's ability to endorse *any* consideration at all is going to make sense.

dependent on our ability to recognize that thing as *appropriately* counting in favor of *x*-ing. In the situations we're concerned with, cases where an agent has no all-things-considered reason to do one thing or another, we cannot simply appeal to any run-of-the-mill pro-consideration in order to break the tie. One needs a *reason* to think, for instance, that the situation at the Stop-n-Go is relevant in a way that being a broad-minded citizen isn't. Unless such a reason can be appealed to, the decision to prefer a relatively insignificant pro-consideration to a relatively significant one is unmotivated. At best, it's a kind of plumping; at worst, it's irrational.

Though Chang anticipates this worry, she doesn't seem to appreciate the deep challenge it poses for her account. She admits that, "The voluntarist reasons you create are arbitrary in the sense that there are no reasons for you to have created those reasons rather than others." She counters, though, that to be "*objectionably* arbitrary the reasons you create should be ones that lead to substantively objectionable conclusions about what you have all-things-considered most or sufficient reasons to do" (2009, 269). In other words, as long as "created reasons" don't give you an all-things-considered reason to do something that wasn't rationally objectionable prior to the act of creation, they aren't objectionably arbitrary. This, however, seems to miss the point. Chang's recommendation is analogous to a theoretical case where we grant that a person is free to believe whatever she wants as long as those new beliefs don't radically revise her previous belief structure. The problem with this response is that beliefs—like reasons—are not the sort of thing that can be created on demand. To the contrary, they are responsive to an agent's beliefs about the relationship between her ends and the means to those ends. Thus, it isn't at all clear that Chang's modified "voluntarism" offers a satisfying solution to our initial Sartrean worry. In the same way that counterfeited money doesn't pay down a debt, stipulated reasons don't make a decision rational. Both solutions avoid a problem at one level by creating a problem at another.

We've now looked at two models that offer distinct takes on the self-transformation puzzle. The Kantian model, intent to explain how norms can be both a reflection of an agent's will and universally obligating, sacrifices the radicalness of self-transformation in order to preserve agent control. This ends up being a problem for the Kantian because it commits her to the unlikely (and heavy-handed) position that agents that don't acknowledge any ethical commitment are, nonetheless, fundamentally ethical at heart. The Sartrean, on the other hand, sensitive to the limits of the Kantian account and intent to explain self-transcendence, sacrifices agent control for the sake of capturing the radicalness condition. For the Sartrean, the explanatory difficulty arises insofar as he maintains that this act of self-transcendence is nonetheless autonomous (or in Chang's case, rationally determinable). Insofar as acts of will that aren't accompanied by reasons (Sartre) and acts of will that are accompanied by *willed* reasons (Chang) fail to overcome the arbitrariness worry, Sartrean accounts are

unable to explain how meaningful agent control is genuinely exercised in cases of radical self-transformation.

One upshot of the Kantian and Sartrean models' respective failures is that it becomes difficult to see how the demands of morality can be unproblematically cashed out as demands of rationality. In a recent paper, philosopher Michael Cholbi acknowledges this worry and responds by recommending a putative third solution to our puzzle: that radical transformation should be evaluated according to "its *retrospective* reasonableness to the converted" (2011, 547, emphasis added).[34] This, argues Cholbi, is a way to admit to the limitations of the Kantian and Sartrean models, without also conceding moral rationalism. Since Cholbi's account is premised on the failure of the two dominant explanatory models, his solution bears consideration.

5 A Third Way? Retrospective Reasonableness

Cholbi advocates what he calls a "Kuhnian" model[35] of transformation. He explains, "Perhaps the mechanism of the egoist's conversion is the metaphorical 'flipping of the gestalt switch.' Perhaps Aristotle was correct that it is by the performance of virtuous acts that an agent comes to be virtuous and to appreciate virtuous actions for its own sake" (2011, 546). On this model, an egoistic agent is slowly habituated to virtue by *acting like* a virtuous person, mimicking the latter's behavior in spite of the fact that he doesn't yet identify virtue as something he has an all-things-considered reason to pursue. Eventually, after sufficient habituation, the vicious person acquires a new set of *virtuous* dispositions and, with it, the reasons that follow.[36]

[34] The suggestion that Cholbi's account provides a "third way" may seem to cut against my earlier suggestion that the Kantian and Sartrean positions mutually exhaust the explanatory options for the theorist looking to preserve agent control. As I will soon argue, Cholbi doesn't actually constitute a genuine third option insofar as his construal of the "rationality condition" is divorced from questions of agent attribution.

[35] Interestingly, ethical conversion doesn't include the one element that would make the comparison between it and scientific models apt. Cholbi writes: "[I]n theoretical paradigm shifts in the sciences, the rationality of the paradigm shift can be explained by appeal to explanatory or theoretical vales (accuracy, simplicity, fertility, and so on), even if there is no algorithm or uncontroversial rank ordering of these values that would decide the controversy between two rival theories. But it is less clear what could play the role of these theoretical values in justifying the conversion from rational egoism to a moral way of life" (2011, 553). Though this worry is initially raised as a hypothetical objection to his account, Cholbi acknowledges that he does not have a satisfying response.

[36] In *After Virtue*, Alasdair MacIntyre provides an example that nicely illustrates what Cholbi has in mind. The case involves teaching chess to a child who doesn't yet appreciate chess's internal goods. In order to motivate the endeavor, the child is offered candy to play and even more candy if he wins. Eventually, after playing many candy motivated matches, MacIntyre hypothesizes that the child might come to appreciate chess for its own sake; he will have been *transformed* into a chess player. See MacIntyre 2007, 188. As I argue below, the

While Cholbi's model has the advantage of being empirically informed,[37] it isn't yet clear how the sort of transformation he describes counts as rational in the sense we are interested in. How, for instance, does an agent who is not yet capable of appreciating the goods of a virtuous life come to recognize that he wants to reform himself into a virtuous person? The key, thinks Cholbi, is "retrospective evaluation." He writes:

> [A radical transformation's] rationality does not depend on the causal mechanism of belief change, but on its retrospective reasonableness to the converted, on her rationally informed willingness to avow the new belief even in the face of defense of the old belief. Hence, agents undergoing conversion do not even need to be aware that they are undergoing a cognitive shift as the conversion occurs. All that is required in order for egoist conversion to count in favor of moral rationalism is that the conversion provide the egoist with an appreciation of moral reasons. (2011, 547)

According to Cholbi, it doesn't matter how or why the ethical transformation occurs as long as the agent approves of his new self. If the converted agent can honestly say that his current self is better than his former self, his transformation counts as *rational*.

There are several potential problems with Cholbi's recommendation. The first concerns the way his appeal to retrospective evaluation shifts the discussion away from an evaluation of *prospective* transformation. Recall that we're interested in the rationality of conversion because we want to explain how transformation can be autonomous, that is, how transformations *follow from* an agent's reasons. Appealing to an agent's post-conversion evaluation doesn't help us gain purchase on this question.

Another problem with Cholbi's recommendation is that it threatens to make *all* radical transformations rational transformations. Recall, that in cases of radical transformation it is an agent's foundational value that is at stake and that this value *constitutes* what an agent takes to be good. This means that an agent is never in a position to give a negative all-things-considered evaluation of his foundational value. This kind of critical stance

transformation from lover of candy to lover of chess is not rational because the child doesn't engage in the exercise with the intent of becoming a chess player. It might even terrify the child to think that repeated chess playing may lead to a state of affairs where he will enjoy playing chess *sans* candy. This case is structurally similar to our earlier Zosima and Gregor case. Choosing to act ethical in order to satisfy hedonistic desires, and then subsequently becoming ethical through habituation, isn't an instance of choosing to become ethical.

[37] One of the great virtues of Cholbi's account is its interaction with the psychology literature. One especially relevant study is conducted by Jonathan Haidt, who concludes that transformations in moral outlook typically involve non-inferential transformations in belief. If deductive arguments play any justificatory role, it's *post hoc*. See Haidt 2001. Other relevant studies include Turiel et al. 1987 and Lieberman 2000.

would require an agent to simultaneously confess that some thing *x* has ultimate value for him *and* that, all in all, he wishes it didn't. While it seems compatible with desiring *x* (or perhaps even *prima facie* valuing *x*) that a person might also wish that he didn't desire (or value it), it's more difficult to make sense of how an agent can consistently value something *ultimately* and still—all things considered—desire that he didn't. This is significant for our evaluation of Cholbi's recommendation because it suggests that whenever a radical transformation occurs, and one foundational value is substituted for another, the agent will have a favorable view of the newly adopted value. This, coupled with Cholbi's suggestion that retrospective approval is sufficient to call a transformation rational, suggests that all radical transformations will count as "rational," this ubiquity taking much of the punch out of an achievement term typically tasked with doing considerable explanatory lifting.

Finally, in a similar vein, even if we grant that ethics is rational in the way that Cholbi defends, this is a far cry from what would have to be shown in order to defend what is traditionally called "moral rationalism." In keeping with the roughly internalist assumptions our debate is premised on, Cholbi agrees that conventional accounts of radical transformation struggle to explain how conversion can be rational when measured against the pre-transformation elements of an agent's motivational set. His response is to judge a transformation's rationality against an agent's *post*-transformation motivational set. However, this sense of rational isn't sufficiently robust to support moral rationalism, a thesis about what it is *objectively* rational to do.[38] In this regard, the weakness of Cholbi's solution is best brought out by considering that a transformation from morality to egoism stands just as much a chance of being "rational" (in his sense) as the transformation from egoism to morality.

Though Cholbi doesn't explain how radical transformations are rational in the sense we're concerned with, his account does move our discussion forward in one very important respect. He reminds us that, as a point of empirical fact,[39] the actual mechanism of radical transformation isn't a person's reasons. Radical transformations occur when some external influence interrupts an agent's practical life.

6 Conversion by "Inspiration"

In the paper's introduction, I suggest that our intuitions (both lay and philosophical) indicate that radical self-transformations sometimes occur. Many of us are willing to grant, for instance, that a selfish bachelor can, independently of any intervening factors, choose to become a moral ascetic.

[38] It should be clear that I'm not suggesting that internalist frameworks are necessarily incapable of accounting for objective rationality. To the contrary, Christine Korsgaard (1996; 2009) and (more recently) Julia Markovitz (2014) are interested in precisely this.

[39] See footnote 38.

Furthermore, some philosophers take such cases to be central to how we think about the scope and claims of morality. In order for morality to be everything that it claims for itself, agents like Zosima must have a reason to be moral; it can't be the case that choosing to be moral is (for any agent) a quintessentially self-undermining act.

Following up on these intuitions, we have examined three different ways in which to model transformation such that it counts as both *radical* and *autonomous*. Each model was deemed insufficient in some important respect: the Kantian model denied radical transformation all together, while the Sartrean and (Cholbi's) Kuhnian models failed to show how radical transformation can be credibly understood as self-authored. If we remain committed to the claim that radical transformations occur (as I think we should), then it is clear that we need an alternative explanation of how they come to be. If they aren't self-authored, then—it seems—they must have their source in factors that are external to an agent's volitional structure.[40] This external interruption is paradigmatically *heteronomous* because it provokes a change that the agent does not have the resources to rationally enact prior to the encounter. In this sense, radical transformations are *a-rational*.[41]

Consider again the two examples we explored above. In the case of Zosima, it seems we should conclude one of two things: either Zosima doesn't undergo a radical transformation—in which case his change is a result of finally giving sufficient weight to a fundamental value that he first develops as a child and has, for most of his adult life, ignored[42]—or that his radical transformation is caused by the traumatic experience of his own cruelty toward his servant. Such an event, combined with implicit

[40] To be clear, I arrive at this conclusion by an argument from elimination:

(1) Radical transformations occur.
(2) They are either self-authored or caused by an external source.
(3) They are not self-authored.
(4) They are caused by an external source.

[41] It is important to remember that the standard of rationality that I'm appealing to is an *internal* one. If it is true, as Philippa Foot thought, that "acting morally is part of practical rationality" (2001, 9), I am willing to concede that a radical transformation that doesn't follow from an agent's desires and values can still count as rational. In this case, a transformation caused by a traumatic head injury might count as rational if it transforms a person from a morally apathetic person to a morally earnest one. The problem is that this incredibly substantive account of practical rationality is, in itself, highly controversial. Furthermore, as I point out in footnote 11, L. A. Paul has recently made a persuasive case for preferring accounts of subjective rationality in situations where the rationality of a person's life decisions are concerned. This follows, she thinks, largely because we are in such cases interested in whether a person exercises agent autonomy.

[42] Within the novel itself, there is good reason to suspect that this is what is really going on. Zosima's realization that he is "guilty before everyone and for everyone" is a reiteration of a claim that impresses Zosima as child, a claim made by his brother on the latter's death bed. If this is the right analysis of Zosima's case, then Zosima can be said to undergo a "reformation" (that is, a return to a foundational value) not a conversion.

insecurities about his more general program of cruelty, seems sufficient to precipitate a radical reorientation of his value set.

A similar analysis can be given in the case of Walter. If Walter genuinely undergoes a radical transformation, we should attribute it not to his decision to get married, but to his life-changing encounter with Allison. Like Zosima's experience of his own cruelty, Walter's experience of a healthy and rewarding relationship radically challenges his deep-seated selfishness.

One interesting question this analysis raises is why some encounters spur radical transformation, while other similar encounters don't.[43] One possible explanation that I hint at above, is that radical transformation depends, in some way, on already present desires—even if alienated—in a person's motivational set. Imagine that pre-transformation Walter actually has a first-order desire to be in a long-term romantic relationship. Though this desire clashes with his higher-order value to remain commitment-free, its presence may nonetheless play an enabling role in his transformation. This is significant because it suggests that some background conditions (like already present desires) may play a necessary role in the radical transformation process, providing a normative foothold for radical transformations to occur.[44] These lower-order desires may also provide an evaluative link that establishes a connection between an agent's pre-transformational and post-transformational selves.

7 Conclusion: Some Implications

Practically speaking, the a-rationality of radical transformation is important insofar as it suggests that our foundational values are, in a very deep sense, *contingent*. Not only is it the case that what is rational for me to value may not be what is rational for you to value, what is rational for my current self to value may not be what is rational for a close possible self to value. Like Walter's introduction to Allison, our deepest value commitments—commitments that we make great personal sacrifices for—hinge on chance experiences. This realization stands to be destabilizing insofar as the attitude of "valuing" seems to involve acknowledging an object's non-relative "goodness." When, for instance, I claim that some thing x is valuable, I seem to also be making a claim about how other people should relate to x—namely, they too *should* take x to be good.

This worry is brought out when we consider the structural similarity certain arguments against religious commitment have to arguments that

[43] Thanks to an anonymous referee for bringing this possibility to my attention.

[44] It is important to note that even if radical transformation depends on already present elements of an agent's motivation set, this does not make the transformation any more rational. For pre-transformation Walter, becoming a person who is defined by a commitment to a life-long sacrificial relationship is irrational regardless of whether he has a first-order desire to enter into such a relationship.

can be brought against ethical commitment[45] more generally. As Philip Kitcher's most recent version of the argument goes, the very fact that a person's religious values would have been different had she been exposed to different cultural influences, is sufficient reason for that person to give up her commitment to those particular values (Kitcher 2014, 8). This follows from the fact that the variables that led to the formation of her commitments are not truth conducive. However, as Alvin Plantinga (2015) has recently argued, it seems like Western democratic values are vulnerable to the same kind of attack. If Philip Kitcher, for instance, had been raised as an Australian aborigine, then he would likely have a significantly different value system than the one he currently possesses. Perhaps *he* should consider abandoning his value commitments. Though Plantinga makes a different inference, namely that since Kitcher's ethical commitments are unobjectionable Plantinga's religious commitments are too, it's not clear why we shouldn't conclude that *both* sets of values are on shaky rational ground. Kitcher's original claim still packs a normative punch.

This issue isn't just the concern of philosophers either. In Karl Ove Knausgaard's acclaimed memoir *My Struggle*, we see how the perceived contingency of a person's value commitments can lead to a kind of value-nihilism. In the memoir's second volume, Knausgaard writes:

> Everyday life, with its duties and routines, was something I endured, not a thing I enjoyed, nor something that was meaningful or that made me happy. This had nothing to do with a lack of desire to wash floors or change diapers but rather something more fundamental: the life around me was not meaningful. I always longed to be away from it. So the life I led was not my own. I tried to make it mine, *this was my struggle*, because of course I wanted it, but I failed, the longing for something else undermined all my efforts. (Knausgaard 2013, 67)

In later passages, Knausgaard confirms that his longing for "something else" is precipitated by the knowledge that, if events had been slightly different, he could very easily be living a different life. Certain chance happenings, like falling in love or the birth of his children, come together to form the core of his identity and commitments, but—for all that—still feel accidental. The lives that *could have been* continually haunt the life that is.

The debate between Kitcher and Plantinga, as well as the crisis that Knausgaard speaks to, are an indication of the perceived seriousness (both at a philosophical and a cultural level) of value contingency. The fact that we are products of transformative experiences that we exercise little control over, does (and, if Kitcher is right, *should*) affect the way we relate to our values. The question is *how* should it affect us. Should we retreat from

[45] By "ethical," I mean non-religious value commitments. In this regard, I count Walter's commitment to his marriage an ethical commitment.

our ethical commitments in the way that Kitcher hopes religious people should retreat from theirs? Should we stand firm and confident as Plantinga recommends? Both of these options seem, in various ways, misguided. For instance, it's not at all clear that an attitude of "valuing" is conceptually wed to any accompanying epistemic attitude. This means that a person might simultaneously acknowledge that he has no good epistemic reason to think some thing x is a non-relative good, while still coherently valuing it. This is significant because it allows a person to take seriously the kinds of epistemic worries Kitcher raises while remaining earnestly committed to his values in the way that Plantinga encourages. Changing the traditional sense and context of the attitude, we might call this orientation a kind of practical "faith."

For many of us, our attitude toward the things we value (moral or otherwise) can and should be one of "faith." When we come to appreciate that our foundational values are contingent, we aren't suddenly liberated from the normative hold they exercise over us. We are, however, prompted to view them with a kind of humility, to appreciate that even the things we take to be most dear are, in a certain respect, tenuous. While for some philosophers this necessarily marks a move toward value-nihilism, others in the Western tradition have suggested that this can, at least potentially, be a catalyst for ethical passion.[46] We step out into the moral landscape with value commitments that we can neither deny nor justify, and this—perhaps counterintuitively—is the beginning of true responsibility.

Ryan Kemp
E-mail : ryan.kemp@wheaton.edu

References:

Chang, Ruth. 2009. "Voluntarist Reasons and the Sources of Normativity." In *Reasons for Action*, edited by David Sobel and Steven Wall, 243–271. New York, NY: Cambridge University Press.

Chang, Ruth. 2013. "Grounding Practical Normativity: Going Hybrid." *Philosophical Studies* 164 (1): 163–187. http://dx.doi.org/10.1007/s11098-013-0092-z.

Cholbi, Michael. 2011. "The Moral Conversion of Rational Egoists." *Social Theory and Practice* 37 (4): 533–556. http://dx.doi.org/10.5840/soctheorpract201137432.

Cohen, G. A. 1996. "Reason, Humanity, and the Moral Law." In *The Sources of Normativity*, edited by Onora O'neill, 167–188. Cambridge: Cambridge University Press.

Dostoevsky, Fyodor. 2002. *The Brothers Karamazov*. New York, NY: Farrar, Straus, and Giroux.

Fischer, John Martin. 2010. "Responsibility and Autonomy." In *A Companion to the Philosophy of Action*, edited by Timothy O'Connor and Constantine Sandis, 309–316. Oxford: Blackwell Publishing.

[46] This is especially true for philosophers in the existentialist tradition. See, for instance, Søren Kierkegaard (2006) and Jean-Paul Sartre (2007).

Acknowledgements This paper has benefited from a number of helpful comments. In this regard, special thanks to: Karl Ameriks, Ryan Hammond, Dan Immerman, Sam Newlands, Jordan Rodgers, Ben Rossi, Fred Rush, Will Smith, Dan Sportiello, and my anonymous reviewers.

Foot, Philippa. 2001. *Natural Goodness*. Oxford: Oxford University Press.
Frankfurt, Harry. 1998. "Freedom of the Will and the Concept of a Person." In *The Importance of What We Care About*, 11–25. Cambridge: Cambridge University Press.
Frankfurt, Harry. 1999. "Autonomy, Necessity, and Love." In *Necessity, Volition, and Love*, 129–141. Cambridge: Cambridge University Press.
Gewirth, Alan. 1978. *Reason and Morality*. Chicago, IL: University of Chicago Press.
Haidt, Jonathan. 2001. "The Emotional Dog and Its Rational Tail: A Social Intuitionist Approach to Moral Judgment." *Psychological Review* 108: 814–834. http://dx.doi.org/10.1037/0033-295X.108.4.814.
Hume, David. 2000. *A Treatise of Human Nature*. Edited by David Fate Norton and Mary Norton. New York, NY: Oxford University Press.
Kant, Immanuel. 1996. *Groundwork of the Metaphysics of Morals*. In *Practical Philosophy*, edited by Mary Gregor and Allen Wood, 37–108. Cambridge: Cambridge University Press.
Katsafanas, Paul. 2013. *Agency and the Foundations of Ethics: Nietzschean Constitutivism*. Oxford: Oxford University Press.
Kierkegaard, Søren. 2006. *Fear and Trembling*. Edited by C. Stephen Evans and Sylvia Walsh. Cambridge: Cambridge University Press.
Kitcher, Philip. 2014. *Life After Faith: The Case for Secular Humanism*. New Haven, CT: Yale University Press.
Knausgaard, Karl Ove. 2013. *My Struggle: Book 2*. New York, NY: Farrar, Straus and Giroux.
Korsgaard, Christine. 1986. "Skepticism about Practical Reason." *The Journal of Philosophy* 83 (1): 5–25. http://dx.doi.org/10.2307/2026464.
Korsgaard, Christine. 1996. *The Sources of Normativity*. Cambridge: Cambridge University Press.
Korsgaard, Christine. 2008. "Realism and Constructivism in Twentieth-Century Moral Philosophy." In *The Constitution of Agency: Essays on Practical Reason and Moral Psychology*, 302–326. Oxford: Oxford University Press.
Korsgaard, Christine. 2009. *Self-Constitution: Agency, Identity, and Integrity*. Oxford: Oxford University Press.
Kosch, Michelle. 2006. *Freedom and Reason in Kant, Schelling, and Kierkegaard*. Oxford: Oxford University Press.
Lieberman, M. D. 2000. "Intuition: A Social Cognitive Neuroscience Approach." *Psychology Bulletin* 126: 109–137. http://dx.doi.org/10.1037/0033-2909.126.1.109.
MacIntyre, Alasdair. 1986. "The Intelligibility of Action." In *Rationality, Relativism and the Human Sciences*, edited by J. Margolis, M. Krausz, and R. M. Burian, 63–80. Dordrecht: Martinus Nijhoff Publishers.
MacIntyre, Alasdair. 2007. *After Virtue: A Study in Moral Theory*. 3rd edn. South Bend, IN: University of Notre Dame Press.
Markovits, Julia. 2014. *Moral Reason*. Oxford: Oxford University Press.
Nagel, Thomas. 1970. *The Possibility of Altruism*. Princeton, NJ: Princeton University Press.
Paul, L. A. 2014. *Transformative Experience*. Oxford: Oxford University Press.
Paul, L. A. 2015. "What You Can't Expect When You're Expecting." *Res Philosophica* 92 (2): 149–170. http://dx.doi.org/10.11612/resphil.2015.92.2.1.
Plantinga, Alvin. 2015. "Review of Philip Kitcher's *Life After Faith: The Case for Secular Humanism*." *Notre Dame Philosophical Review* https://ndpr.nd.edu/news/54977-life-after-faith-the-case-for-secular-humanism/.
Sartre, Jean Paul. 1993. *Being and Nothingness*. New York, NY: Washington Square Press.
Sartre, Jean Paul. 2007. *Existentialism Is a Humanism*. New Haven, CT: Yale University Press.
Street, Sharon. 2012. "Coming to Terms with Contingency: Humean Constructivism about Practical Reason." In *Constructivism in Practical Philosophy*, edited by James Lenman and Yonatan Shemmer, 40–59. Oxford: Oxford University Press.

Turiel, Elliot, Melanie Killen, and Charles C. Helwig. 1987. "Morality: Its Structure, Function, and Vagaries." In *The Emergence of Morality in Young Children*, edited by Jerome Kagan and Sharon Lamb, 155–243. Chicago, IL: Chicago University Press.

Velleman, J. David. 2000. "The Possibility of Practical Reason." In *The Possibility of Practical Reason*, 170–199. Ann Arbor, MI: Scholarly Publishing Office, University of Michigan Library.

Warren, Robert Penn. 2001. *All the King's Men*. New York, NY: Houghton Mifflin Harcourt.

Wilkerson, William. 2010. "Time and Ambiguity: Reassessing Merleau-Ponty on Sartrean Freedom." *Journal of the History of Philosophy* 48 (2): 207–234. http://dx.doi.org/10.1353/hph.0.0213.

Williams, Bernard. 1981. "Internal and External Reasons." In *Moral Luck: Philosophical Papers 1973–1980*, 101–113. Cambridge: Cambridge University Press.

Trans*formative Experiences

Rachel McKinnon

Abstract: What happens when we consider transformative experiences from the perspective of gender transitions? In this paper I suggest that at least two insights emerge. First, trans* persons' experiences of gender transitions show some limitations to L. A. Paul's (2015) decision theoretic account of transformative decisions. This will involve exploring some of the phenomenology of coming to know that one is trans, and in coming to decide to transition. Second, what epistemological effects are there to undergoing a transformative experience? By connecting some experiences of gender transitions to feminist standpoint epistemology, I argue that radical changes in one's identity and social location also radically affects one's access to knowledge in ways not widely appreciated in contemporary epistemology.

Increasing attention is being paid to the various decisions we face that change our lives, recently arising in light of L. A. Paul's (2015) groundbreaking work. These decisions have important implications in how we think about decisions (e.g., decision theory) and the phenomenology of these choices. However, there are also important contributions to epistemological questions to be made by focusing on transformative experiences. Moreover, much can be gained by focusing on some of the experiences of trans people[1] deciding to undertake a gender transition. As I argue in this

[1] I should make a few notes on my language choices in this paper. As I also note in McKinnon 2014, I will generally use the language of "trans women" to refer only to transsexual women, and "trans* women," which is the emerging convention, to be the more inclusive term that refers to all forms of transgender women, including genderqueer, genderfuckers, bi-gender, and so on. The generic "trans*" denotes maximal inclusivity, including trans masculine people, agender people, and so on. The primary focus of this paper, though, is on trans women's experiences. What I have to say will apply, in varying degrees to other trans* identities.

I also want to note that I do not ascribe to a gender binary where there are only men and women (even if these categories include trans men and trans women). Moreover, I don't fully ascribe to the distinction between gender (as social and perhaps mental) and sex (as biological). Regarding language, I will use male, female, man, and woman to describe gender identities, whether cisgender or transgender. I'll typically refer to transgender women as "trans women" when it serves my purposes, and cisgender women as "cis women." When I use the general form "woman" or "female," I mean to include both cisgender and transgender women. I know that this is controversial. Fully justifying this is well beyond the scope of this

Res Philosophica, *Vol. 92, No. 2, April 2015, pp. 419–440*
http://dx.doi.org/10.11612/resphil.2015.92.2.12

paper, when we consider trans people's decisions to transition through the lens of transformative experiences and decisions, two insights emerge. The first concerns an important limitation to Paul's account of the normative decision theory of transformative experiences. The second concerns the implications for epistemology when we consider the effects of radically changing one's social identity and location, which is a feature of some (perhaps even many) transformative experiences. This paper is thus composed of two related projects, tied together by considering trans experiences of gender transition viz. transformative experiences.

I begin by first explaining Paul's account of transformative experiences, with a focus on her view on the normative decision theory of undertaking such experiences. I then argue that gender transitions count as—perhaps paradigmatic—instances of transformative experiences. I then show that deciding to undertake a gender transition shows important limitations to Paul's view of the decision theory of transformative experiences. In short, I argue that while one may not know what it will be like after one transitions, one may know what it will be like if one does not.

The remainder of the paper takes up the epistemic effects of radically changing one's social identity and location, which is something that almost universally occurs when one undertakes a gender transition. I make this case by connecting transformative experiences to feminist standpoint epistemology. In rough outline, I argue that radical changes to one's social identity and location give one important new access to knowledge that was unavailable or prohibitively difficult to obtain prior to the change. This has broader implications for epistemology: these changes will happen, to greater and lesser degrees, whenever one radically changes one's social identity or location. Moreover, this has political implications, particularly for anti-racist, anti-sexist, and anti-oppression projects.

1 Gender Transitions as Transformative Experiences

In "What You Can't Expect When You're Expecting," L. A. Paul endorses, for the sake of argument, a normative decision theory. In deciding whether or not to transition, for example, one must determine the possible actions, the possible outcomes of those actions, the probabilities of the possible outcomes, the values (in terms of utility) of each of the possible outcomes, and then one should choose the option that maximizes one's expected utility.

paper. However, one worry I have is that making a relatively clear distinction between, for example, "female" and "woman" is cissexist. Let's say that we grant that trans women are women (gender term). Are they female (sex/biological term)? Let's say that we grant that those on hormone replacement therapy (HRT) and post-genital surgery are female. That's problematic for a whole host of reasons, not least of which is the financial burden that such medical interventions cost (they're often prohibitively expensive, which raises class issues, and other intersectional issues). Such a distinction, I think, often seems to make intersex people invisible and placed into "gray areas" of the applications of the concepts in problematic ways.

Paul argues that some experiences are "epistemically transformative." She argues that some knowledge involves "what it's like" experiences, and one cannot have the knowledge of, for example, what it's like to see red unless one has had that experience. So for someone who has never seen red, such as Mary the neuroscientist in Jackson (1986), when she sees red for the first time, even if she knows that red has a particular wavelength, she's epistemically transformed. As Paul writes, "[b]efore she leaves her room [to see red for the first time], she cannot project forward to get a sense of what it will be like for her to see red, since she cannot project from what she knows about her other experiences to know what it is like to see [red]" (2015, 7).

Paul continues:

> Before she leaves her room, because she doesn't know what it's like to see red, or indeed what it is like to see any sort of color at all, she doesn't know what feelings and thoughts she'll experience as the result of seeing red. And so she doesn't know whether it'll be her favorite color, or whether it'll be fun to see red, or whether it'll be joyous to see red, or frightening to see it, or whatever. (2015, 7)

Moreover, she won't know what it'll be like to have whatever emotions she might have from the experience (and subsequently). Essentially, Mary can't know what it'll be like to be herself after her transformative experience. So Paul is arguing that one cannot be rational in deciding what to do when faced with deciding about undertaking a transformative experiences because one cannot know"what it's like," or what it will be like, after one has made the choice. More precisely, though, Paul is arguing that one cannot know what one will be like—what one's preferences will be—after a transformative experience. For example, in deciding whether to have children or not at age 32, I can't know what it will be like for me to be childless at 50, just as I can't know what it will be like for me to have an 18 year old child when I'm 50.[2] Moreover, I can't know what I will be like at 50 with a child, or even what I'll be like at 50 without a child. I can't know whether my preferences will change and, if they do, what they would be.

Just as Paul argues that having one's first child is both epistemically and personally transformative, so is a gender transition. An experience is personally transformative when "it may change your personal phenomenology in deep and far-reaching ways. A personally transformative experience radically changes what it is like to be you, perhaps by replacing your core

[2] I think this isn't quite right, though. I think that I can have a much better idea of what it will be like for me to be childless at 50 than of what it might be like for me to have an 18 year old child when I'm 50: the former is an easier projection of my life at present, whereas the latter involves a transformative experience (having and raising a child). However, there is certainly room for disagreement here. For a useful discussion, see Arvan Forthcoming. I return to this point in section 2.

preferences with very different ones" (Paul 2015, 8).[3] Gender permeates our lives, often in ways that those who've never wrestled with their gender identity don't realize. From our gendered names, to pronouns, to what we wear, and to how people relate to us, gender inflects all facets of our lives. Changing one's gender—say, from one binaristic identity to another—will radically change one's life.[4]

For someone who transitions from, for example, a relatively stereotypical masculine male identity to a relatively stereotypical feminine female identity, nearly everything about her experiences will change.[5] Moving through the world where people attribute a male gender is very different from moving through the world where people attribute a female gender, particularly in sexist, patriarchal societies such as ours. Men are typically afforded more space than women; women are more likely to be ignored in conversation; people are more comfortable being touched (casually on the arm during conversation, for example) by women; and so on.[6] The way one relates with social and legal institutions is changed, particularly if the person has to navigate the often complicated systems of changing their sex/gender marker on identification such as a driver's license, birth certificate, or social security number. It may change one's tastes in clothing, or at least one's ability to express and participate in various clothing and gender presentation preferences. And in almost all cases, one's post-transition preferences will have shifted over time: it's impossible to predict what gender presentation preferences one will have post transition, and how those choices will feel like, and how they will affect others' interactions with oneself. For example, how do people react to someone claiming a femme lesbian identity compared to a femme hetero identity? How will one's athleticism be treated post-transition? How will one's newfound hobbies of, say, sewing and baking be viewed?

There may also be biological or physiological changes, particularly if one receives hormone replacement therapy or transition-related surgeries.[7] The latter will certainly change how one has sex with partners and how one experiences sex (including masturbation). It changes the experiences of even more basic bodily functions such as going to the bathroom. A gender transition, then, is a paradigmatic instance of what Paul refers to

[3] For a useful discussion on ways that this changes what one knows, particularly viz. sensual knowledge, see Shotwell 2011.

[4] I can't stress enough, though, that not all people are binary-identified, and that not all trans people transition from one binary identity to another. People's experiences will differ, but insofar as one changes one's socially recognized gender—e.g., in claiming a trans identity—one will radically change one's experiences.

[5] This is not at all to say that all these things must change, or that they should, nor is it to say that they will for everyone who transitions.

[6] A useful, accessible discussion of this can be found in Vincent 2006.

[7] It's crucial to note that not everyone has access to safe and adequate transition-related medical care. It's also crucial to note that not all trans people want any transition-related medical interventions such as HRT or surgeries.

as a transformative experience. The decision to transition, then, is also a paradigmatic transformative decision. However, as I argue in the next section, when we consider the rationality of choosing to transition, we'll see that Paul's account of the normative decision theory of transformative decisions gets the gender transition cases wrong.

2 Implications for Decision Theory

Insofar as one must be able to at least attach approximate utilities to each of the possible outcomes of a transformative experience, Paul is correct that one can't make a normatively rational choice in transformative experience decisions. However, this doesn't mean that one can't know (or reasonably believe) what the expected utility of choosing not to have the transformative experience will be.[8] And that's where I'd like to place some of my focus: for many, but certainly not all, trans people contemplating a gender transition, they know (roughly, though some may know exactly) the expected utility of not transitioning. The upshot is that this shows an important limitation of Paul's decision theoretic account of transformative experiences.

To simplify things, let's assume that in deciding to transition, one has the decision between two mutually exclusive options: transition and not transition. Each decision option will have many possible outcomes, but let's simplify further and assume that one will either be happy or not happy for each decision option. So we have two decision options, each with two possible outcomes, for four possible outcomes: transition-happy, transition-unhappy, not transition-happy, and not transition-unhappy.

The suicide and depression statistics for trans* people are distressing. In a number of studies, the percentage of trans* people surveyed who've attempted suicide at least once is around 41%.[9] For many but not all trans people, the available options are either transition or commit suicide.[10] This means that the "not transition-happy" outcome is so unlikely as to be effectively impossible. In cases structured such as these, the outcomes

[8] This then connects to my brief discussion in section 1, and footnote 2, of being better epistemically positioned to know what it will be like not to have a child than what it will be like to have a child

[9] It's higher for those without family support, for those with more oppressed intersectional identities, and lower for those with family support and with less oppressed intersectional identities. See: http://williamsinstitute.law.ucla.edu/wp-content/uploads/AFSP-Williams-Suicide-Report-Final.pdf, last accessed September 12, 2014. The reader may note that I switch from discussing "trans*" people to "trans" people. The reason is that not all trans* people decide what we would understand as a "gender transition"; however, it's a common feature of what it means to be trans.

[10] I want to make it as clear as possible that not all trans people have the pre-transition experience of considering suicide as the only available option to attempting to transition. I make no claim whether such experiences constitute "most" of trans people's experiences. I do want to note, though, that the decision theory situation I'm setting up will also work if the outcome options for the agent aren't simply <suicide or transition>, but also include <deeply unhappy life or transition>. I thank an anonymous referee for the latter point.

of deciding not to transition are known: the probability of "not transition-unhappy" is essentially 1 for many trans people, and the probability of "not transition-happy" is essentially 0.[11] And the disutility of the "not transition-unhappy" outcome is large. What this means is that the decision situation is between known suffering (or suicide, depending on how we characterize it), and a gamble at an unknown probability of happiness (of an unknown magnitude).

Many of the experiences for trans people who identify with a gender significantly different than their birth-assigned gender (say, someone assigned male at birth but who identifies as female) are that once one comes to understand oneself as trans (and to identify with their gender identity), there's a feeling of an existential need to transition, and to do it as soon as possible. It becomes all-consuming, often in surprising ways. We can tragically see this in the experiences of those who we've lost to suicide: they often report that when they realized (or strongly believed) that they couldn't transition, often due to a lack of family and social support and acceptance, they saw no other future than one of misery living as their birth-assigned gender.

In light of these features of the decision to transition for many trans people, consider an analogy. Suppose I have to place a bet with my life. I know I'll lose if I bet on red. But I have an unknown non-zero chance of winning some unknown amount by betting on black. If I care about winning, the only rational choice is to bet on black. So I should do that, according to normative decision theory. The same is true for many trans people contemplating transition.

This has important implications for Paul's decision theoretic account of transformative decisions and experiences.[12] Paul's description of the (normative) decision theory of transformative experiences is importantly incomplete. She's quite right that in most transformative experiences, one can't know—or even reasonably guess—whether one will likely be happy (or unhappy) after the experience. Moreover, one is epistemically blocked from knowing one's post-transformative experience preferences required to complete the utility calculus and perform a rational decision. However, some trans people's experiences of transition as the only option other than suicide (or of a life of extreme unhappiness) shows us that in contexts such as these, choosing to undertake a transformative experience becomes rational.[13] Just as it's rational to bet on red in the aforementioned case, it's

[11] Of course, this won't be true of all trans people contemplating transition.

[12] Now, one might think that gender transitions aren't best characterized in terms of Paul's analysis of transformative experiences. The decision to transition shares all of the key features of Paul's analysis of transformative experiences, as I discussed at length in section 1, and I can't think of a good argument for excluding gender transitions as, properly speaking, transformative in Paul's terms. I thank an anonymous referee for suggesting this point.

[13] In a sense, this sort of bet is what poker players (and gamblers, more generally) call a "freeroll." These bets involve structures where one will be no worse off for "losing" the bet than one is before one takes the best, and since one has a non-zero chance of winning, the

rational for trans people to transition if their only other option (that they judge worth pursuing) is suicide.[14] Paul's account of the decision theory, and thus the rationality, of transformative experiences needs to account for cases structured such as gender transitions, where one can effectively know the probability and cost associated with not deciding to undertake the transformative option. However, I make no comment on how that ought to be done.

3 Trans*formative Knowledge

But what of the coming to know about one's trans* identity, and the potential attendant decision to transition? And what does the phenomenology of that coming to know look like?[15] Phenomenologically, an experience common to many trans people at the beginning of their transition is an identifiable instant where one goes from not considering oneself trans (that is, anything other than the binaristic gender one was assigned at birth), to opening oneself up to the possibility of being trans, to knowing that one is trans.[16]

Why would there be so many stories sharing this experience, though? One might think that coming to realize that one is trans (and that one ought to transition) would be much like what has sometimes been called a "feminist awakening?" As Clara Fischer (2014) describes it, a feminist awakening is a "transformative experience from nonfeminist (un)concsiouness to

bets always have a positive expected value. Accepting freeroll bets then, ceteris paribus, is always rational. Transition for the trans people I've described often shares the structure of a freeroll: not transitioning essentially guarantees deep unhappiness, so the worst that could happen post-transition is to be just as unhappy as one would be without transitioning. And, to put it starkly and darkly, if transition doesn't work out, if one was already considering suicide as the only option other than transition, that option is still available if the transition doesn't improve one's life. However, people's quality of life almost universally improves post-transition, particularly if there is adequate family and social support. So it's an especially good bet.

[14] It's certainly possible that someone could go from not being suicidal pre-transition to suicidal (or otherwise deeply unhappy) post-transition. Such cases exist but are extremely rare. Most studies, and more are being released with increasing frequency, show that the vast majority of trans people's quality of life is improved by transitioning. My thanks to Marcus Arvan for raising this worry.

[15] It's important to stress once again, though, that not all trans* people decide to undertake any form of transition, let alone seek medical interventions. While I'm generally focusing on trans people who do so decide, some who would otherwise want to may decide against it due to the imagined or quite real social, economic, and physical costs of such a decision. As I've already noted, the suicide rates among trans* people are shockingly high. Part of this is attributable to those who lack good social support for their transition.

[16] One might wonder about the epistemology of this instant: where does it come from, and what's going on, epistemically speaking, in that moment (or range of moments)? I don't have much to say on this topic. For the purposes of this paper, I leave it largely mysterious. However, I think that the ensuing discussion of William James and the shift from an option changing from dead to live is illuminating.

feminist consciousness" (122). That is, it's when one comes to self-identify as a feminist. Fischer argues that "coming to feminist consciousness is not an abrupt, sudden event, but rather a protracted experience, being rooted in the contradictions of oppressive systems, manifested in feelings of uncertainty and unease" (140). As I'm arguing, though, many people's "trans awakening," as it were, are unlike Fischer's description of a slow, gradual feminist awakening. Understanding why will be important, as I'll argue that many trans women's "trans awakening" was shortly followed by their feminist awakening. And how these experiences differ will tell us something important about epistemology.

In understanding how coming to know that one is trans may be abrupt, though perhaps at the end of a long struggle with one's unease with one's birth-assigned gender, William James's (2014 [1897]) discussion of a genuine hypothesis, and particularly his distinction between live and dead options, is helpful. For James a live hypothesis or option, for an agent, is one that is a legitimate candidate for belief by the agent. Not all possibly (or even necessarily) true propositions are live options for all agents. James's example was belief in God: for some atheists, it's simply not the case that they'll possibly believe in God.[17]

Until I came to know myself as trans, one might say that I considered myself cis (well, I didn't know about the concept of cisgender, so I merely didn't consider myself trans). The truth is that I was long aware that trans people existed (though only through terrible, stereotyped media portrayals in movies like *Ace Ventura: Pet Detective* or daytime television like *The Maury Povich Show*). And while I experienced a distinct and persistent discomfort with my gendered self starting around age 12, I didn't once consider being trans as even a possible explanation. That is, it simply wasn't a live hypothesis for me. However, I can distinctly remember the moment (even the exact date) where I first opened up being trans as a live option (after much research and reflection on trans narratives).[18] And in the very same instant, I went from opening myself up to the possibility to knowing that it was the case: I was trans. The rest, they say, is history.

What's important is that this is a very common experience among those who undergo a gender transition: years of doubt, avoidance, of not viewing

[17] He raised this as an objection, and I think a convincing one, to Pascal's Wager argument directed at atheists. Pascal effectively argued that atheists convinced by the wager argument, but who didn't yet believe in God, should go through the motions as if they believed in God, and eventually they would come to believe. James was raising a problem for this argument.

[18] I think it's important to flag that I don't share the "traditional" trans narrative: knowing from approximately age 3, not engaging in behaviors expected for one's birth-assigned gender, having a post-transition heterosexual orientation, the feeling of being "trapped in the wrong body," and so on. This made coming to know difficult, since most widely circulated trans narratives portray exclusively the "traditional" narrative, and so it was hard to find stories and experiences that matched mine. Thankfully, though, that is slowly changing, as more voices are speaking, and a more accurate representation of the vast variety of experiences is being shared and read. For a discussion of some reasons we should reject this as a standard trans narrative, see Bettcher 2014.

being trans as a live option. But the moment it's opened up as an option, the phenomenology of the transition to knowing is abrupt and almost instantaneous. Moreover, were being trans a live hypothesis to me at 12, I'm confident that I would have come to know much earlier than I did. The epistemic roadblock was, in a real sense, merely that being trans wasn't a live option for quite a long time.

In what follows, I will turn my lens to what trans experiences can teach us about epistemology. In particular, I will consider how viewing gender transitions through feminist standpoint epistemology can teach us something important about how radically changing one's social identity and location can create new, and very different, opportunities for knowledge.

4 Radical Changes in Epistemic Standpoint

Relatively much has been written in feminist standpoint epistemology of the importance of one's social identity, location, or situatedness for access to various instances or forms of knowledge. A typical view, for example, is that members of oppressed groups are often in a better epistemic position to see the oppressive nature of social institutions. However, relatively little has been written on what effects changes in one's situatedness, whether minor or radical, have on knowers.[19] My purpose in this section is to explore these effects, focusing on gender transitions as a case study.

Feminist standpoint epistemologies (FSEs) focus on three central theses: situated knowledge, epistemic privilege, and achievement (Harding 1991, 1993; Wylie 2001, 2003, 2004; Pohlhaus Jr. 2002; Rolin 2006; Intemann 2010; Crasnow 2013). All of these theses are related, but they have some distinct attributes. It's important to note that there are many different feminist standpoint epistemologies, hence the plural. Currently endorsed epistemologies shift and change over time, particularly as new knowers enter the conversations to question assumptions of extant theories.[20] Each of these theses can be brought to bear on insights gained from radical shifts in one's situatedness.

For much of epistemology's history, it was thought that politics or one's identity, such as biases and prejudices, could only contribute to block objectivity in science, and to knowledge acquisition in general. FSEs—and feminist standpoint empiricism—turn that assumption on its head by acknowledging how one's identity can be a critical resource in creating knowledge. In fact, feminist epistemologists and feminist philosophers of science started to see gaps in our scientific knowledge (and knowledge

[19] As I'll discuss, two notable exceptions are Kukla and Ruetsche 2002 and Shotwell 2011.

[20] Importantly, much of what FSEs have to contribute is consistent with much of "mainstream," mostly atomistic epistemology, particularly when we focus on definitions of knowledge. However, FSEs tend to categorically reject the concept of an atomistic knower; that is, one who can adopt the view from nowhere, to use Nagel's (1986) phrasing, and know some proposition without any reference to their situatedness.

more broadly) due to traditional atomistic epistemologies that considered, for example, the gender of the researcher irrelevant. So, far from being merely a liability, in some contexts one's situatedness is an asset for creating knowledge. But what do we mean by situatedness, and what is FSEs' situatedness thesis?

Each person has a complicated intersectional identity, composed of various socially and biologically constructed factors.[21] These factors include race, gender and gender identity, sexual orientation, socioeconomic status, education, religious affiliation, nationality, and so on. These also include perceived versions of these statuses. For example, someone who is black might appear, and thus be racialized, as white. These are sometimes called "invisible" identities.[22]

What matters for the situated knowledge thesis is that one's social location as, say, a cisgender black heterosexual woman, as a member of an oppressed class, may allow her to "recognize that many of the concepts and procedures adopted by [a] discipline are problematic when her colleagues do not, precisely because she is able to see the objects of study both with the eyes of a researcher trained in the discipline and through her own experience from a marginalized social location" (Crasnow 2013, 417, discussing Collins 1986).[23]

The situated knowledge thesis goes hand in hand, I think, with the epistemic privilege thesis. The latter is the idea that those with a particular situatedness—particularly those with oppressed intersectional identities—have, as a consequence of having their identity within a social structure, an epistemic advantage in accessing certain kinds of knowledge, especially of the structures of oppression themselves.[24] For example, if we want to

[21] However, I am of the view that any biological feature, such as race, sex/gender, and so on is also inherently socially constructed. What it means for someone to be black, mixed race, a man, or a woman (or neither!) inherently depends on social decisions, almost always implicit and undisclosed. For some useful discussions of Intersectionality, see Crenshaw 1991 and Garry 2008, 2012.

[22] See Alcoff 2005. This is sometimes also discussed in the stereotype threat literature. See McKinnon 2014.

[23] I think it's critical to note that this insight was long ago discussed in terms of "double consciousness." Collins and Crasnow both use the phrase "double vision," which I worry is ableist.

[24] Talia Mae Bettcher (2009) argues that, setting aside the epistemic privilege question aside, there's a moral duty to give trans people (and, by extension, those with margianalized identities) first person testimonial authority. It's also important to note that one view in feminist standpoint epistemology is that we should aim to understand social structures and identities from the perspective of those with the relevant identities and situatedness. See, for example, Harding 1991, 2006. Gaile Pohlhaus (2012) makes a similar point. She argues that "if a person's social position makes her vulnerable to particular others, she must know what will be expected, noticed by, and of concern to those in relation to whim she is vulnerable, whereas the reverse is not true. Finally, when one is marginally positioned, the epistemic resources used by most knowers in one's society for knowing the world will be less suited to those situations in which marginally situated knowers find themselves on account of being marginal" (Pohlhaus Jr. 2012, 717).

know about problems with how women's testimony (not strictly only in legal contexts) is often treated with insufficient epistemic authority (that is, we're interested in understanding what Fricker (2007) calls testimonial injustice) then we should ask women, not men. Since women inhabit the relevant social locations for the knowledge we seek to gain, they're the better sources due to their inhabiting that identity within that structure of oppression. Moreover, if we want to understand how black people suffer epistemic injustice, we should ask black people about their experiences, not white people.

Here's an illustrative example of both the situated knowledge and epistemic privilege theses working together. I'm fairly active on social media platforms such as Facebook. An acquaintance had posted a story about a recent study of hormone therapy treatment for trans women, specifically focusing on the importance of antiandrogens (testosterone blockers) in concert with estrogen therapy. A cisgender male physician, who works with trans patients, made a comment about how antiandrogens are critical in treating "bio-males." I asked him to define this term, which he took to refer to "someone who has functional testes with average production levels of testosterone in their system." This is deeply problematic language to use in describing trans women, but he couldn't understand what might be objectionable about this term and its use.

This is oppressive terminology, particularly in describing trans women. There is no set of necessary and sufficient conditions for a body to count as male or female.[25] The categories are socially constructed in that science alone can't tell us how to classify all people into the binary categories of "male" and "female." There will necessarily be borderline and unclear cases. For example, like the physician, most people take someone with functioning testes to count as clearly male. However, there are a number of intersex conditions where someone appears otherwise female, but have functioning testes that produce testosterone. Are we to count these people as male or female? Biology alone can't answer that for us. Moreover, what of a trans woman who has had genital surgery (which includes removing the testes)? Is she suddenly no longer a "bio-male"? If so, then this privileges those who desire (as not all trans women do) and can afford (as even fewer can) genital surgery. Rather, the physician should use more descriptive terms such as "trans women who have functioning testes" rather than "bio-male." The latter is oppressive in ways that the former is not.[26]

The key point is that the cisgender male physician couldn't understand how "bio-male" might be oppressive, let alone offensive. However, as a trans woman myself, I've had to deal with the ways that being socially

[25] A useful recent discussion can be found in Karkazis et al. 2012.

[26] For two recent blog posts/online articles on this topic, see http://www.autostraddle.com/let-it-go-for-the-last-time-trans-women-were-not-born-boys-255055/ and http://www.metamorpho-sis.com/blog/2013/04/40-hey-youd-know-right-how-do-guys-think.html, both last accessed September 30, 2014.

labeled "male," particularly in reference to my biology (chromosomes, hormones, presence or absence of testes, etc.) operates within social and medical systems to create oppression. For example, in many jurisdictions, one must have genital surgery in order to change the sex/gender marker on one's identification such as a driver's license, health care card, passport, social security number, and so on. Conceiving of gender in terms of someone's genitals, then, affects whether their gender—as female, in this case[27]—can be legally and socially recognized. And this produces oppression. Someone without the double consciousness of both a trans woman who has struggled with the relevant systems of oppression, and one well trained as a researcher (who also engages in trans advocacy and activism) is less well epistemically positioned to understand "bio-male" as oppressive.[28] And unsurprisingly, this is what happened. Even upon explaining it, the physician struggled to grasp the seriousness of the problem. Importantly, this person considers himself an advocate for trans health care (and he is). But he remains epistemically impoverished due to his situatedness.

Finally, the achievement thesis is that knowledge isn't something passively gained from the world. Rather, knowledge is gained through struggling with the world, with particular attention to one's situatedness and social structures. This connects nicely to my comments earlier about the phenomenology of a gender transition. It effectively forces one (in transitioning from one instantiation of masculinity to an instantiation of femininity, in particular I think) to struggle with the world in ways one hadn't before.

It's a contingent fact of our world that navigating the world in differently gendered bodies and personalities presents people with different experiences. Men typically have some version of male privilege, which may entail a likelihood of being paid more than a woman of equal qualifications, on one hand, or having his utterances given a higher baseline level of respect, on another hand. Of course, this is not to say that all men will benefit from male privilege, or that they will do so equally. Intersectionality matters, as always. However, to take but one of many examples from my own experiences, pre-transition, I noticed that my questions at philosophical conferences were taken up by speakers and afforded a level of respect appropriate to taking a question seriously.[29] However, on one vivid occasion, as the only women in the audience of a session on decision theory (an area in which I have some expertise), I asked a fairly simple question about the basis of the speaker's argument. He didn't understand what I was asking. I re-phrased. He still didn't understand. I re-phrased one more time.

[27] Bettcher (2007; 2009) writes about how gender presentation in our North American society is often about genitals. It's important to remind the reader that I don't subscribe to the sex/gender distinction at all.

[28] Here I'm thinking of the work of W. E. B. Du Bois (2007 [1903]).

[29] I take up the relationship between gender, language, and norms of language in the final chapter of McKinnon 2015.

Still nothing. A man (and soon to become friend, particularly due to our bonding over this experience) reiterated my question, almost verbatim of the form of my first asking, and suddenly the speaker understood and gave an answer.

This is an experience that many women have experienced. One need only look at the many posts on blogs such as the "What it's like to be a woman in philosophy" blog www.beingawomaninphilosophy.com.[30] And we may grant that many (cis) men are well aware that sexism exists, and may know, propositionally, that women face this sort of discrimination and epistemic injustice. However, I've personally experienced a particularly transformative kind of change in my perceptions of events like this. Sexism stands out in a way it didn't before: being forced to struggle against implicit bias, sterotype threat, attributional ambiguity, harassment, and all the social ills disproportionately visited upon women has changed my epistemic access to how things are in the world. And while such knowledge is, I think, strictly speaking available to (cis) men, their not having to struggle with it in the same way as many women leads to the men being epistemically disadvantaged on these issues. And while, pre-transition, I was ostensibly aware of sexism and many of its manifestations, I didn't fully understand what it felt like to experience sexism, and what it felt like to inhabit a world of structural oppression. So part of the shift in my understanding of sexism has come from experiencing the "what it's like" of sexism in addition to having the propositional knowledge that these forms of sexism exist.[31]

Critically, though, this coming-to-know and understand the presence and operation of sexism and misogyny on women isn't limited to raising one's personal awareness, or the awareness of only those who share the same intersectional identity. The insights that, in this case, trans women gain into sexism and misogyny can aid in cis women's understanding too. For example, some of my experiences of the same conversational spaces before-and-after transition give insights into how deeply gendered expectations of speech permeate our worlds. And sharing these observations with cis women can (and indeed has) led to them coming to understand these features of their spaces in ways they may not have previously appreciated.

This is something that I think has been given insufficient attention in the various, mostly feminist, literatures touching on standpoint epistemology. Relatively little has been said about the epistemic and political implications

[30] For further discussions of these issues, see Hutchison and Jenkins 2013.

[31] But while I think that there are parallels with, for example, a person of color experiencing structural racism, I don't take my experiences understanding sexism and misogyny (and, indeed, transmisogyny) to give me critical insight into what it's like to be not white in our (Western, North American) world. For example, as I argue below, while I may understand the ways in which certain spaces are "white spaces," I lack the "what-it's-like" understanding of what it means to inhabit such spaces as a non-white person. I will fail to notice subtle (and not-so-subtle) ways in which white spaces fail to make way for non-white bodies and participation. I will likely also fail to notice subtle (and not-so-subtle) ways in which white spaces facilitate my presence and participation.

of what happens when someone is able to shift their social situatedness. Rebecca Kukla and Laura Ruetsche (2002) discuss how people can struggle to change their "contingent second natures." However, their work focuses on changing how one interacts with knowledge, often through much personal struggle, but doesn't focus on how changing who one is—that is, one's identity and social situatedness—may affect one's position as a knower. So in an important sense, their work focuses on less radical changes than what I'm interested in discussing.

Alexis Shotwell (2011), to some extent, discusses changes in identity and how this may affect one as an epistemic agent. She focuses, though, more heavily on the implicit understandings, such as one's knowledge of oneself, one's embodiment, emotions, and so on, than she does of how one may change one's ability to come to know things about the world apart from oneself. Instead, I want to focus on what Avery Gordon (1997) has called "transformative recognition." She writes, "Being haunted draws us affectively, sometimes against our will and always a bit magically, into the structure of feeling of a reality we come to experience, not as cold knowledge, but as transformative recognition" (8). And this is what a gender transition is like: one is forced to struggle with radically different experiences. One becomes changed as a knower, almost magically, often whether one wants to or not. So while changes in one's contingent second natures typically involves much conscious effort, the epistemic changes one faces when one radically changes one's situatedness, such as through a gender transition, is different in both degree and kind.

I think we can learn something important about what it means to come to know things—about oneself, others, and the world—by considering what happens when people radically change their situatedness. I grant, of course, that such changes are hard to undertake, and there aren't many instances where people can undergo such changes. It's not as if one can change one's race, for example, in the same way one can change one's gender identity.

One might question this, though.[32] In *Black Like Me* (1961), John Howard Griffin does for attempting to go undercover viz. race what Norah Vincent does viz. gender. Griffin, a white man, takes on the appearance and persona of a black man. Essentially, Griffin undertakes a project of passing as a black man.[33] In the preface, Griffin writes that, "This may not

[32] I thank an anonymous referee for raising this issue.

[33] Issues of passing come up in a number of fraught ways, and a black person can "pass" as a white person, for instance; similarly, a cis man can "pass" as a woman under the right conditions. Issues here, though, are that "passing as" connotes that one is not authentically the identity that one passes as. This can be fraught, for example, given that there is a large focus in some trans communities (and certainly in many media obsessions of trans identities and lives) on "passing" but not in the "passing as" sense. Trans women, for example, simply are women, after all. Passing in this sense is much more closely connected to what Kessler and McKenna (1978) refer to as gender attribution. A trans person's efforts to "pass" as their authentic gender, then, involves attempts to alter the gender attributed to them by others

be all of it. It may not cover all the questions, but it is what it is like to be a Negro in a land where we keep the Negro down. Some Whites will say this is not really it. They will say this is the white man's experience as a Negro in the South, not the Negro's. But this is picayunish, and we no longer have time for that" (1961, i).

Griffin's project involved only changing the pigmentation on his visible skin. He didn't change his name or any other details. Certainly, this would be easier to get away with in the pre-internet era. A journalist like Griffin would be easily searchable, and images of him as a white man would be easily discoverable. He travelled to New Orleans, arranged to have his skin pigmentation changed, and began his experiment.

This undoubtedly was an edifying experience for him (and thus for his readers). In a sense, I have been arguing for the position that Griffin takes: "How else except by becoming a Negro could a white man hope to learn the truth?" (1961, 1), noting "that the best way to find out if we had second-class citizens and what there plight was, would be to become one of them" (1961, 3). In a sense, it's an avowal of standpoint epistemology. He took himself to be an expert on race issues, but realized that his knowledge barely scratched the surface of the reality of being black in the US south.[34] Insights were gained into what it means to lack white privilege, and to suffer anti-black racism. During his experiment, Griffin gained access to information that as a person racialized as white he did not have previously. For example, black people would mention or discuss features of their experiences of white people that they would only tend to say to other racialized-as-black people. One consequence of my gender transition—and along with it passing privilege—was access to similar information that women tend only to share with other women.

And yet, it's hard not to think that Griffin took himself to be gaining new insights a little too easily, too quickly, and without enough struggle. He claims to be coming to perceive things he never did before partly because "I had seen [the ghetto] before from the high altitude of one who could look down and pity. Now I belonged here and the view was different" (1961, 19). But how different and why? He hasn't yet suffered any anti-black racism. There's nothing epistemically blocking him from seeing things as he does at this point were he still racialized-as-white: he just wasn't looking before.

One problem with his experiment is that Griffin failed to properly inhabit a body racialized-as black. He lacked the appropriate situatedness of being black. He had the option of reversing the blackness of his skin and immediately transforming back into a white man were it to serve his

(insofar as this is even under their control), and so is fundamentally different from Griffin's project, since Griffin continued to be a white man even when he was attempting to pass as a black man. I raise this largely to set it aside.

[34] A consequence of the view I'm offering in this paper is that his "expertise" viz. race issues in the United States at the time, since he was a white man, was relatively impoverished.

purposes. When he recounts the first time a slur was directed at him, he writes, "I learned a strange thing—that in a jumble of unintelligible talk, the word 'nigger' leaps out with electric clarity. You always hear it and it always stings" (1961, 22). This is questionable: does the sting that he feels share the what-it's-like that it would for someone who is actually black? I sincerely doubt this. The slur was certainly directed at his body (which was racialized-as-black) but not him as a person since he was fully aware that he was just playing the part of a black man for his experiment.

All the while, Griffin knew that his long-term economic and social prospects weren't affected by anti-black prejudice and implicit biases.[35] He could code-switch when appropriate: he could break character and reveal his whiteness, and talk in a racialized-as-white way, or produce proof of his whiteness (such as a photograph). It was merely an experiment for Griffin. Being a trans woman, though, is no experiment. Trans women are women. While there is to some extent the possibility to "de-transition," these cases are extremely rare and always traumatic. Much more importantly, though, Griffin's experiences lack the historical situatedness of experiencing, struggling against, and indeed suffering anti-black racism. Racism wasn't directed at him, but only at the façade that he created as part of his experiment.

Philosophers have made related attempts to show an analogy between a gender transition and at least the possibility of race transitions (or being so-called "transracial" or receiving "transracial surgery" or other medical interventions in parallel to the various available gender analogues, such as what Griffin underwent for his experiment). The idea here is that race and gender are analogues. So we have, for example, the claim by Christine Overall (2004) that "if transsexual surgery is morally acceptable . . . then transracial surgery should be morally acceptable" (184). But I think it's wrong to consider gender and race analogues in this sense.

Cressida Heyes (2005, 2009) notes that "to the extent that the creation of particular subjectivities is a necessarily historical process, in which certain possibilities become sedimented by years of social practice, sex [or gender] and race have emerged looking rather different" (2009, 141). In brief, her view is that "an individual's racial identity derives from her biological ancestors undermines the possibility of changing race, in ways that contrast with sex-gender" (142). That is, "race is taken to be inherited in a way that sex [or gender] is not" (144). Moreover, "[w]ith race inhering both in the body and in ancestry, and transracialism lacking a diagnostic mechanism, the marketing of race-altering body modifications cannot play to individual essence to the extent that sex change can" (144).

[35] One might note that Griffin and his supporters were concerned that he would suffer anti-black racist effects were his experiment to become known to, for example, hate groups such as the Ku Klux Klan. However, these effects would attach to Griffin as perhaps a "black sympathizer" rather than as a black man.

There are thus a number of important disanalogies between radical shifts in one's social (and thus epistemic) location viz. race and gender, and this largely explains my focus on gender transitions rather than on race transitions. This presents an unfortunate barrier, I think, to anti-racist epistemologies, for example. George Yancy (2012; 2014; 2015; Unpublished) has argued that one critical aspect of white anti-racist development must involve white people developing the sort of double-consciousness that DuBois (2007 [1903]) articulated that black people develop.[36] On this view, white people need both to come to see themselves through their own eyes as white people and to see themselves through the eyes of black people and other people of color. Even the first part, seeing themselves through their own eyes as white people is a significant step in societies where whiteness is normalized and treated as the default. However, if I'm right that one must experience the "what-it's-like" of an oppressed group identity, then white people's double-consciousness will fail to achieve the depth of understanding capable in black double-consciousness.

In short, I suspect that it's easier for one to come to see what it's like to be oppressed by obtaining the identity and situatedness of the oppressed. It's harder, although I think and hope not impossible, to come to know these same things without such a radical transformation. But fortunately, I also think, we should attend to the experiences of those who do—or are able to—undertake such radical shifts in their situatedness, and what their experiences can teach us about, for example, epistemology.

I noted in the previous section that a common experience for trans women who undertake a gender transition is that their transition is often (though certainly not always) closely followed by some form of a "feminist awakening." The lens of feminist standpoint epistemology can help explain this. When one used to inhabit the world with a male gender attribution, one lacks the what-it's-like quality of misogyny and sexism.[37] One might be relatively well aware of the concept of misogyny and its effects on women. However, actually inhabiting the world as a woman and having patriarchal forces operate on oneself is a different matter. In short, it's the difference between knowing the rules of a sport and actually playing. Some things just have to be experienced. An anecdote will help clarify what I mean.[38]

Prior to my transition, I was somewhat aware of stories where women would say that they were often talked over, ignored, or variously excluded from conversations. However, I had yet to perceive any instances of this. This is not at all to say that I was never present for such instances: were I present, and one of these instances happened, I simply didn't perceive it as such. However, within a few months of transition ("thanks" largely to passing privilege, where people routinely perceive me as a cisgender woman even, paradoxically, many who know about my trans status), I was part of

[36] Nussbaum (1997) discusses how literature may serve this sort of function.
[37] For an insightful discussion of issues of gender attribution, see Kessler and McKenna 1978.
[38] I recount a number of these as well as others on my blog, www.metamorpho-sis.com.

a three person conversation in a lunch room. I was asked a question about pedagogy by one of my male interlocutors, and in the middle of my answer, when I paused to take a breath, he physically turned his body to begin a new conversation only with the other male interlocutor. Since we were arranged in a triangle, this effectively excluded me from the conversation. Neither of them turned to me while speaking to indicate my inclusion in "their" conversation. "My" conversation was clearly over, as the topic was dropped, and my story cut off. So the exclusion was two-fold: the topic I was participating in was dropped and a new one taken up that was focused on the third person's interests, and both male interlocutors physically shifted to face each other, turning the triangle into a pair, with me to the side.

This is, of course, a very minor observation. But it was one of the first of many, of increasing severity. It was surreal: I was aware that women are often excluded from conversations in exactly the way I just was, but I hadn't perceived it before it happened to me. Changing my social identity and location to being a woman changed my situatedness and it changed how I struggle against subtle forms of misogyny and sexism. This changed, and began to sharpen, my ability to even perceive such instances as instances of sexism. The social change led to epistemic changes. I don't claim that these changes were inevitable, though: the transition didn't guarantee that I would have these epistemic changes. However, it certainly facilitated the changes.

Does this mean that having the "what-it's-like" of experiencing misogyny is necessary to have a feminist awakening? Certainly not. Anyone can be a feminist. However, what I'm claiming is that one's access to various forms of knowledge does depend on one's social identity and location. In order to more deeply grasp what it means to be a woman in our Western societies, one does need to be a woman. Feminist standpoint epistemology helps explain this.

Returning to Clara Fischer's discussion of personal change and feminist awakenings, she writes that "[f]eminist understandings permeate almost every aspect of one's existence now [after one's feminist awakening], as previously unproblematic norms become problematized and reassessed in feminist terms. Issues surrounding the body, sexuality, work, family life, and so on, all come to be seen in a different light, or through what feminists call 'gendered lenses'" (2014, 124). As I've been arguing, the same almost universally happens for those who undertake a gender transition: the same events in the world take place, but one perceives them from a different epistemic standpoint.

Consider one more analogy. Suppose that one has experienced most of one's life as a predator animal. One is an expert hunter and is adept at perceiving features of one's environment (such as the direction of wind, lay of the land, and so on) such that one can be fairly good at predicting where one's prey will move. But suppose that this person is suddenly transformed

into what was very recently their prey animal. On my view, they will begin to perceive the same world in importantly different ways. They may come to understand why the prey attempts to evade the predator as they do. They may begin to perceive areas as good hiding places, ones that the predators don't tend to notice. And they'll start to notice these because their situatedness has changed.

Turning to a real-world example, many men are at least somewhat aware that women often feel unsafe walking home alone at night, particularly after dark. They may even have some understanding why women experience this fear. But they tend to lack the "what-it's-like" experience that women tend to have. And lacking that "what-it's-like" has epistemic effects. For example, many men aren't aware that many women choose to wear footwear on walks home at night in which they can more easily run if they need to.

Now consider a trans woman who transitions in her 20s. I've spoken to many trans women who transitioned in their 20s or 30s who've had the experience where pre-transition they had no real concern about walking home on a particular route at night. But post-transition, they were acutely afraid of that same route, and they changed their behaviors accordingly. What was taken to be known—that women experience fear and concern about walking home alone in the dark—took on a new depth of understanding when the same agent occupied the social identity and position of a woman being confronted with walking home alone in the dark.

So by way of conclusion, I think we can learn something important about what it means to come to know things—about oneself, others, and the world—by considering what happens when people radically change their situatedness. While one may have access to some kinds of knowledge given one's situatedness, one is epistemically disadvantaged—or even blocked—from other kinds of knowledge grounded in other social identities and locations. And I've applied feminist standpoint epistemology in order to make sense of this.

5 Conclusion

In this paper I've raised the question about what we can learn viz. transformative experiences through the lens of gender transitions. Gender transitions are a paradigmatic case of transformative experiences, in Paul's sense. I first argued that considering the rationality of gender transitions for some trans* people shows an important limitation to Paul's account of the decision theory for transformative experiences. On the one hand, one might have some understanding of what it would be like were one not to undertake a transformative experience. On the other hand, and more importantly, in some situations such as some gender transitions, one may rationally choose the transformative experience since the cost of not so choosing is so high.

I second considered some of the ways in which gender transitions are transformative. Specifically, I argued that radical changes in one's social identity and location can lead to radical shifts in one's access to various forms of knowledge. Prior to the transformative experience, one was epistemically disadvantaged—or even blocked—from knowledge that becomes facilitated (though not guaranteed) after the experience. One of the upshots of this is that philosophers will have to re-consider the consequences of including first-person accounts in work related to various intersectional identities. Another is that it raises a worry about various anti-oppression and anti-racist projects that partly depend on those with more powerful social identities and locations developing a sort of epistemic double-consciousness we normally associate with those with the less powerful identities and locations. Both of these, I suspect, will result in a foregrounding of epistemic trust in first-person reports of people with the relevant intersectional identities.

Rachel McKinnon
E-mail : rachelvmckinnon@gmail.com

References:

Alcoff, Linda. 2005. *Visible Identities: Race, Gender, and the Self*. Oxford: Oxford University Press.

Arvan, Marcus. Forthcoming. "How to Rationally Approach Life's Transformative Experiences." *Philosophical Psychology* .

Bettcher, Talia Mae. 2007. "Evil Deceivers and Make-Believers: On Transphobic Violence and the Politics of Illusion." *Hypatia* 22 (3): 43–65.

Bettcher, Talia Mae. 2009. "Trans Identities and First Person Authority." In *"You've Changed": Sex Reassignment and Personal Identity*, edited by Laurie Shrage, 98–120. Oxford: Oxford University Press.

Bettcher, Talia Mae. 2014. "Trapped in the Wrong Theory: Re-Thinking Trans Oppression and Resistance." *Signs* 39 (2): 383–406. http://dx.doi.org/10.1086/673088.

Collins, Patricia Hill. 1986. "Learning from the Outsider Within: The Sociological Significance of Black Feminist Thought." *Social Problems* 33: S14–S32. http://dx.doi.org/10.2307/800672.

Crasnow, Sharon. 2013. "Feminist Philosophy of Science: Values and Objectivity." *Philosophy Compass* 8 (4): 413–423. http://dx.doi.org/10.1111/phc3.12023.

Crenshaw, Kimberlé. 1991. "Mapping the Margins: Intersectionality, Identity Politics, and Violence Against Women of Color." *Stanford Law Review* 43: 1241–1299. http://dx.doi.org/10.2307/1229039.

Du Bois, W. E. B. 2007 [1903]. *The Souls of Black Folk*. Oxford: Oxford University Press.

Fischer, Clara. 2014. *Gendered Readings of Change: A Feminist-Pragmatics Approach*. New York, NY: Palgrave Macmillan.

Fricker, Miranda. 2007. *Epistemic Injustice: Power and the Ethics of Knowing*. Oxford: Oxford University Press.

Garry, Ann. 2008. "Intersections, Social Change, and "Engaged" Theories: Implications of North American Feminism." *Pacific and American Studies* 8: 99–111.

Acknowledgements I'd like to thank Marcus Arvan, Liam Bright, L. A. Paul, and Alexis Shotwell for their helpful conversations on these topics.

Garry, Ann. 2012. "Who Is Included? Intersectionality, Metaphors, and the Multiplicity of Gender." In *Out from the Shadows: Analytical Feminist Contributions to Traditional Philosophy*, edited by Sharon L. Crasnow and Anita M. Superson, 493–530. Oxford: Oxford University Press.

Gordon, Avery. 1997. *Ghostly Matters: Haunting and the Sociological Imagination*. Minneapolis, MN: University of Minneapolis Press.

Griffin, John Howard. 1961. *Black Like Me*. Boston, MA: Houghton Mifflin.

Harding, Sandra. 1991. *Whose Science? Whose Knowledge? Thinking from Women's Lives*. Ithaca, NY: Cornell University Press.

Harding, Sandra. 1993. "Rethinking Standpoint Epistemology: What is "Strong Objectivity"?" In *Feminist Epistemologies*, edited by Linda Alcoff and Elizabeth Potter, 49–82. New York, NY: Routledge.

Harding, Sandra. 2006. "Transformation vs Resistance Identity Projects: Epistemological Resources for Social Justice Movements." In *Identity Politics Reconsidered*, edited by Linda Alcoff, 246–263. New York, NY: Palgrave Macmillan.

Heyes, Cressida. 2005. "Changing Race, Changing Sex: The Ethics of Self-transformation." *Journal of Social Philosophy* 37 (2): 266–282. http://dx.doi.org/10.1111/j.1467-9833.2006.00332.x.

Heyes, Cressida. 2009. "Changing Race, Changing Sex: The Ethics of Self-transformation." In *You've Changed: Ses Reassignment and Personal Identity*, edited by Laruie Shrage, 135–154. Oxford: Oxford University Press.

Hutchison, Katrina and Fiona Jenkins. 2013. *Women in Philosophy: What Needs to Change?* Oxford: Oxford University Press.

Intemann, Kristen. 2010. "Twenty-Five Years of Feminist Empiricism and Standpoint Theory: Where are we Now?" *Hypatia* 25 (4): 778–796. http://dx.doi.org/10.1111/j.1527-2001.2010.01138.x.

Jackson, Frank. 1986. "What Mary Didn't Know." *The Journal of Philosophy* 83 (5): 291–295. http://dx.doi.org/10.2307/2026143.

James, William. 2014 [1897]. *The Will to Believe and Other Essays in Popular Philosophy*. Cambridge: Cambridge University Press.

Karkazis, Katrina, Rebecca Jordan-Young, Georgiann Davis, and Silvia Camporesi. 2012. "Out of Bounds? A Critique of the New Policies on Hyperandrogenism in Elite Female Athletes." *American Journal of Bioethics* 12 (7): 3–16. http://dx.doi.org/10.1080/15265161.2012.680533.

Kessler, Suzanne and Wendy McKenna. 1978. *Gender: An Ethnomethodological Approach*. Chicago, IL: University of Chicago Press.

Kukla, Rebecca and Laura Ruetsche. 2002. "Contingent Knowers and Virtuous Knowers: Could Epistemology be 'Gendered'?" *Canadian Journal of Philosophy* 32 (3): 389–418.

McKinnon, Rachel. 2014. "Stereotype Threat and Attributional Ambiguity for Trans Women." *Hypatia* 29 (1): 857–872.

McKinnon, Rachel. 2015. *The Norms of Assertion: Truth, Lies, and Warrant*. New York, NY: Palgrave Macmillan.

Nagel, Thomas. 1986. *The View from Nowhere*. Oxford: Oxford University Press.

Nussbaum, Martha. 1997. *Poetic Justice: The Literary Imagination and Public Life*. Boston, MA: Beacon Press.

Overall, Christine. 2004. "Transsexualism and "transracialism"." *Social Philosophy Today* 20: 183–193.

Paul, L. A. 2015. "What You Can't Expect When You're Expecting." *Res Philosophica* 92 (2): 149–170. http://dx.doi.org/10.11612/resphil.2015.92.2.1.

Pohlhaus Jr., Gaile. 2002. "Knowing Communities: An Investigation of Harding's Standpoint Epistemology." *Social Epistemology* 16 (3): 283–293. http://dx.doi.org/10.1080/0269172022000025633.

Pohlhaus Jr., Gaile. 2012. "Relational Knowing and Epistemic Injustice: Toward a Theory of *Willful Hermeneutical Ignorance*." *Hypatia* 27 (4): 715–735. http://dx.doi.org/10.1111/j.

1527-2001.2011.01222.x.
Rolin, Kristina. 2006. "The Bias Paradox in Feminist Standpoint Epistemology." *Episteme* 3 (1–2): 125–136. http://dx.doi.org/10.3366/epi.2006.3.1-2.125.
Shotwell, Alexis. 2011. *Knowing Otherwise: Race, Gender, and Implicit Understanding*. University Park, PA: Pennsylvania State University Press.
Vincent, Norah. 2006. *Self-Made Man: One Woman's Year Disguised as a Man*. New York, NY: Penguin.
Wylie, Alison. 2001. "Doing Social Science as a Feminist: The Engendering of Archaeology." In *Feminism in Twentieth Century Science, Technology, and Medicine*, edited by A. N. H. Creager, E. Lunbeck, and L. Schiebinger, 23–45. Chicago, IL: University of Chicago Press.
Wylie, Alison. 2003. "Why Standpoint Matters." In *Science and Other Cultures: Issues in Philosphies of Science and Technology*, edited by R. Figueroa and S. Harding, 26–48. New York, NY: Routledge.
Wylie, Alison. 2004. "Why Standpoint Matters." In *The Feminist Standpoint Theory Reader*, edited by S. Harding. New York, NY: Routledge.
Yancy, George. 2012. *Look, A White! Philosophical Essays on Whiteness*. Philadelphia, PA: Temple University Press.
Yancy, George, ed. 2015. *White Self-criticality beyond Anti-racism: How Does It Feel to Be a White Problem?* Lanham, MD: Lexington.
Yancy, George. Unpublished. "Travon Martin, Philosophy, and White Spaces."
Yancy, George and Maria del Guadalupe Davison, eds. 2014. *Exploring Race in Predominantly White Classrooms: Scholars of Color Reflect*. New York, NY: Routledge.

How You Can Reasonably Form Expectations When You're Expecting

Nathaniel Sharadin

Abstract: L. A. Paul has argued that an ordinary, natural way of making a decision—by reflecting on the phenomenal character of the experiences one will have as a result of that decision—cannot yield rational decision in certain cases. Paul's argument turns on the (in principle) epistemically inaccessible phenomenal character of certain experiences. In this paper I argue that, even granting Paul a range of assumptions, her argument doesn't work to establish its conclusion. This is because, as I argue, the phenomenal character of an experience supervenes on epistemically accessible facts about its non-phenomenal character plus what the deciding agent is like. Because there are principles that link the non-phenomenal character of experiences (together with what a particular agent is like) to the phenomenal character of experiences, agents can reasonably form expectations about the valence of the phenomenal character of the experiences that they are deciding whether to undergo. These reasonable expectations are, I argue, enough to make the ordinary, natural way of making a decision yield rational decision.

1 Introduction

Sometimes, when we are trying to make a choice, we reflect on the phenomenal character of the possible outcomes of our choice. That is, we reflect on "what things would be like" if we chose one way or another. This is a natural way of deciding what to do. In this journal, L. A. Paul (2015) has argued that this natural approach to decision is inapt for a range of choices where the phenomenal character of the outcomes of our choices is inaccessible to us. In particular, Paul argues that this way of deciding cannot yield a rational choice when it comes to deciding whether or not

Res Philosophica, *Vol. 92, No. 2, April 2015, pp. 441–452*
http://dx.doi.org/10.11612/resphil.2015.92.2.2

to have a child. This argument has been picked up widely by the popular press.[1] I think it is mistaken.

Paul's argument turns on the (in principle) epistemically inaccessible phenomenal character of certain experiences. Here I argue that, even granting Paul a range of assumptions, her argument doesn't work to establish its conclusion. This is because, as I will argue, the phenomenal character of an experience supervenes on its non-phenomenal character plus what the deciding agent is like. Hence, because agents can and often do have epistemic access to these facts, and because there are principles that link the non-phenomenal character of experiences (together with what a particular agent is like) to the phenomenal character of experiences, agents can reasonably form expectations about the valence of the phenomenal character of the experiences that they are deciding whether to undergo. And these reasonable expectations about whether the phenomenal character of an experience will be positive or negative are, I argue, enough to make the ordinary, natural way of making a decision yield rational decision. My focus throughout will be, as in Paul's argument, on the decision to procreate. In the conclusion I'll briefly explain how my argument can be extended to other similar decisions. Before all that, Paul's argument.

2 Paul's Argument

According to Paul, reflecting on what it would be like to have a child cannot rationally yield either the decision to have a child or the decision to remain childless. This is because, according to Paul, the phenomenal character of the outcome where you have a child is unavailable to you: you cannot know what it is like to have a child, and so your choice cannot be rationally grounded in considerations of what it would be like. Having a child is, in her words, "epistemically transformative": information about what it is like to have a child is *in principle* unavailable to childless adults; in order to have an epistemic grip on the phenomenal character of having a child, you must therefore actually have one (2015, 8–9). Of course, what this means is that prospective parents cannot rationally appeal to considerations concerning what it would be like to have a child in order to decide whether to have one, since they do not—and cannot—know what, in fact, it would be like. Paul explains (2015, 11):

> The trouble comes from the fact that, because having one's first child is epistemically transformative, one cannot determine the value of what it's like to have one's own child before actually having her. This means that the subjective unpredictability attending the act of having one's first child makes the story about family planning into little more than

[1] Gopnik (2013); Burkeman (2013a,b); Rothman (2013); Lombrozo (2013a,b); Marshall (2013); Moran (2013); Bartlett (2013).

> pleasant fiction. Because you cannot know the value of the relevant outcome, there is no rationally acceptable value you can assign to it. The problem is not that a prospective parent can only grasp the approximate values of the outcomes of her act, for then, at least, she might have some hope of meeting our norms for ordinary decision-making. The problem is that she cannot determine the values with any degree of accuracy at all.

You might object: the distinctive and, let's grant, epistemically inaccessible, phenomenal character of the outcome of having one's own child is just one feature of that outcome. There are other features of the outcome of having one's own child that prospective parents *can* know about. Some of these are phenomenal but not distinctive to having one's *own* child, and some are not phenomenal at all. For instance, prospective parents can know that having a child will be expensive. This would be a nonphenomenal feature of the outcome. They can also know that having a child will affect their sleep patterns in certain predictable ways. This is a phenomenal feature of the outcome of having one's own child, but it is one that even prospective parents can be phenomenally acquainted with; after all, they can have the relevant experience by agreeing to watch someone else's newborn, or by setting alarms to go off at random intervals throughout the night. These are just two examples: there are other nonphenomenal and phenomenal features of the outcome of having one's own child that prospective parents can know about. Here, then, is the worry: Can't prospective parents rationally decide whether to have a child on the basis of reflecting on what it would be like to have a child in terms of these epistemically accessible (phenomenal and nonphenomenal) features of what it is like to have one?

Paul anticipates this objection. Her reply is that the *distinctive* phenomenal character of the outcome of having a child—the one that is epistemically inaccessible to prospective parents—is likely to *swamp* all other considerations in determining whether it is rational to have a child. That is, she grants that there might be nondistinctive phenomenal or nonphenomenal features of the outcome of having one's own child that are epistemically accessible by prospective parents. But, she claims, these features of the outcome of having one's own child will likely be swamped by the distinctive phenomenal character of the outcome: as she says, "even if other [features of the] outcomes are relevant, the value of the phenomenal outcome, when it occurs, might be so positive or so negative that none of the values of the other relevant outcomes matter" (2015, 17). Paul's idea is plausible: what really matters, when it comes to determining whether it is rational to have a child, is not whether you will lose sleep (and what that feels like), nor whether it will be expensive (and what it's like have less disposable income), but instead what it's *distintictively* like to have a child. Deciding to have a child isn't simply a matter of deciding whether to get less sleep or

be less prodigal: After all, you might know everything about the physical and fiscal cost of children and still wonder whether or not it makes sense to have a child. So the distinctive phenomenal character of the outcome of having a child plausibly swamps other considerations when it comes to deciding whether to have a child.

Here Paul's comparison with ordering food from an unfamiliar menu is helpful. If someone else is paying for your meal, then we can suppose that the only thing that matters is what the food tastes like. What it is like to taste the food, as we might put it, swamps all the other considerations. Unfortunately, this puts you in in an unenviable situation. This is because, since you do not know what it's like to taste the items on the menu, there seems to be nothing that could rationally ground your decision for one item over another. You could flip a coin. But then you wouldn't be deciding based on considerations having to do with what it will be like to taste the food you'll thereby have ordered. And so your choice would be disanalogous to the decision to have a child based on considerations having to do with what it will be like to actually have one. Indeed, it would be like deciding whether to have a child by flipping a coin. So if all that matters is what the food tastes like, and you, *ex hypothesi*, have no idea what that is like, then it seems you're rationally at a loss. So it goes with children: if what really matters is the distinctive phenomenal character of what it is like to have a child, and you have no idea what that is like, then it seems you're rationally at a loss.

3 Linking Principles and Rational Expectation

Let's grant that the distinctive phenomenal character of the outcome of having a child always swamps all other considerations when it comes to whether it is rational to decide to have a child. As Paul points out, there are two ways this might happen: swamping can happen in either direction (2015, 17). Either the distinctive phenomenal character of having a child of one's own might be such that, whatever negative nondistinctive phenomenal and nonphenomenal features of the outcome of having one's own child, the "what it is like" to have a child of one's own makes the net value of the outcome of having one's own child positive. Intuitively, this is what happens when the loss in sleep and financial hardship of having a child of one's own are outweighed by the distinctive phenomenal character of the experience. Or the distinctive phenomenal character of having a child of one's own might be such that, no matter what positive nondistinctive phenomenal and nonphenomenal features of the outcome of having one's own child there might be, the "what it is like" to have a child of one's own makes the net value of the outcome of having a child negative. Either way, swamping is a matter of the distinctive phenomenal character of having a child of one's own being the final arbiter of whether the overall value of having one's own child is positive or negative. If the distinctive phenomenal character is one

way, that value will be positive; if it is another, it will be negative. In order to capture this idea, I'll say that the distinctive phenomenal character of having one's own child is *valenced*: it is either positive or negative. Let's also grant that the distintinctive phenomenal character of the outcome of having one's own child is in principle epistemically inaccessible to prospective parents. That is, there is no way for prospective parents to know the content of the phenomenological experience of "what it is like" to have a child of their own before they in fact do. Does it follow from these two assumptions that prospective parents cannot rationally decide whether to have a child by reflecting on what it would be like to have one? It does not.

For notice that, if the distinctive phenomenal character is the final arbiter of whether the net value of the outcome of having one's own child is positive or negative, then in order to know whether it is rational to have a child prospective parents do not need to know the content of that phenomenal character, and they do not need to be assign any particular value to that experience: they only need to know its valence—whether it is positive or negative. And that, I claim, is something prospective parents *can* know—or at least, something about which they can form reasonable expectations. Let me explain.

A new experience that is the same in respect of its non-phenomenal character can strike two individuals differently in terms of its phenomenal character. This is because the phenomenal character of a new experience will be shaped not just by features of the experience itself but also by features of the individual undergoing it. And where two individuals differ in the relevant features, the same experience will, in terms of its phenomenology, be correspondingly different. For example, borrowing from Paul (who in turn borrows from David Lewis [1990]), we can imagine two individuals, *A* and *B*, tasting Vegemite for the first time. Now, Vegemite has certain non-phenomenal, physical characteristics that contribute to what the experience of tasting it is like. Nonetheless, what *A*'s experience of tasting it is like might differ widely from what *B*'s experience of tasting it is like: *A* might find it pleasantly savory and salty, whereas *B* might find it overwhelmingly heady. And this difference in the character of *A* and *B*'s phenomenology of tasting Vegemite for the first time will not be due to a difference in the Vegemite tasted. It will be due, instead, to a difference in some relevant features of *A* and *B*. Perhaps *A* likes strongly flavored things, and *B* doesn't. In fact, we can go further than this. We can say that the particular phenomenal character of a new experience for an agent always supervenes on two things:

(1) The non-phenomenal character of the experience.
(2) What the agent is like.

Call this claim, the claim that the phenomenal character of an experience for an agent depends both on what the experience is like, non-phenomenally

speaking, and what the agent is like, *Phenomenal Supervenience*. Phenomenal Supervenience isn't a controversial claim: it simply says that what it is like to have an experience will depend both on features of that experience and on features of the agent undergoing the experience. How does this bear on Paul's argument?

Notice first there are facts that can take us from knowledge of (1) and (2) in a particular case to a prediction about the valence of the phenomenal character of the relevant experience in that case.[2] For instance, we know what Vegemite is non-phenomenally like and so know what the non-phenomenal character of the experience of tasting Vegemite is like. Suppose in addition to knowing what Vegemite is like I know that *B*, who is about to taste Vegemite for the first time, is a "supertaster." (A supertaster is someone who, probably due at least in part to genetic factors, has a significantly increased number and sensitivity of tastebuds. Such people are much less likely to enjoy strongly flavored foods such as Vegemite (Prescott et al. 2001). Given this knowledge of what Vegemite is like (it is very salty) and what *B* is like (she is a supertaster), we can safely predict—though not, of course, with certainty—that *B*'s experience of tasting Vegemite for the first time will be negatively valenced. This is not to say that we can predict what the content of *B*'s phenomenal experience will be like. Nor is it to say that we can assign a particular value to that experience. It is to say that we can rationally predict that *B* will not enjoy the experience—we can predict that the experience will, overall, be negatively valenced. And, importantly, *B* can predict this fact too, so long as she knows the relevant facts about Vegemite and herself. Such statistical facts are what I will call *linking principles*: they link knowledge of (1) and (2) to predictions of the valence of the phenomenal character of an experience.

Of course, Paul isn't concerned to argue that we can't rationally decide to try Vegemite for the first time. But the example is illustrative. Despite the fact that prior to trying Vegemite for the first time *B* cannot know what it is like to taste Vegemite, *B* can safely predict that she will not like it—she can predict that, all things considered, it will be a negatively valenced phenomenal experience for her. This is because of the existence of a principle supported by empirical facts about the connection between what she is like and what the phenomenal character of her experience will be like, given what the non-phenomenal character of that experience is like. What the example illustrates is that that Paul's argument turns on denying the existence of any such linking principles in the case of having a child. There are two problems with denying the existence of such linking principles.

The first problem is that denying the existence of such principles seems to commit us to the claim that the experience of having a child is not just *epistemically* transformative, but also *personally* so. To see this, notice

[2] For a similar idea, see Dougherty et al. 2015, 307.

that in the absence of personal transformation there must in every case be, as a matter of simple causal necessity, some principle linking what a particular experience is like non-phenomenally and what an agent is like to how that experience will affect them. We could deny that this was so in any particular case of an experience if we thought that part of the experience's effect on a person was to transform them into someone different. Then the effect that an experience had on any particular individual could not be "read off" what the experience is like together with what the agent is like, for what the agent is like would depend on how the experience affected them. This would amount to a rejection of Phenomenal Supervenience in a particular case. But this way of denying the existence of linking principles is unavailable to Paul. This is because Paul explicitly denies that having a child is always personally transformative (2015, fn. 21).

Of course, it is open to Paul to argue that, although having a child is not always personally transformative, it often is; that is, that the probability that having a child will be personally transformative is high. Indeed, she seems to suggest this: "[T]he claim that having a child is epistemically transformative does not entail that it is also personally transformative: *for most people, it is*. For some people, it isn't" (2015, fn. 21, emphasis added). It might appear that, if true, this idea would rescue Paul's argument. After all, if prospective parents can reasonably expect themselves to be *personally* (and not just epistemically) transformed by an experience, then principles linking the non-phenomenal character of an experience with what they are like *right now* won't be any help at all. And so Paul's conclusion would appear to follow: prospective parents couldn't rationally decide in the ordinary, natural way. But there are two reasons why this appearance is misleading: the idea that parenthood is often—or, as we'll see, even *always*—personally transformative won't save the argument.[3]

First, if the probability that parenthood will be personally transformative is less than one, then in any particular case it makes sense to ask: What is the probability that *this* case of parenthood will be personally transformative *for this agent*? Whether an experience is or is not likely to be personally transformative is not a random affair. The deliberate decision to order catfish rather than trout is not likely to transform me into a different person. But it might: it might, for instance, if I thought of myself (prior to ordering the catfish) as someone who strictly adhered to traditional Jewish dietary laws forbidding the consumption of fish without scales. The decision to abandon these laws has a good chance of making me into a "different person" in the relevant sense. What this case illustrates is that whether or not a decision (and the experience that is its natural upshot) is likely to be personally transformative can itself be something about which we form reasonable expectations. You can see where this is going. We can form reasonable expectations about whether an experience will in fact be

[3] Thanks to an anonymous referee for pressing me to be clearer on this point.

personally transformative by thinking about what sort of person an agent is right now, and how experiences of the relevant sort affect people like that (i.e., whether these experiences are likely to be personally transformative for that sort of agent).[4]

The situation is even worse than it appears. Even if you think, implausibly in my view, that we can never safely predict *whether* some experience will be personally transformative, we have been given no reason to think that there are no further linking principles that tell us *how* an agent is likely to be personally transformed by an experience. That is, we have not been given any reason to deny the existence of principles that say how an agent that is thus-and-so *right now* is likely to be after the agent undergoes some experience. And we have positive reason for thinking there are such principles, for the same reasons we have for thinking there are the first sort of linking principles. How an agent is (likely to be) transformed by an experience supervenes on how the agent is right now and what the experience is like. Selfish cads do not become selfless altruists by adopting puppies, though they may become less selfish. Personal transformation may be commonplace, but Damascene conversion is not. So, even if we grant that we can't have any reasonable expectations regarding *whether* an experience will be personally transformative for an agent, we can form reasonable expectations about *how* an agent is likely to be transformed by it. And then we're back to the races: with these expectations in hand we can form expectations about the valence of the phenomenal character of the experience not for the agent as she is now, but for the agent as she is likely to be afterward.[5]

I said there were two problems with denying the existence of principles linking the non-phenomenal character of having a child with the valence of the phenomenal experience thereof. The first problem, as we just saw, was that denying the existence of such principles seems to commit one either to the (implausible) claim that having a child is always personally transformative, or to the (equally implausible) claim that we can't reasonably form expectations about whether some experience will be personally transformative or how it will be so. The second problem is that there

[4] After all, that's why it sometimes makes sense to say to some of the people you know that, say, reading a particular book will change their life, and why it never makes sense to say such a thing to every person you know.

[5] At the very least, this narrows the scope of Paul's argument to those decisions where two conditions are met: (i) it is plausible that there is a high (or certain) probability in this particular case that the decision will be personally transformative; and (ii) there is no way to reasonably form expectations about how an agent will be transformed. It might well be that the decision to procreate is sometimes like this for some people, but I seriously doubt that that the decision to procreate in general is like this. And while decisions other than prospective parenthood might also sometimes be like this for some people, I only somewhat more tentatively doubt whether there any interesting decisions to which Paul's argument is meant to apply (one's choice of career, one's choice of spouse) that are like this in general. (Thanks to an anonymous referee for suggesting this point.)

manifestly *are* such linking principles, and we know what some of them are. For just one example, depression on the part of either parent, but especially maternal depression, is linked to both affective and behavioral disorders on the part of children (Lovejoy et al. 2000; Tan and Ray 2005).[6] And parents of affectively or behaviorally disordered children report significantly higher rates of stress and lower levels of subjective well-being—as good a measure as any of the valence of the phenomenal character of their experience of what it is like to have a child (Tan and Ray 2005, 77). What this means is that if we know (1) what it is non-phenomenally like to have a child and (2) that some agent is depressed (or socioeconomically disadvantaged, see fn. 5), then we have at least some reason to expect that the phenomenal character of the experience of having a child will, for that agent, be negatively valenced. Of course, that reason to believe the phenomenal character of the experience of having a child will be negatively valenced might not be decisive. There might be further reasons to expect the experience will be negative that contribute to our expectation—or, indeed, there might be countervailing reasons, grounded in other linking principles, to expect that it will be positive. The point is just that such expectations are sometimes warranted. They are warranted on the grounds that, given knowledge of the non-phenomenal features of an experience and knowledge of what some particular agent is like, we can justifiably believe facts about how that experience is likely to phenomenally affect the agent. This is true in the case of the experience of having a child no less than it's true in the case of trying Vegemite for the first time. This should come as no surprise at all: what people are like helps determine how things turn out for them. And, thanks to years of pyscho- and sociological research, we can often safely predict how things will turn out for an agent given enough psycho- or sociological information about them.

Of course, the situation is no different when it comes to ourselves than it is in the case of others. Or at least, it is not relevantly different. Just as I can know that, given that some agent is depressed, the phenomenal character of her experience of having a child is unlikely to be positive, I can know of *myself* that, given I am depressed, the phenomenal character is unlikely to be positive. And so, *ceteris paribus*, I can safely predict that it would be unwise, just now at least, for me to have that experience.

Here is another way to put the same point. In the course of her argument, Paul claims that agents deciding whether to procreate cannot use reports of the phenomenal character of similar experiences garnered from other agents because those reports will not be able to impart the distinctive phenomenal character of the relevant experience. That's how having children for the first time is like tasting Vegemite for the first time—you can't know what it's like just from hearing about it from evangelical Australians. And

[6] Similar findings have (perhaps unsurprisingly, and probably relatedly) connected persistent poverty and overall socioeconomic disadvantage with cognitive, affective, and behaviorial disorders on the part of children. See McLoyd 1998.

since that distinctive phenomenal character is what matters to making the decision in a rational way, prospective parents can't rationally decide whether to have children of their own. In effect, what I've just argued is that, even granting that agents deciding whether to procreate can't access the distinctive phenomenal character of having a child of their own, they can form reasonable expectations about the valence of that experience given the existence of empirical principles that link the sort of people they are now to the sort of phenomenal experiences they are likely to have if they have a child of their own.

4 Concluding Remarks

The upshot of the argument, then, is this: Paul is quite correct to highlight the epistemically transformative nature of the experience of having a child. Prospective parents cannot know what it is phenomenally like to have a child of their own before they do so, just as prospective diners cannot know what it is like to taste Vegemite before they do. As Paul points out, this means that prospective parents cannot rationally decide to have a child by reflecting on the phenomenal character of that experience: it's in principle epistemically inaccessible to them. But this does not mean that prospective parents cannot rationally decide to have a child by reflecting on what it is like to have a child. It just means they have to take a somewhat circuitous route: prospective parents must reflect on the non-phenomenal features of the experience, on what they themselves are like, and on the principles that link how they are to how the experience is likely to affect them. By doing so, prospective parents can form rational expectations about how the experience of having a child is likely to phenomenally affect them, and so can form rational expectations about the phenomenal valence of that experience. This means that, given Paul's assumption (which, here at least, I grant) that the phenomenal character of the experience is really the final arbiter of whether or not it is rational to decide to have a child, prospective parents can rationally decide whether or not to procreate.

Although Paul's argument focuses on the decision to procreate, it is not limited to that decision. If correct, Paul's argument might apply equally well to the decision to change careers, start a new hobby, engage in a new romantic relationship, or become interested in a new cuisine. What my argument shows is that Paul's argument doesn't work in the case of the decision to have a child because we think there are linking principles that can help us form reasonable expectations about how experiences will phenomenally affect us. But, I think, the same goes for these other areas of decision: they too have linking principles. For example, if you know what philately is non-phenomenally like, and you know pretty well what *you* are like, then you can form reasonable expectations about how the distinctive phenomenology of philately (if there is one) is likely to strike you. (I, for one, am pretty sure I would not like it.)

Let me close by highlighting two features of the view I've defended here. First, it can still be useful for agents to reflect directly (i.e., not via linking principles) on what they think some new experience will be like. This is because when a childless agent reflects on what it would be like to have a child, even if this reflection is epistemically unreliable, it sometimes has a certain valence—the agent might experience the reflection itself as overall positive, overall negative, or somehow mixed. And even if this reflection has little to do with what in fact it will be like for the agent to have a child of their own (that's the hinge on which Paul's argument tries to turn), it can still have important evidential value. This is because it helps reveal the antecedent attitude the agent has toward the experience. And the antecedent attitude an agent has toward an experience is part of who she is—and that, as we already know, will affect how the experience will strike her. And so reflecting on what a new experience might be like can still play a role in rationally deciding whether to undergo that experience. Not because it can provide good information about what the experience will actually be like, but because it can provide good information about what we ourselves are like.[7]

Second, note that the account I've given here of how it can be rational to decide under conditions of uncertainty about the phenomenal character of a new experience (e.g., having a child of one's own) by reflecting on what it is like to do so squares very nicely with our practice of giving and asking for advice when deliberating about whether to take the plunge. For instance, when we want to know whether it is rational for us to have a child, we not only ask people who have had children what it is like, we ask people who we think are a lot like ourselves in relevant respects (e.g., people with comparable socioeconomic status or similar values).[8] This is presumably because we think that the way the experience affects others is a good guide to how it will affect us: we think there are principles that link the way people are (not just socioeconomically, but also in terms of their values, commitments, cares, and so on) with how experiences affect them. And so we can use others' experiences to guide rational expectations about the value—or at least the valence—of the phenomenal character of the experience we are deliberating about whether to have.

Nathaniel Sharadin
E-mail : sharadin@unc.edu

References:

Bartlett, Tom. 2013. "Maybe You Should Have a Baby." *The Chronicle of Higher Education*. http://chronicle.com/blogs/percolator/maybe-you-should-have-a-baby/32379.

[7] I'm grateful to an anonymous referee for this point.
[8] Compare Harman 2015, 328.

Acknowledgements Thanks to Finnur Dellsén, Megan Mitchell, Kate Nolfi, Wesley Sauret, and two anonymous referees for their helpful feedback.

Burkeman, Oliver. 2013a. "This Column Will Change Your Life: Transformative Experiences." *The Guardian*. http://www.theguardian.com/lifeandstyle/2013/apr/06/this-column-change-life-transformative-experiences.

Burkeman, Oliver. 2013b. "We're Truly in the dark when it comes to life's most important decisions." *Business Insider*. http://www.businessinsider.com/the-problem-with-advice-2013-4.

Dougherty, Tom, Sophie Horwitz, and Paulina Sliwa. 2015. "Expecting the Unexpected." *Res Philosophica* 92 (2): 301–321. http://dx.doi.org/10.11612/resphil.2015.92.2.5.

Gopnik, Alison. 2013. "Is It Possible to Reason About Having a Child?" *The Wall Street Journal*. http://online.wsj.com/news/articles/SB10001424127887324432404579052901271445142.

Harman, Elizabeth. 2015. "Transformative Experience and Reliance on Moral Testimony." *Res Philosophica* 92 (2): 323–339. http://dx.doi.org/10.11612/resphil.2015.92.2.8.

Lewis, David. 1990. "What Experience Teaches." In *Mind and Cognition: A Reader*, edited by William Lycan, 499–519. Oxford: Blackwell.

Lombrozo, Tania. 2013a. "Is Having a Child a Rational Decision?" *NPR*. http://www.npr.org/blogs/13.7/2013/03/11/173977133/is-having-a-child-a-rational-decision.

Lombrozo, Tania. 2013b. "Is it Rational to Have a Child? Can Psychology Tell Us?" *Psychology Today*. http://www.psychologytoday.com/blog/explananda/201303/is-it-rational-have-child-can-psychology-tell-us.

Lovejoy, M. Christine, Patricia A. Graczyk, Elizabeth O'Hare, and George Neuman. 2000. "Maternal Depression and Parenting Behavior: A Meta-analytic Review." *Clinical Psychology Review* 20 (5): 561–592. http://dx.doi.org/10.1016/S0272-7358(98)00100-7.

Marshall, Richard. 2013. "Metaphysical (Interview with L.A. Paul)." *3:AM Magazine*. http://www.3ammagazine.com/3am/metaphysical/.

McLoyd, Vonnie C. 1998. "Socioeconomic disadvantage and child development." *American Psychologist* 53 (2): 185–204. http://dx.doi.org/10.1037/0003-066X.53.2.185.

Moran, Collete. 2013. "The Decision Whether to Have Children Cannot Be Made Rationally." *The National Review*. http://www.nationalreview.com/home-front/358115/decision-whether-have-children-cannot-be-made-rationally-colette-moran.

Paul, L. A. 2015. "What You Can't Expect When You're Expecting." *Res Philosophica* 92 (2): 149–170. http://dx.doi.org/10.11612/resphil.2015.92.2.1.

Prescott, John, Nini Ripandelli, and Ian Wakeling. 2001. "Binary Taste Mixture Interactions in PROP Non-tasters, Medium-tasters and Super-tasters." *Chemical Senses* 26 (8): 993–1003. http://dx.doi.org/10.1093/chemse/26.8.993.

Rothman, Joshua. 2013. "The Impossible Decision." *The New Yorker*. http://www.newyorker.com/books/page-turner/the-impossible-decision.

Tan, Susan and Joseph Ray. 2005. "Depression in the young, parental depression and parenting stress." *Australasian Psychiatry* 13 (1): 76–79. http://dx.doi.org/10.1080/j.1440-1665.2004.02155.x.

Change Your Look, Change Your Luck: Religious Self-Transformation and Brute Luck Egalitarianism

Muhammad Velji

Abstract: My intention in this paper is to reframe the practice of veiling as an embodied practice of self-development and self-transformation. I argue that practices like these cannot be handled by the choice/chance distinction relied on by those who would restrict religious minority accommodations. Embodied self-transformation necessarily means a change in personal identity and this means the religious believer cannot know if they will need religious accommodation when they begin their journey of piety. Even some luck egalitarians would find leaning exclusively on preference and choice to find who should be burdened with paying the full costs of certain choices in one's life too morally harsh to be justifiable. I end by briefly illustrating an alternative way to think about religious accommodation that does not rely on the choice/chance distinction.

In Québec, at the end of 2013 and start of 2014, the then elected separatist party proposed a bill initially called the "Québec Charter of Values."[1] It would have effectively banned religious symbols such as all forms of the Islamic veil, the Sikh turban, the Jewish kippah, large crosses, and other "conspicuous"[2] religious symbols from being worn by public servants. A constant question that showed up in the discourse surrounding the charter was why some minorities received exceptions from laws and some did not. One justification for differentiating between who got exemptions and who did not was that certain minorities, such as the disabled, were unfortunate in that they had not chosen their disability and therefore deserved accommodation. On the other hand, religious people chose their religion and the form their religious practice takes and therefore needed

[1] "Bill nº60: Charter Affirming the Values of State Secularism And Religious Neutrality And The Equality Between Women And Men, And Providing A Framework For Accommodation Requests," http://www.assnat.qc.ca/en/travaux-parlementaires/projets-loi/projet-loi-60-40-1.html.

[2] "5. In the exercise of their functions, personnel members of public bodies must not wear objects such as headgear, clothing, jewelry or other adornments which, by their conspicuous nature, overtly indicate a religious affiliation."

Res Philosophica, *Vol. 92, No. 2, April 2015, pp. 453–471*
http://dx.doi.org/10.11612/resphil.2015.92.2.13

no accommodation but should conform like everyone else to general laws. These questions of fortune and misfortune also showed up in discourse around the French ban on the veil and bans proposed for religious animal slaughter in Denmark.[3] Specifically in townhall meetings that Charles Taylor and Gerard Bouchard conducted in Québec to get a sense of the public's opinions on how minorities in Québec were to be handled, this theme was present:

> During our consultations, a number of participants called into question the legitimacy of accommodation requests for religious reasons. The rightfulness of an adjustment that allows, for example, a female or a male student to wear a headscarf or a kirpan, respectively, is not obvious to everyone. Similar exemptions may be granted for health reasons: a young girl must cover her head on her physician's orders or a diabetic child must bring a syringe and a needle to school. No one would dream of objecting to such exceptions. We also know that accommodation aimed at ensuring the equality of pregnant women or the physically disabled is readily accepted. Québec (and Western) public opinion thus reacts much more harshly to requests motivated by religious belief. One of the most frequent arguments put forward to explain why requests justified by religious reasons and those motivated by health reasons cannot be put on an equal footing is that *individuals who are disabled or sick have not chosen their condition while believers appear to have a choice between renouncing their religion or reinterpreting it in a manner that makes accommodation requests superfluous*. (2008, 143, emphasis added)

This position, linking accommodation and choice, can be found represented in a particular branch of anglo-american political philosophy called "Luck Egalitarianism." Although often linked only to just distribution patterns rather than the issue of minority religious accommodation, it has, since the 1970s, become a sophisticated position that attempts to appeal to both the left and the right by being egalitarian yet also sensitive to responsibility. To many liberals it seems intuitively right that a gambler who squandered all their money should have a weaker entitlement to claim benefits than someone who was born into poverty. The reason for this is that the gambler is presumably more responsible for their own deprivation. To Richard Arneson, responsibility plays a fairly straightforward regulatory role in shaping people's entitlements. If someone is responsible for their own deprivation then they and not anyone else should suffer the burdens associated with that deprivation, otherwise "some individuals [who] behave

[3] See Valenta 2012.

culpably irresponsibly, again and again, [will end up] draining resources that should go to other members of society" (Arneson 2000, 349).

My intention is to reframe the practice of veiling as an embodied practice of self-development and self-transformation and argue that practices like these cannot be handled by the choice/chance distinction. Embodied self-transformation necessarily means a change in personal identity and this means the religious believer cannot know if they will need religious accommodation when they begin their journey of piety. Even some luck egalitarians would find leaning exclusively on preference and choice to find who should be burdened with paying the full costs of certain choices in one's life too morally harsh to be justifiable. I end by briefly illustrating an alternative way to think about religious accommodation that does not rely the choice/chance distinction.

It must be emphasized that my goal here is not to undermine *all* luck egalitarian positions; my aim here is to attack the popular intuition that the distinction between chance and choice is morally relevant to broad debates about multiculturalism as a normative ideal and minority religious accommodation. Along with becoming more sophisticated, luck egalitarianism has multiplied into a spectrum of philosophical positions.[4] Many of these positions will not be touched by my argument either because some argue that luck egalitarianism only applies to economic, distributive justice (Tan 2008, 670), while some others rely on the choice/chance distinction in talking about minority accommodations but think that strong valuation choices are not choices but chance (Cohen 2004) and finally some luck egalitarians think that religious practices are choices, yet luck egalitarianism should be considered a *pro tanto* theory that can be overruled by other values, such as burdens that are too costly (Tomlin 2013, Knight 2009). We will come back to this third position later when dealing with the criticism that people should be responsible for actions that they identify with.

The view my argument wishes to reach are those that embody best the intuition illustrated by those in the Québec town hall meetings. This view Peter Vallentyne calls Brute luck egalitarianism (2008, 58) and Patrick Tomlin calls 'canonical' luck egalitarianism[5] (2013, 395). This view holds simply that those states and events that the agent could not deliberately influence should be equalized or accommodated but that the effects that are attributable to the agent's choice[6] need not be (Vallentyne 2008, 58).[7]

[4] I am grateful to the anonymous referee who pushed me to deal with this multiplicity of views.

[5] I will use both these terms interchangeably.

[6] Many canonical luck egalitarian theories will replace the word "choice" with "responsibility," while Dworkin and Barry will replace "choice" with "preference."

[7] A perfect example of this canonical view is Eric Rakowski who denies Cohen's portrayal of expensive tastes that were cultivated when young as involuntary. Before these beliefs became deep beliefs, and *even if* these beliefs are deep, if the agent "engendered this interest and permitted it to become pronounced, aware of the costs . . . then it seems only right that [they] should answer for [their] choice" (Rakowski 1991, 56). He concludes that "to the extent that

My critique then is not just of this one luck egalitarian view, but also strong responsibility-sensitive views of equal opportunity such as Dworkin and also Brian Barry[8]. In a section of Barry's book *Culture and Equality*, he gives a sustained argument against giving minority religious accommodations by arguing against Biku Parekh's (and one would imagine Cohen's) position that we should consider religious practices as involuntary. He does this by appealing to the same intuitions that those in the Québec town halls made, by making a comparison to a group he does think should be given accommodations because of brute luck, the disabled. Barry is convinced that

> the position of somebody who is unable to drive a car as a result of physical disability is totally different from that of somebody who is unable to drive a car because doing so would be contrary to the tenets of his or her religion. To suggest that they are similarly situated is in fact offensive to both parties. Someone who needs a wheelchair to get around will be quite right to resent the suggestion that this need should be assimilated to an expensive taste. And somebody who freely embraces a religious belief that prohibits certain activities will rightly deny the imputation that this is to be seen as analogous to the unwelcome burden of a physical disability. (Barry 2001, 37)[9]

My task in this paper is to answer the canonical luck egalitarian and Barry's challenge without falling into the counter-intuitive explanations that Parekh and Cohen provide about the involuntariness of cultural practices. My argument is a move to look beyond the choice/chance distinction rather than merely moving the cut between whether religious practices are a choice or chance toward the latter position.

1 Diachronic Critiques of Canonical Luck Egalitarianism

Before I begin my own argument about self-transformation and responsibility, I would like to bring out an argument that will remain implicit

people elect to expose themselves to, preserve, or suppress certain desires, the more or less expensive preferences they develop are beyond the bounds of justice: no correction need or should be made for them" (1991, 57).

[8] In his article on Barry's final book, *Why Social Justice Matters*, Arneson summarizes Barry's strong responsibility-sensitive views of equal opportunity. "In the language of personal responsibility, Barry's view is that, if people start with equal opportunities and some voluntarily undertake courses of action from this equal starting point that leave them worse off than others, the loss that falls on the individual in consequence of such voluntary choice is her responsibility. It is not the responsibility of society to make good the loss" (Arneson 2007, 397). I am grateful for the anonymous referee who pushed me to make a better connection between Barry and the Brute luck egalitarian position I am critiquing.

[9] I will not critique here the ableist assumptions of this kind of example.

throughout my paper: that canonical, "static," luck egalitarianism already has a problem coming to terms with diachronic aspects of responsibility over a lifetime. Clare Chambers points out that while canonical luck egalitarians pour through the histories of individuals trying to parse what in their lives is chance and what choice, certain choices have their inegalitarian effects in the future not in the past (2009, 376). At some point in a person's life there is a moment that is just assumed by the canonical luck egalitarian as the point where choices should no longer be compensated for. Chambers calls this a Moment of Equal Opportunity (MEO). Chambers shows that present choices amplify their impact on the chooser's life. Correctly chosen big decisions open more opportunities while choosing wrongly, both relatively to choosing right and in an absolute sense, closes more and more opportunities. This is done in a way that disproportionately burdens the agent who made the initial choice. Instead MEOs must be done many times over a lifetime.[10] Concretely, Chambers does not see how this could happen in practice. Chambers describes the dilemma thusly, "it is not at all clear how equality of opportunity *can* be applied throughout a person's life, since doing so poses serious problems of epistemology, efficiency and incentives, and leads to counter-intuitive results . . . *theories of equality of opportunity are inconsistent if they support [only one] MEO and unrealizable if they do not*" (2009, 378). For canonical luck egalitiarians who are concerned about giving religious accommodations to minorities, there is really only one MEO and that is when the woman chose to veil or the man chose to wear a turban rather than a motorcycle helmet. But as Chambers points out, this hides, both, that over a life time many MEOs should be considered and also that the amplification of the cost of a single choice over time is problematic for an egalitarian theory.

Chambers has shown, as time moves forward, the costs of taking responsibility of a present choice becomes disproportionately large. Patrick Tomlin[11] takes this and combines it with another argument about personal identity. He argues that over time, responsibility can diminish just by the fact that people should not be held responsible for their choices forever (Tomlin 2013, 400). This argument is linked to an idea of personal identity over time. "Identity isn't enough to acquire responsibility, 'suitable reflectiveness' of agency is required too. If I am responsible because I am related to the act in a certain way then I don't see why I should be thought to be responsible at some later time unless I am still related to the act in the relevant way. If a person has changed such that whatever it was that made the action suitably reflective of their agency at the time has diminished or disappeared, then it seems plausible to think that this kind of change will diminish or extinguish responsibility" (403). This makes each present choice doubly problematic for Brute luck egalitarians: at the same

[10] This is something Dworkin makes clear in his critique of "starting gate theories" (1981, 310) yet as I will get to, Chambers still has him in a dilemma.

[11] I am grateful for the anonymous referee who suggested this article to me.

time as the burdens of choosing unfairly amplify as we travel forward in time, the agent's responsibility for that action diminishes since that person's continuity with the person who made that initial choice diminishes.

This diachronic critique will be in the background as I begin my own argument against static canonical luck egalitarianism. But I do not take on all of Tomlin's argument since Tomlin (unlike Chambers) attempts to keep the choice/chance distinction by proposing a "dynamic luck egalitarianism" (Tomlin 2013, 400). There is an important difference in my argument from Tomlin's personal identity argument. This difference is that I will not be considering the case of just *any* decision that predictably and gradually becomes less of a responsibility as a person gains temporal distance from the decision. My case is a case of pious self-transformation where the link of the person before and after the transformation is strikingly different, almost a break in personal identity such that it is unpredictable, in a much shorter time, how much responsibility the pious believer has.

2 Moving from a Third Person to a First Person View of Agency

The big theoretical shift I would like to introduce before moving on to my argument about self transformation is a move from a third person view to the first person view of agency. Bernard Williams has a contrasting view of responsibility than the one that brute luck egalitarians use in order to support their choice/chance distinction. The objective, third person way of looking at luck that scaffolds brute luck egalitarianism, Williams calls "incident luck." The first person, agentic way of looking at luck he calls "constitutive luck." Williams finds that incident and constitutive luck problematize morality in two different ways. Incident luck undermines the idea that we can always determine before we act, which of our choices are justifiable. Constitutive luck undermines the assumption of equality regarding our capacities for moral agency (Williams 1981, 21). When looking at actions and practices of a person, luck egalitarians reflect a concern about luck's threat to autonomy. Williams, however, concentrated more on character and agency. He was more concerned about threats to a person's integrity. Considerations involved with the concept of integrity involve consistency, coherence, and commitment. Whereas for luck egalitarians, autonomy involves considerations of independence and avoiding the contamination of heteronomy. Williams's skepticism regarding the advisability of planning in advance for one's life as a whole turns on the vulnerability of the luck of our very identity. Because who we become is not immune to luck, our knowledge from now of what will be in our interests in the future is limited. Contingencies of our development that are inaccessible at the moment of making critical choices threaten our integrity and interfere with our carrying through on obligations and commitments. The problem of agency and integrity are such that despite the admitted contingencies and luck of our constitution, we still cannot help but feel that we should not

betray commitments central to our identities. This different conception of agency makes it not only impossible to separate brute luck aspects from option luck aspects of action, but makes this separation morally irrelevant. This concentration on constitutive luck is grounded in the ethical theories of ancient philosophers such as Aristotle. What Williams, but also many religious traditions, inherit from the ancient conception of agency is that we are not born responsible but have, at most, potential for becoming agents. As Claudia Card points out, this agency is realizable to a greater or lesser extent with luck and hard *work* (1996, 24).

I emphasize this shift from backward third person to forward first person looking responsibility because when brute luck egalitarians think about religion, they look at it from the objective third person view. Additionally there is also a tendency to interpret the habitual, collective, embodied practices of religious devotion of those influenced by this first person, agentic conception as Protestantized, individual, duties of conscience. By defining religion as a matter of belief or faith, a tradition comes to be treated as "a cognitive framework, not as a practical mode of living, not as techniques for teaching the body and mind to cultivate specific virtues and abilities that have been authorized, passed on, and reformulated down the generations" (Asad 2001, 216). When religion is treated in this cognitive way, every religious believer then has complete access to their belief and can choose which among these beliefs conform to the state's general law. For example, Barry argues that the state should not accommodate ritual slaughter since if "faced with a meatless future, some Jews and Muslims may well decide that their faith needs to be reinterpreted so as to permit the consumption of humanely slaughtered animals" (Barry 2001, 35). Reinterpretation here is construed as an act of autonomous will that all religious believers exercise.

Contrast this view to another way of looking at religious practice, grounded in ancient philosophy. Michel Foucault argues that this ancient conception of subjectivity, reappropriating a term from Pierre Hadot (1981) called "spirituality," is the practice or exercise through which the subject carries out the necessary transformations on themselves in order to have access to religious and cultural agency. This access to religious subjectivity is not given to the subject by right. Self-transformation, self-development, modification of one's existence and to some extent becoming other than oneself are the "price to be paid" (Foucault 2005, 15) for this access. As Card argues as well, subjectivity is a kind of work. This is a work of the self on the self for which one takes responsibility in a long labor of *askesis* (religious practice of self-discipline) (Foucault 2005, 16) but not necessarily in knowing what exactly the outcome of this labor will be.

Brute luck egalitarians who argue against minority religious accommodations are correct that religious practices are not involuntarily compelling like coughing when you are sick. But in order to secure certain religious accommodations, those practicing veiling, wearing a turban, carrying a

kirpan, and even wanting minarets in Switzerland have had to take up a type of discourse that makes religion a matter of brute luck. This reifies religion as monolithic, objective, and imposed. As Susan Mendus (2002, 34) argues, in order to reply to attacks by luck egalitarians, thinkers such as Biku Parekh have had to make religious practice, while not entirely beyond human control, sufficiently intractable. If we are to understand why religious people demand accommodation, we cannot just think of religion as a series of imperatives. The reasons that certain Muslims have for asking for accommodations for prayer at their jobs or schools cannot just be reduced to codified rules. It assumes that religions lay down certain binding rules and that the exercise of religion consists only of obeying those rules. Douglas Laycock observes bitingly, "it is as though all of religious experience were reduced to the Book of Leviticus. It is the view of religion held by many secularized adults, who left the church in their youth after hearing much preaching about sin and failing to experience any benefits" (1990, 24). The pietists, like the ancient Greeks, conform to particular norms not because they are obliged by universally recognized laws to do so, but because they aspire to a particular *telos* or ideal of self: the pious self. So we can say that the pietists are engaged in practices of self-creation through particular ways of inhabiting norms (Weir 2013, 131).

For some Muslim women, veiling is an unavoidable means to the particular end of being pious. Veiling is not the end in itself. What these women are claiming is that by not allowing them to veil, the state is frustrating a larger goal of transformation and the ability to practice their religion beyond the bare minimum. Saba Mahmood, an anthropologist who has studied veiled women in Egypt, compares the practice of veiling to a pianist who submits herself to the often painful regime of disciplinary practice, as well as to the hierarchical structures of apprenticeship, in order to acquire the ability and requisite agency to play the instrument with mastery. Her agency is predicated upon her ability to be taught, a condition classically referred to as docility (Mahmood 2005, 29). What is considered suffering under the veil actually enables certain capacities that can be exercised, for some women at least, in no other way than through the veil. One cannot simply argue that those women who choose to veil should find another way. The veil is a critical marker of piety and the ineluctable means by which she trains herself to be pious. While wearing the veil serves at first as a means to tutor herself in the attributes that make up piety, it is also simultaneously integral to the practice of piety: "one cannot simply discard the veil once a modest deportment has been acquired, because the veil itself is part of what defines that deportment" (158). The veil is not a mere means; it is, instead, constitutive of becoming a pious person. Piety is not a finished state, but a continuing activity. If we take the goal of the woman who veils to be a transformative activity, then taking away her veil destroys her ability to concretely become the person she chooses to be through carrying out those actions that express her own purposes and needs.

Mahmood suggests that Muslim women, regardless of whether they veil or not, in practicing to become a pious Muslim, create religious desire through a set of disciplinary acts like athletes train their body. That is to say, desire in this model is not antecedent to, or the cause of moral action, but its *product* (2012, 231). Through the use of the veil, the goal (piety) is also one of the means by which desire is cultivated and gradually made realizable. In this Aristotelian model of ethical pedagogy, external, performative acts like veiling are understood to create corresponding inward dispositions. The way the veil creates this inward disposition is through habitus. Habitus, in this older Aristotelian tradition, is understood to be an acquired excellence learned through repeated practice until that practice leaves a permanent mark on the bodily character of the person (Mahmood 2005, 136).

3 The Problem that Self-Transformation Poses for Luck Egalitarianism

With this, more embodied, view of the practice of veiling, it becomes harder to argue that embarking on the labor of piety is like taking a gamble where the believer "loses" if their journey leads them to a religious practice that runs afoul of general laws. Some people may go through the self-development required to access religious agency and yet will not need to veil, wear a turban, or need special accommodation to go to a mosque on Fridays. Yet inevitably there will be others whose self-transformation calls upon them to do one of these practices that the state does not want to accommodate. This is to say that, before they began their journey toward a more pious subjectivity, their identity might not have been complete by wearing a veil, yet somewhere along the way, they changed so much that wearing a veil turned from an option to something much more mandatory.[12]

What I am describing here falls somewhere between what Edna Ullmann-Margalit calls a "conversion" and "drifting" toward a big decision (2006). There are two ways that the self-transformative characteristics in becoming a pious subject affects the luck egalitarian critique of minority religious accommodations. First, since piety in Islam is about training bodily habit, like drifting to a big decision, it is a subtly incremental process. It is a process such that, although the veiling subject is agential, she does not know in what way her piety will lead and whether she will end up taking up the practice of veiling in her journey toward piety. This is because her transformation will be so great, nothing other than going through and

[12] Mayanthi Fernando, an anthropologist who works with veiled women in France after the veil ban reports that "the temporal gap between beginning to pray and beginning to veil was common to most of the practicing Muslim women I knew, who prayed regularly for months and sometimes years before putting on the headscarf. Such a gap highlights the intellectual and bodily disciplinary process through which these young Muslim women worked on themselves by undertaking one step in a series of necessary practices to induce the desire for the next step toward becoming a properly pious Muslim" (Fernando 2010, 25).

experiencing this transformation will be adequate for her to know whether her piety will or will not include veiling. Secondly, the process of bodily self-transformation is like a conversion in that it has an irrevocable quality to it.

Luck egalitarianism places such weight on the distinction between choice and chance because it assumes that the choice to become pious and then to veil fits certain paradigmatic decisional procedures that weigh the value of one's future experiences. Allowing accommodations only for those actions judged not to be a choice is supposed to disincentivize people attempting to make themselves exceptions to the law. They should not be accommodated if they choose their "expensive" lifestyle, and are therefore asking for more than their fair share. For a brute luck egalitarian, not being the exception to a general law should be an integral part of making the decision to become pious. To take into account the law of secularism in the public sphere and still decide to veil is considered irrational or selfish and so the individual who chooses to veil must take responsibility for their actions. The problem with this assumption is that the agent making this kind of decision is not in the epistemic condition to make this decision until after the process is over. So holding them responsible to the point of punishment is not responsibility-sensitive in the way any luck egalitarian would want.

The problem is of the epistemically impoverished starting position of anyone who would like to begin to be pious. There are two reasons for this, first is the minutely incremental nature of becoming pious and the second is the problem that piety cannot be known without going through the process of becoming pious. The decision to become a pious Muslim is not necessarily like a conversion as described by Ullman-Margalit. It is not always an instantaneous gestalt switch where one is "blinded by the light of the compelling new truth" (Ullmann-Margalit 2006, 162). This gives the process too much of a cognitive, Protestantized aspect. In reality, since it is about training the mundane, everyday habits, it is more like Ullman-Margrit's idea of "drift" decision making (170). It is only from the retroactive perspective that one could see that the self-transformation undertaken has not just been one of minute degrees, but taken altogether is a change of kind and quality (rather than quantity). Becoming pious is incremental in nature and is the continual activity of a series of small mundane decisions to change certain everyday habits without any single stage ultimately being the one where the pious woman decides to either wear or not wear the veil. The problems that Brute luck egalitarianism has with diachronic aspect of responsibility comes back to haunt it since with this kind of "drift" decision, there is no MEO to be identified.

The problem of experience as it is related to transformative practices is illustrated well by L. A. Paul with the example of pregnancy. Paul argues that "what it's like" knowledge, such as the phenomenal knowledge a person who had never seen color might experience when seeing red for the first time, is a kind of knowledge only accessible via experience (2015, 6).

In deciding to have a child, the mother does not know the phenomenal feeling of this experience. She does not know "what emotions, beliefs, desires, and dispositions will be caused by what it's like for her to" raise a child (7). Even if she has tried to babysit and gain experience with children in an attempt to simulate this experience and she thinks she will feel joy, there is still the lacuna in her knowledge that she still does not know what it is like to experience feeling the joy while raising the child until she actually goes through that process.

This point is further complicated in the case of religious self-transformation. In the case of pregnancy, there is an epistemically transformative experience in having and raising a child that may also include a personally transformative experience. This personal transformation is only incidental, though.[13] For example, some parents, when they experience the epistemic transformation of raising a child may realize they do not need to change themselves. This could be because they can afford to pay others for the labor it takes to raise a child, so their activities and routines may go unchanged and they may be the same people after as they were before. But it is *in the nature* of working on religious piety that the self cannot remain the same. In this case, piety involves not changing to find a true "I" that was always present but dormant within, but to transcend the "I," to become different than the "I" that was (Mahmood 2005, 148). The types of bodily and emotional work one must go through to progress toward the character of a pious person involves working on one's desires. One's actions and decisions do not come from natural feeling, but instead they *create* them through training one's habits, memory, desire, and emotions (Mahmood 2005, 157). As I have argued, religion is not just a series of clear imperatives and so what one will become through self-transformation is not dictated by consulting holy texts or a religious leader. On what path she may end upon is not known at the beginning. This point is not restricted to women who veil. For instance, this type of experience of training one's self to transform is reported by Cressida Heyes in her own experience through yoga. She describes how yoga pushed her to the edge of her physical capacity while also pushing her through emotional pain, often experiencing innumerable rounds of violent sobbing. Through this bodily self-discipline, she felt herself change in unexpected ways, especially since yoga "isn't charted in the way that normalized discipline is: there are no leaflets or narratives or diagnoses waiting to tell me who I am and what will happen next" (Heyes 2007, 129).

This kind of bodily self-transformation constituently involves a discontinuity of the self, a change in one's beliefs, desires as well as one's cognitive and evaluative systems. This can change your personal phenomenology in deep and far reaching ways, decentering what beliefs and preferences you may have had with very different ones. This brings out the problem

[13] Nathanial Sharadin (2015, 7) comes to this conclusion as well.

that these kinds of big decisions hold for Brute luck egalitarianism. How does one evaluate this kind of choice rationally? The problem for luck egalitarians with this argument is that it is different from the problem of experience in that it is not about new knowledge about the world but that we probably will not know our future personality. As Williams argued, this is the problem of constitutive luck and integrity. There must necessarily be difficulty in trying to decide for the future person you will be since, if one is training oneself to become pious correctly, there will be no continuity with that person. Yet whether that future person had a choice and is therefore responsible is predicated on the continuity of the person making the decision. This can be illustrated by Ullman-Margalit's story of the person who hesitates to have children because they do not want to become the boring type of personality he or she encounters in people who have had children. Yet after the experience of having children, this same person approves of their new, boring personality (2006, 167n10). How do we evaluate this? If there were no child, this person would not have the new preferences, yet in having the child the old preferences seem invalid from a second-order perspective. It is not that making these kinds of self-transformative decisions is irrational since we have no clear path as to what the rational procedure would be instead. Even Brute luck egalitarians will concede that being merely causally responsible for an outcome like one billiard ball hitting another is not sufficient for agential responsibility, since the agent may reasonably have been unaware that her choice had the effect in question. One may be responsible for the foreseeable causal effects of one's choices, but one is not agent-responsible for all the causal effects of one's actions (Vallentyne 2008, 58). An agent is not broadly agent responsible for an outcome if there was no way she could have known her choice would produce the outcome since this affects an agent's disposition to choose and can thereby affect the baseline for the allocation of responsibility (Vallentyne 2011, 178). Being agent-responsible is the kind of responsibility that must be focussed on when talking about minority religious accommodations since it is the one that best justifies why an agent should be forced to carry the burden of responsibility of an action.

To expect the religious agent to not take the necessary first steps in being pious because there is a chance she may or may not veil is to ask this agent to forgo everything but the *minimum* in practicing their religion "just in case" she may find imperative the need to wear something religiously ostentatious in the public sphere. For many believers, the attempt to distinguish what is required from what grows organically out of the religious experience is an utterly alien question. In most faiths, serious believers rarely concentrate their efforts on identifying the minimum that God requires (Laycock 1990, 26).

Finally, becoming pious is also "valvic" in the sense that there is an irrevocability to it. Once one puts many years of practice into piety, it is very difficult to lose the subconscious habits one has cultivated and

replace them with new ones. One's habits are tied to everyday routines and the self-transformation involved in piety will alter the nature of all your relationships with others, with yourself, and with the world, in all of the practices of your daily life (Weir 2013, 133). And once one practices piety with a veil, this "training" one puts oneself through is not just about body learning but about learning a new body sense. As one of Mahmood's veiled subjects attests, while before she may have been relaxed with her hair showing, her "body literally comes to feel uncomfortable if [she does] not veil" (Mahmood 2005, 157). The *telos* of this bodily training is such that veiling should "attain the status of an almost physiological need that is to be fulfilled without conscious reflection" (139),[14] which would explain the uncomfortable feeling when unveiled in public. This is why we must take seriously, and not assume that it is hyperbole, when we read about the Québec woman who states when asked how she would feel if a law forced her to take off her veil: "ce n'est pas banal . . . ça fait partie intégrante de moi . . . Si on me l'enlève, c'est comme si on m'amputait."[15] There is also a normative aspect to this irrevocability. It is a conversion process in the sense that the now pious individual will look back upon their previous life in a negative light.

4 An Objection Based on Responsibility as Identifying with One's Preferences

While I think my argument is effective against canonical luck egalitarianism, there is still one objection open to those with strongly responsibility-sensitive views of equal opportunity such as Dworkin and also Barry who do not identify with luck egalitarianism. Dworkin and Barry get around epistemic problems that I have just raised by arguing that we can skip the obsessive parsing of choice and chance in the history of individuals. According to Barry, "people are responsible for their preferences whenever they are content with them. How these preferences originated is irrelevant, and the ease with which they could be changed is relevant only in this way: that we would have to question the sincerity of your claim not to want to have the preferences you actually do have if it were easy to have the preferences you actually do have if it were easy for you to change" (1991, 156).[16] Barry finds religious belief exemplary of this principle. He thinks those in lower economic classes or with disabilities are not content and would prefer not to be in that class or have that disability and are therefore

[14] This particular quote is in reference to daily prayer, but Mahmood makes it clear that this applies to wearing the veil as well.

[15] "It is not trivial . . . [The veil] is an integral part of me . . . If one were to remove it, it would be like an amputation." http://www.lapresse.ca/le-soleil/opinions/chroniqueurs/201309/19/01-4691194-vous-etes-comme-des-religieuses.php

[16] Dworkin expresses this argument very simply and in a negative way, "a "taste" is a handicap [due accommodation and compensation] if one would prefer not to have it" (2004, 392n31).

not responsible and should be allowed compensation or accommodation. But the case of a religious believer is very different since as long as you continued to be a religious believer "you could hardly complain that it was bad luck to have the preferences you had, since you would not have wished things to be any different" (Barry 1991, 157).

The problem with this idea of identifying with one's preferences causing certain responsibilities being demanded of an individual is that we do not know the scope of this responsibility. Serena Olsaretti points out that it is just assumed by luck egalitarians that preference and responsibility are tightly connected concepts that entail each other. When people with strong responsibility-sensitive views talk about their commitment to holding individuals responsible, they neglect to say *what*, precisely, they are committed to holding individuals responsible for. Instead, they imply that it is self-evident what the consequences of people's choices and actions are and which ones they could justifiably be held responsible (Olsaretti 2009, 169). In order to do this, we minimally need knowledge of counterfactual situations to specific actions. This is to say, when people are not responsible for something, there may not be anything determinate, either to be found or constructed, that they would be responsible for *instead*. Susan Hurley gives an example that echoes Williams's worry about constitutive luck, that, "if Sam had not had the deprived childhood that makes his current low income bad luck for him, what would he have been responsible for instead? He might have chosen to be a workaholic or a surfer, or anything in between. I call this the indeterminacy problem" (Hurley 2003, 162). There are too many things people would choose if they could, under various counterfactual conditions. So the question then becomes, is the choice of becoming pious really something that should be actively disincentivized by the state?

The way responsibility is cast by Dworkin and Barry, is that it rewards the minimally religious individual who is able to convert what little freedom she has been given to her into a higher level of satisfaction. As Joseph Heath notes, this "frugality is not rewarded because it is intrinsically good, it is rewarded because the frugal individual imposes fewer costs on others, and therefore needs to moderate her desires to a lesser degree" (1998, 185). To then expect her to forgo or to restrict satisfaction of that preference because it is expensive is, therefore, to ask her to accept an alienation from what is deep in her. Religious people do not regret the duties imposed on them by their reading of their religion, such as modesty in public or the extra lengths they have to go for the sake of piety, but they do regret the disadvantages that attend to their religion in the context of Western society. The point is not that religious beliefs per se are unrevisable or uncontestable, but that resource considerations provide the wrong sorts of grounds for motivating people to revise their religious beliefs or their commitment to their community (McGann 2012, 13).

My argument here against Dworkin and Barry is a species of argument called the Harshness objection (see Voigt 2007) used against those who define responsibility as identification with preferences. To only be able to practice the bare minimum just in the case this leads to a conflict that needs an accommodation is too harsh a penalty to inflict. The irony of the harshness objection is that as one begins to identify more deeply with one's preference, according to Barry and Dworkin, one gains more and more responsibility for that preference yet it also makes the cost of that responsibility more and more harsh. As David Miller argues against Barry's comparison of religious practice with disability, "the opportunity to do X, in other words, is not just the physical possibility of doing X. At the very least, it is the possibility of doing X without incurring excessive costs" (2002, 51). As Tomlin points out, as an egalitarian, one cannot be monomaniacal about responsibility-sensitivity because we should remember its *pro tanto* nature (2013, 397). And therefore Carl Knight concedes that the brute luck/ option luck distinction cannot carry all the justificatory load. This highlights "the possibility that egalitarian justice, especially as depicted by Dworkin, treats the bearers of valuational judgment-based expensive tastes in unduly harsh fashion" (2009, 497).

5 Getting Beyond Choice/Chance

Again, I emphasize, the argument presented here should not be taken to show that although it is intuitive that we think certain minority religious practices like veiling are a choice, that veiling is *really* unchosen and is therefore brute luck and should be accommodated. This assumes that the chance/choice distinction held here is the appropriate view of egalitarianism. The argument presented is meant to trouble this distinction. I can agree with those at Québec town halls or Brian Barry that this practice of veiling did not "happen" to these women. Practicing veiling is not an involuntary action nor is culture reified enough that it "causes" them to veil. These women's agency was integral throughout the process. One might even say, through their self-discipline they were able to increase and unlock new capacities that extend their agency.

But in having to defend the veil as a choice, it becomes impossible for women who veil to articulate, in a way that is intelligible to the secular public, the fact that the practice of veiling is indispensable to their religiosity *and* their sense of self because these are seen as inimical to each other. My point is that Brute luck egalitarianism is not able to cover the type of agency expressed by women who veil and this is not just about being descriptively wrong in an academic sense. By rendering veiled women's distinct configuration of agency conditioned by authority unintelligible, they open approaches of critiquing these religious practices that should not

be legitimate. In France for instance,[17] the argument was made that "there are a thousand ways for a Muslim woman who aspires to wear the veil to wear it on the *inside* without wearing it on the outside" (Fernando 2010, 26). This kind of argument draws on the assumption that the relationship between conscience and practice is a semiotic relationship of signification. For them, religious practices like veiling are outward manifestations of an *already* constituted conscience. According to this logic, banning a practice does not constitute a violation of religious liberty because it has no effect on the believer's *conscience*. If I had relied on a theory of transformation based just on phenomenal and experiential transformation rather than self-transformation as a disciplined, embodied work, the door might have been reopened for this kind of choice/chance argument. The inner, cognitive belief becomes involuntary and unchangeable and therefore inviolable but a religion's external manifestations are variable, optional and chosen and therefore do not have to be accommodated. Not relying on cognitivist assumptions makes a difference in the real world because it is this logic of unchosen internal belief versus chosen external manifestation of religiosity that is used the European Court of Human Rights (ECHR) to justify not overturning bans on veiling in France, Switzerland, and Turkey.[18]

Finally, one might justifiably ask how we can think about minority religious accommodation without choice/chance or internal/external belief to arbitrate which accommodations are legitimate and which are not? The first step in making this gestalt switch is to stop thinking of this issue as one of accommodation at all. The language of "accommodation" is not part of a theory of justice but implies what Anna Galeotti terms a modus vivendi, a pragmatic compromise, which can be accorded today and denied tomorrow, a concession of discretionary power (2002, 43). This lack of secureness in their ability to pursue their religious ends means that minority groups will begin to lack confidence in the majority's capacity or willingness to be responsive to their concerns and so there is a failure of trust (Carens and Williams 1998, 168). A concentration on which accommodations the majority should allow hides the history of exclusion of the minority in question either because they are latecomers on the scene, or because they were previously oppressed or invisible. So these conflicts of accommodation (veiling, wearing a turban on a motorcycle, Sikh child taking a knife to school, kosher/halal slaughter of animals etc.) are not actually about deep moral disagreement but rather concern asymmetries in social standing, status, respect, and public recognition (Galeotti 2002, 5). These conflicts then precipitate negative majoritarian perceptions of traits, habits, and practices of minority groups which are singled out as "different"

[17] Although this originates in France, there have been similar arguments made in support of headscarf bans in other countries

[18] See *Dogru v. France*, no. 27058/05, ECHR 2008; *Dahlab v. Switzerland* (dec.), no. 42393/98, ECHR 2001-V; and *Leyla Şahin v. Turkey* (just satisfaction) [GC], no. 44774/98, ECHR 2005-XI.

and excluded from what the majority defines as standard forms of behavior (Galeotti 2002, 10).

Jeremy Waldron asks us to consider certain practices from a different view than the majority. Take for example, the situation where some children get together with an older adult and he supplies them with alcohol. What about the situation where a priest passes a cup of wine to young communicants. Are these the same action or different actions? A man is found in a public place with a knife concealed on his person. Is this knife a dangerous and offensive weapon? Or does it belong to a Sikh, carrying a kirpan, in fulfillment of religious obligation (Waldron 2002, 4)? The first is understood by most in the West as part of a recognized, innocuous everyday occurrence, while the second is usually considered an accommodation. It is not just that religious minorities should be allowed accommodations because there is a history of oppression but because otherwise the practices of religious minorities will never be integrated into the unproblematic traditions of the majority. There have been accusations that multicultural minority traditions are too rigid and closed off from "liberatory" norms of Western culture. Yet if we continue to talk about "allowing" religious practices as "accommodations" for religious minorities, Western tradition becomes rigid and not open to the inscription of different norms as part of "our" heritage. Because the event of communion is entwined with Western tradition such that it becomes a background practice, allowing underaged children to consume wine is considered natural rather than as an "accommodation" to the equality of law. There is no hyperbole that this may become a gateway to alcohol addiction and that we must paternally take these children's safety into consideration first. Yet this is the kind of discourse that surrounds the case of Sikh's wanting to wear turbans instead of motorcycle helmets. The rationale that Québec and Italy[19] give to why they allow a large cross to adorn Québec's parliament and Italy's schools, yet must ban all other religious symbols from public office, is that the cross is part of Québec and Italy's patrimony. Yet if we take the case of the history of Jews in Québec, their history and traditions are entwined with Québec's for over a hundred years. Why should the Jewish kippah then not be considered as just another part of Québec's patrimony? How long does it take for "their" traditions turn into "our" traditions?

Muhammad Velji
E-mail : muhammad.velji@gmail.com

[19] See *Lautsi and Others v. Italy* (just satisfaction) [GC], no. 30814/06, ECHR 2011.

Acknowledgements I would like to thank Alia Al-Saji and Daniel Weinstock for their patience in going through earlier drafts as well as pushing me conceptually. I would also like to thank the two anonymous referees for their suggestions and Douglas Hanes for coming up with the title of the paper.

References:

Arneson, Richard J. 2000. "Luck Egalitarianism and Prioritarianism." *Ethics* 110 (2): 339–349. http://dx.doi.org/10.1086/233272.

Arneson, Richard J. 2007. "Does Social Justice Matter? Brian Barry's Applied Political Philosophy." *Ethics* 117 (3): 391–412. http://dx.doi.org/10.1086/511732.

Asad, Talal. 2001. "Reading a Modern Classic: W. C. Smith's 'The Meaning and End of Religion'." *History of Religions* 40 (3): 205–222. http://dx.doi.org/10.1086/463633.

Barry, Brian. 1991. "Chance, Choice, and Justice." In *Liberty and Justice: Essays in Political Theory 2*. Oxford: Clarendon Press.

Barry, Brian. 2001. *Culture And Equality: An Egalitarian Critique Of Multiculturalism*. Cambridge, MA: Harvard University Press.

Card, Claudia. 1996. *The Unnatural Lottery: Character And Moral Luck*. Philadelphia, PA: Temple University Press.

Carens, Joseph H. and Melissa S. Williams. 1998. "Muslim Minorities in Liberal Democracies: The Politics of Misrecognition." In *Secularism And Its Critics*, edited by Rajeev Bhargava. Delhi: Oxford University Press.

Chambers, Clare. 2009. "Each Outcome Is Another Opportunity: Problems with the Moment of Equal Opportunity." *Politics, Philosophy & Economics* 8 (4): 374–400. http://dx.doi.org/10.1177/1470594X09343066.

Cohen, G. A. 2004. "Expensive Tastes Rides Again." In *Dworkin and His Critics: With Replies by Dworkin*, edited by Justine Burley. Malden, MA: Blackwell Publishing.

Dworkin, Ronald. 1981. "What is Equality? Part 2: Equality of Resources." *Philosophy and Public Affairs* 10 (4): 283–345.

Dworkin, Ronald. 2004. "Ronald Dworkin Replies." In *Dworkin and His Critics: With Replies by Dworkin*, edited by Justine Burley. Malden, MA: Blackwell Publishing.

Fernando, Mayanthi L. 2010. "Reconfiguring Freedom: Muslim Piety and the Limits of Secular Law and Public Discourse in France." *American Ethnologist* 37 (1): 19–35. http://dx.doi.org/10.1111/j.1548-1425.2010.01239.x.

Foucault, Michel. 2005. *The Hermeneutics Of The Subject: Lectures At The College De France, 1981–1982*. Translated by Frederic Gros. New York, NY: Palgrave-Macmillan.

Galeotti, Anna E. 2002. *Toleration as Recognition*. Cambridge: Cambridge University Press.

Hadot, Pierre. 1981. *Exercices Spirituels et Philosophie Antique*. Paris: Etudes Augustiniennes.

Heath, Joseph. 1998. "Culture: Choice or Circumstance?" *Constellations* 5 (2): 183–200. http://dx.doi.org/10.1111/1467-8675.00087.

Heyes, Cressida J. 2007. *Self Transformations: Foucault, Ethics, and Normalized Bodies*. Oxford: Oxford University Press.

Hurley, Susan L. 2003. *Justice, Luck, and Knowledge*. Cambridge, MA: Harvard University Press.

Knight, Carl. 2009. "Egalitarian Justice and Valuational Judgment." *Journal of Moral Philosophy* 6: 482–498. http://dx.doi.org/10.1163/174046809X12464327133177.

Laycock, Douglas. 1990. "The Remnants of Free Exercise." *The Supreme Court Review* 1990: 1–68.

Mahmood, Saba. 2005. *Politics of Piety: The Islamic Revival And The Feminist Subject*. Princeton, NJ: Princeton University Press.

Mahmood, Saba. 2012. "Ethics and Piety." In *A Companion to Moral Anthropology*, edited by Didier Fassin, 223–241. Hoboken, NJ: Wiley Blackwell.

McGann, Michael. 2012. "Equal Treatment and Exemptions: Cultural Commitments and Expensive Tastes." *Social Theory and Practice* 38 (1): 1–32. http://dx.doi.org/10.5840/soctheorpract20123811.

Mendus, Susan. 2002. "Choice, Chance and Multiculturalism." In *Multiculturalism Reconsidered: 'Culture And Equality' And Its Critics*, edited by P. J. Kelly. Cambridge: Polity Press.

Miller, David. 2002. "Liberalism, Equal Opportunities and Cultural Commitments." In *Multiculturalism Reconsidered: 'Culture and Equality' and Its Critics*, edited by P. J. Kelly. Cambridge: Polity Press.

Olsaretti, Serena. 2009. "Responsibility and the Consequences of Choice." *Proceedings of the Aristotelian Society* 109 (1pt2): 165–188. http://dx.doi.org/10.1111/j.1467-9264.2009.00263.x.

Paul, L. A. 2015. "What You Can't Expect When You're Expecting." *Res Philosophica* 92 (2): 149–170. http://dx.doi.org/10.11612/resphil.2015.92.2.1.

Rakowski, Eric. 1991. *Equal Justice*. Oxford: Clarendon Press.

Sharadin, Nathaniel. 2015. "How You Can Reasonably Form Expectations When You're Expecting." *Res Philosophica* 92 (2): 441–452. http://dx.doi.org/10.11612/resphil.2015.92.2.2.

Tan, Kok-Chor. 2008. "A Defense of Luck Egalitarianism." *The Journal of Philosophy* 105 (11): 665–690. http://dx.doi.org/10.5840/jphil20081051120.

Taylor, Charles and Gerard Bouchard. 2008. *Building the Future: A Time for Reconciliation Report*. Québec: Commision de Consultation sur les Pratiques D'Accommodement Reliees aux Difference Culturalles.

Tomlin, Patrick. 2013. "Choices Chance and Change: Luck Egalitarianism Over Time." *Ethical Theory and Moral Practice* 16: 393–407. http://dx.doi.org/10.1007/s10677-012-9340-0.

Ullmann-Margalit, Edna. 2006. "Big Decisions: Opting, Converting, Drifting." *Royal Institute of Philosophy Supplement* 58: 157–172. http://dx.doi.org/10.1017/S1358246106058085.

Valenta, Markha. 2012. "Pluralist Democracy or Scientistic Monocracy?: Debating Ritual Slaughter." *Erasmus Law Review* 5 (1): 27–41.

Vallentyne, Peter. 2008. "Brute Luck and Responsibility." *Politics, Philosophy and Economics* 7 (1): 57–80. http://dx.doi.org/10.1177/1470594X07085151.

Vallentyne, Peter. 2011. "Responsibility and False Beliefs." In *Responsibility and Distributive Justice*, edited by Carl Knight and Zofia Stemplowska. Oxford: Oxford University Press.

Voigt, Kristin. 2007. "The Harshness Objection: Is Luck Egalitarianism Too Harsh on the Victims of Option Luck?" *Ethical Theory and Moral Practice* 10: 389–407. http://dx.doi.org/10.1007/s10677-006-9060-4.

Waldron, Jeremy. 2002. "One Law for All? The Logic of Cultural Accommodation." *Washington and Lee Law Review* 59: 3–34.

Weir, Allison. 2013. *Identities And Freedom: Feminist Theory Between Power And Connection*. Oxford: Oxford University Press.

Williams, Bernard. 1981. *Moral Luck: Philosophical Papers, 1973–1980*. Cambridge: Cambridge University Press.

Transformative Choice: Discussion and Replies

L. A. Paul

Abstract: In "What You Can't Expect When You're Expecting," I argue that, if you don't know what it's like to be a parent, you cannot make this decision rationally—at least, not if your decision is based on what you think it would be like for you to become a parent. My argument hinges on the idea that becoming a parent is a *transformative experience*. This unique type of experience often transforms people in a deep and personal sense, and in the process, changes their preferences.

In section 1, I will explain transformative experience in terms of radical first-personal epistemic and self change. In section 2, I'll explain the notion of subjective value that I use to develop the decision problem. In section 3, I will discuss the way we ordinarily combine our introspective assessments with testimony and evidence. In section 4, I will discuss the problems for rational decision-making. In section 5, I will explore the problem of first-personally transformed future selves. In section 6, I will engage with the main themes and arguments and ideas of the authors of the papers contributed to this volume.

In "What You Can't Expect When You're Expecting," (2015b) I focus on a very ordinary, but deeply important, life-changing personal decision: whether to have a baby. I argue that, if you don't know what it's like to be a parent, you cannot make this decision rationally—at least, not if your decision is based on what you think it would be like for you to become a parent.

My argument hinges on the idea that becoming a parent is a *transformative experience*. Being a parent is a unique kind of experience that can dramatically change your core personal preferences and the nature of your lived experience. As such, it's the kind of thing that you have to experience in order to know how it will affect you. So if you've never been a parent, you don't know what it's like to be a parent. Having the experience is necessary for you to have the capacity to represent the nature of the outcome—the lived experience of being a parent—that the subjective value of the outcome depends upon.

Res Philosophica, *Vol. 92, No. 2, April 2015, pp. 473–545*
http://dx.doi.org/10.11612/resphil.2015.92.2.15

This unique type of experience often transforms people in a deep and personal sense, and in the process, changes their preferences. The idea isn't that you don't know what it's like to babysit, change diapers, or be very tired before you become a parent. Rather, what you don't know is its most important and distinctive feature: what it will be like to form and occupy the identity-constructing, preference-changing, physically and emotionally overwhelming perspective of being a parent.

If the salient details of the transformative experience of producing and becoming cognitively and emotionally attached to your child are epistemically inaccessible to you before you undergo this type of experience, then you cannot, from your first-personal perspective, imaginatively represent the relevant first-personal nature of the preference changes you will undergo. Because of your lack of experience, you lack the representational capacities needed to imagine, model, and grasp the nature of your future lived experience, and thus, of your future self.

In sum, you must decide whether to form yourself into a parent without knowing what it will be like to become a parent. This puts you, as a prospective parent, into a high-stakes decision problem.[1] The choice to become a parent (assuming your act is successful) is irreversible and will determine the nature of the rest of your life. Yet, the distinctive feature of being a parent, standing in a deep and loving attachment to the child you produce and raise, is epistemically inaccessible to you until you've actually become a parent. The nature and character of this attachment will, in the ordinary case, have a huge effect on your future lived experience. Thus, before you choose, you cannot assess the value of what you are likely to gain against the value of what you may lose.

This raises a special kind of philosophical problem: the choice to have a child asks you to make a decision where you must choose between earlier and later selves, with different sets of preferences, but where your earlier self lacks crucial information about the values, preferences and perspectives of your possible later selves. You cannot first-personally foresee or represent the new self you are making yourself into. We can think of the conceptual change involved as a first-personal version of a Kuhnian paradigm shift: one's first-personal view of oneself is not invariant under the epistemic transformation.[2]

The experience of becoming a parent is not the only kind of experience that can be transformative. It is merely an especially interesting case. It is especially interesting because many people have experienced it, the transition is reasonably abrupt, the experience can be dramatically different

[1] This is true even if you have some testimonial knowledge. While testimonial knowledge alone might license action in some low-stakes cases, it isn't enough, by itself, to license a life-changing action in a high-stakes case. (See Moss Unpublished, who points out how, when an agent moves to a high-stakes context, the contents of her knowledge can license fewer actions than they did before.)

[2] For related comments, see Van Fraassen 1999.

for different people, and an important ordinary approach to the choice (in contemporary Western society) involves deliberation and careful assessment paired with an explicit cultural narrative that urges us to "look within ourselves" to decide whether to make this major transition.

But there are other types of life transformations that can change you in this way. Going to war can be transformative (see Zelcer 2015). Developing from a child into an adult can be transformative. Descending into Alzheimer's can be transformative. Being betrayed, or betraying someone else, can be transformative. Getting divorced can be transformative.[3]

A less common sort of case, but one that illustrates the idea clearly, is the dramatic life change that a congenitally blind adult would experience if he were to gain sight. In cases where we examine changes in sensory capacities, it is intuitively clear that one's life is changed deeply and dramatically by having a distinctive new kind of sensory experience.

In all of these cases, the transformative nature of the experience can affect the real-life decisions we make when undergoing those experiences, and we must grapple with real-life philosophical issues. Once we start to look closely at major life experiences and the choices they involve, we seem to find transformation everywhere.

The wider thesis, then, is that there are distinctive philosophical issues concerning the way that we understand and construct who we are, and these issues arise in a most pressing manner when we contemplate a life-changing choice like whether to have a child. Such choices can change us deeply and permanently.

In this way, the question of whether to become a parent illustrates larger themes about the philosophical issues involved in the way we model, understand, and construct our selves. In my book, *Transformative Experience* (2014), I introduce and develop the notions of transformative experience and transformative choice and frame them in terms of transformative decision-making. There, I discuss the structure of transformative experience, and explore the tension it raises between rational decision-making and authentically forming our future selves.

In cases of transformative decision-making, you cannot grasp the subjective nature of your future lived experience, including the nature of your future self, until you become that future self, and thus you must make a life-changing decision without knowing, in the deepest sense, who you'll become. The book develops and elaborates the structure of transformative decision-making and its implications for a wide range of big life decisions.

Many of the authors in the papers contributed to this volume raise arguments or questions about my argument, or develop the theme of transformative decision-making in interesting and novel ways. I am extremely

[3] Preliminary results from psychological research done by Starmans and Bloom suggests that the transition from childhood to adulthood is transformative. Preliminary results from psychological research done by Nunziato and Cushman suggests several of the kinds of transformations I describe. I thank Christina Starmans, Paul Bloom, Josiah Nunziato and Fiery Cushman for discussion.

grateful to the contributors for their thoughtful engagement with the original argument in "What You Can't Expect When You're Expecting" (2015b) and with their subsequent engagement with the arguments and themes of *Transformative Experience* (2014). Below, I will attempt to address their concerns, reply to their arguments, and engage with their positive theses.

In section 1, I will explain transformative experience in terms of radical first-personal epistemic and self change. In section 2, I'll explain the notion of subjective value that I use to develop the decision problem. In section 3, I will explain the way life-making choices often involve assessments of one's future lived experience in terms of personal, subjective values, and discuss the way we ordinarily combine our introspective assessments with testimony and evidence. In section 4, I will explain how the ordinary, subjective deliberation involved in the decision to become a parent makes the choice transformative, and the problems this causes for rational decision-making. In section 5, I will explore the problem of first-personally transformed future selves.

In section 6, I will engage with the main themes and arguments and ideas of the authors of the papers contributed to this volume. Section 6.1 discusses formal epistemology and decision theory and replies to John Collins, Jennifer Carr, and Thomas Dougherty, Sophie Horowitz, and Paulina Sliwa. Section 6.2 discusses social choice, social justice, and social identity and replies to Rachael Briggs, Elizabeth Barnes, Rachel McKinnon, Ryan Kemp, and Muhammad Velji. Section 6.3 discusses subjective value and happiness and replies to Antti Kauppinen. Section 6.4 discusses decision-making in contemporary ethics, including "I'll be glad I did it" reasoning, reasons, and self-construction, and replies to Elizabeth Harman, Dana Howard and Ruth Chang. Section 6.5 discusses the problems with using contemporary empirical research to choose to have a child and replies to Nathaniel Sharadin.

1 Transformative Experience: Epistemic and Personal

An *epistemically* transformative experience is an experience that teaches you something you could not have learned without having that kind of experience. Having that experience gives you new abilities to imagine, recognize, and cognitively model possible future experiences of that kind. A *personally* transformative experience changes you in some deep and personally fundamental way, for example, by changing your core personal preferences or by changing the way you understand your desires and the kind of person you take yourself to be.

It is important to note that, as I use the phrase, a *transformative experience* is an experience that is both epistemically and personally transformative. Transformative choices and transformative decisions are choices and decisions that centrally involve transformative experiences. In many

cases, it is the degree of the epistemic transformation that creates the corresponding personal transformation—the dramatic epistemic change carries dramatic personal change along with it.

Having a child, at least in the ordinary, traditional way, involves the transformative experience of gestating, producing, and becoming attached to the child you create. In such a case, if you area woman who has a child, you go through a distinctive and unique experience when growing, carrying, and giving birth to the child, and in the process you form a particular, distinctive and unique attachment to the actual newborn you produce. Men can go through a partly similar experience, one without the physical part of gestating and giving birth. For both parents, in the usual case, the attachment is then deepened and developed as they raise their child.

I take the experience of having a child to be unique, because physically producing a child of one's own is unlike any other kind of human experience. As a mother, in an ordinary pregnancy, you grow the child inside yourself, and produce the baby as part of the birth process. As a father, you contribute your genetic material and watch the child grow inside your partner. When a newborn is produced, both parents experience significant hormonal changes and enter new physiological states, all of which help to create the physical realizer for the intensely emotional phenomenology and cognitively rich mental states associated with the birth. These experiences contribute to the forming and strengthening of the attachment relation, and further characteristics of the nature of the attachment manifested between you and your child are determined by the particular properties of the actual child you produce. All of this generates the unique lived experience associated with having one's first child. Raising a child is then a temporally extended process that extends, deepens, and complicates this relationship.[4]

2 Subjective Values

Subjective values, as I understand them, are experientially grounded values attaching to lived experiences.[5] These are the types of values that are involved in transformative decision-making: I describe them as "what it's like" values to emphasize that they necessarily include phenomenal value. But it is important to note that subjective values can be based on more

[4] For simplicity, I focus on the case of parenting one's biological child. Other types of parenting can also be transformative.

[5] To forestall confusion: to say "*x* is grounded in *y*" need not entail that *x* is entirely grounded in *y*. The language is similar to causal language: saying "*c* causes *e*" does not entail that *c* is the *only* cause of *e*. ("The striking caused the match to light" should not be understood to imply that the striking was the only cause. Oxygen in the environment, lack of extreme humidity, etc., are also causes.) So when I say that subjective values ground objective values, or that objective values depend on subjective values, this does *not* entail that subjective value is the only ground for objective value, or that objective values depend solely on subjective values.

than merely qualitative or merely sensory phenomenology: they could also include values arising from nonsensory content. Independent of any esoteric theses about qualia or phenomenology, subjective values are intended to be values that attach to the contentful features of rich, developed experiences embedded in a range of mental states such as beliefs, emotions, and desires.

Thus, as I understand it, the subjective value of a lived experience is not merely a matter of the phenomenal character of the internal characteristics of one's inner life. It's a richer value, a value that includes what it's like to live "*in this*," as John Campbell puts it (Campbell 2015; Paul 2015a). That is, it encompasses the value of what it's like to live in a particular set of circumstances, where those circumstances may include the external environment. (My reply to Kauppinen 2015 below connects with some of these issues.)

This should make it clear that subjective values are not internalist, purely qualitative values. I do not assume that the subjective value of future lived experience is determined merely by the inner, purely qualitative state of the self who is transformed by the new experience. Rather, the information gained by the discovery (even a merely qualitative discovery), functions as a necessary element in the epistemic and personal transformation of the agent. Imagining the subjective value of your future lived experience can also involve an act of *de re* imagination about the nature of your lived experience in the world.

The point of emphasizing the necessary role for experience is that you must be able to cognitively evolve yourself forward under the transformative change involved in order to prospectively grasp the subjective value of life in your possible future circumstances.[6] What it will be like for you to live in those circumstances is informed by and infused with the qualitative information you gain when you undergo the epistemic transformation. It will have a significant effect on how your core personal preferences are formed and developed. In cases of epistemic transformation, experience is needed to teach you what your future could be like, since experience of the relevant kind is needed to give you the capacity to first-personally represent or model your possible future selves in these possible future circumstances.

There are other types of values, of course, such as impersonal moral and political values, that can also come into play when we make big decisions: I do not propose that we ignore these values. When we make big life choices, we should always make them in concert with our best objective moral, legal, and empirical standards. But for the purposes of this

[6] These values are assigned to outcomes, because a normatively rational decision-maker is to choose to perform the act with the highest expected value, given her assignments of values to outcomes and probabilities to states. Outcomes are defined on acts by agents in states, and values for the agent attach to such outcomes. In my discussion below, I will sometimes speak loosely of assigning credences or probabilities to outcomes, where this should be understood more precisely in terms of a probability function defined on conditionals about agents performing acts in states of the world.

discussion, where we are considering an individual's personal choice, we are focusing on subjective values. That is, I am assuming there isn't some external reason that trumps or dominates your choice, making subjective deliberation irrelevant or unnecessary. As a result, the decision centrally involves your preferences concerning your future lived experiences. In the cases of interest, such preferences cannot be eliminated from your deliberations without doing violence to the natural and ordinary way you want to make the choice.

So while questions of morality and social value can apply, the context of concern for transformative decision-making is subjective lived experience.

These are the decisions I am interested in: they define a class of important and interesting life choices where a particular sort of subjective decision is called for. Not all contexts and not all transformative experiences are contexts of transformative choice, but some of the most interesting and important ones are, and they are the focus of my project. In Paul 2014, I characterized these choices as choices about who you take yourself to be and who you want to become. In what follows, I will sometimes describe these choices, understood to occur in these contexts, as *life-making choices*.

3 Subjective Deliberation

On the approach to deliberative choice I am engaging with, when making an important, life-changing decision, you want to knowledgably assess the nature and character of the different possible outcomes of your choice so that you can choose in an informed way. More simply, you want to know what your choice means for you and your future (and for others whose futures depend on your choice). So you want to know what the possible outcomes are of your choice, and you want to know this in a way that allows you to assess and value each of them. In addition to being informed about the natures of the outcomes, you'll need to know how likely each outcome would be, given your act. Once you can assess the natures and values of the possible outcomes (and you know their likelihoods), you can determine your preferences, that is, you can determine what you prefer to have happen as the result of your choice.

Thus, to approach your life-making choice in a reflective, deliberative way, you reflect upon how you want to realize your future and then map out the options involved. You reflect on the ways you might act, on the possible outcomes of your actions, and on what those outcomes, if realized, would be like. Then you determine your preferences about how to act, given how you'd like your life to go.

What's of key interest in this discussion is that you need to be able to prospectively assess the natures of the possible outcomes of your choices in order to evaluate them properly.[7] If you can assess the natures of these

[7] There are important questions about likelihoods of the outcomes, or, speaking more precisely, about their credences. One important question concerns whether we have the knowledge we

possible outcomes, you can determine your preferences about your future and choose accordingly.

One especially effective way to reflect on the natures of different possible outcomes is to imaginatively project different possible futures for yourself, futures that stem from the different possible choices you could make.

Such imaginative projection is a very natural way to approach a major decision, and relies on the ability to cognitively model different possible outcomes.[8] We do it in ordinary contexts all the time. For example, when you research possible apartments to live in, or consider buying a house, to evaluate the best place for you to live, you mentally project yourself into a future where you live in that place. The value of the outcome depends partly on how much it would cost to live there, but it also depends, importantly, on what it would be like to live there. That is, you want to assess what the lived experience would be like in that house or that apartment.

People implicitly recognize this in lots of decision contexts. In the house-hunting example, if you care about where you live and what kind of space you live in, you want to choose carefully and deliberately. Ideally you'll visit each promising place, examine it inside and out, and attempt to assess what it would be like to live there by imagining or somehow representing yourself actually living there. And if you are house hunting as part of a decision between job offers in different cities, you'll want to attempt to imagine what it would be like to live in that city, with that job, in that house. You'll try to project yourself into each possibility in order to assess and compare: should I take *this* job, and *this* apartment? Or should I choose *that* job and *that* apartment?

The more important this decision is—that is, the more important it is to you where you live and work—the more important it is that you know as much as possible about each outcome. Why? Because the better your assessment of the subjective value of the lived experience of being in that home, in that city, with that job, the better informed your decision will be.

So an important part of your deliberation involves determining the values of the possible outcomes of your action. To assess the subjective values of these possible outcomes, you need to be epistemically acquainted with them in the right way (see Lewis 1989). In particular, what is necessary for the right sort of epistemic acquaintance is that you represent the nature of the lived experience of the outcome to yourself under the subjective mode of presentation. This gives you the acquaintance you need to grasp and assess its subjective value. Arguably, the best way to manifest this representational capacity is to imaginatively represent the nature of the lived experience of the outcome. This will allow you to stand in a relation of imaginative acquaintance to that outcome, and grasp its subjective value.

need to have in order to attach credences to outcomes in a way that will license our actions. (See Moss Unpublished.) This issue will surface indirectly in a number of places, including my exchange with Dougherty et al. (2015) discussed below.

[8] See my exchange (Paul 2015a) with Campbell (2015) for related discussion.

Is imaginative acquaintance the only way to manifest this representational capacity? I'd prefer to hold that imaginative acquaintance is the most natural and important route to assessing subjective value, but to allow for the remote possibility that there might be others. What is required for subjective valuation, however we arrive to it, is a grasp on the nature of the outcome from the subjective perspective. That is, we must represent the outcome under a suitably first-personal mode of presentation. In order to grasp the value of lived experience, we need to acquaint ourselves with it, by imaginatively representing (or perhaps by representing some other way) the nature of the experience using the subjective, or first-personal, mode of presentation.

The importance of the subjective mode of presentation is familiar from discussions in the philosophy of mind: to grasp the subjective value of an outcome involving seeing red, I need to be able to represent seeing red from my first-personal, conscious perspective. Imaginative acquaintance is the usual basis for how we represent: for example, I can represent the experience of seeing red because I am imaginatively acquainted with what it's like to see red. And, at least in ordinary contexts, experience of the right sort is necessary for us to have the ability to represent outcomes of that sort under the subjective mode of presentation.

I've been emphasizing the role of first-personal representation in assessing the subjective values of outcomes. But of course, we can get some kinds of information from other sources, such as testimony from friends and relatives and advice from experts. As you attempt to predict how you'll respond to an experience, and, correspondingly, decide how to act, you should take into account any reliable outside testimony and empirical evidence that bears on the question of what to do. In particular, you might hope to get descriptive information from these sources, such as descriptions of the various possible outcomes and other information about the numerical magnitude and valence of possible values.

We need to refine this idea in the context of this discussion, however. Recall that, in the first instance, what you want to know is the subjective values of your outcomes. Unfortunately, what testimony gives you is something different. What friends and relatives can do is describe their own experiences and outcomes. What experts can tell you is about the valence and magnitude, via numerical specifications or descriptions, of the subjective values of each outcome for the average member of a population that is relevantly similar to you. What you are getting, then, is not information that can, by itself, allow you to represent your own outcomes under the subjective mode of presentation. Rather, you are getting descriptive information about possible subjective values that is intended to aid and guide you in your representation and grasp of your own subjective values.

What you hope to do is to use this information, the general evidential facts and the testimonial evidence of those close to you, to determine what is right for you, in your situation, with regard to your life-making choice.

That is, you want to consider testimony and evidence of people similar to you when you think about how you'll respond to the experience you are considering undergoing. You'll use this to help you to know what the experience will be like for you, given what you know about the situation and what you know about yourself (which includes whatever you might learn about yourself from how others respond), so you can predict how you'll respond—that is, so you can grasp your own subjective values.

Going back to our house-hunting example: when you think about where to work and live, you should consider the testimony of friends and relatives. You should consult with others who live or who have lived in each place. Based on their experience and on what they believe about you, your sources might even give you testimony about what they think you'd prefer. You could also get official statistics about the job, about the crime rate, about the length of the commute, and about any amenities the neighborhood offers. To decide in a deliberative, informed way, you'll want to make use of these external facts and testimony, but you'll make use of them by combining them with your sense of who you are and of your own representation of how it would be for you to live and work in each place.

That is, your comparison of the subjective values of these outcomes is still based on the attempt to first-personally project yourself into each possible future. It's just that your attempt to do this should be as informed as possible by the testimony and evidence available to you.

So you weigh the testimony and evidence and consider how to apply it to your own case. Metaphorically, you survey the landscape presented by the data and testimony, and attempt to find yourself in it. You use your knowledge of how other people respond, paired with your own, first-personal assessment of who you are and how similar you take yourself to be to the others who you have data about, to prospectively assess how you think you'd respond to the experience you are considering undergoing.[9] Formally, we might say that you use information and testimony to update your (prior) assessments of how you'll respond to the experience.

This should make it clear that taking testimony into account when you assess outcomes is not simply *replacing* what you think with what friends and family tell you. Likewise, taking scientific evidence into account doesn't mean you unreflectively *replace* how you think you'll respond in this situation with what the scientific expert tells you about how people like you tend to respond.

One of the reasons why thinking for yourself is so important in these cases is because you are making a special kind of high-stakes decision. As I have been emphasizing, transformative experiences are a distinctive kind of experience, a kind of experience that forms—or re-forms—who you are. In other words, transformative choices are life-making choices.

[9] Of course, you may also want to predict how your act can affect others, including how it affects the lived experience of those who are close to you.

This is because transformative experiences are self-making experiences: a distinctive feature of a transformative experience is that the dramatic epistemic change involved also involves about a change in the agent's self. When making a transformative choice, it's not just about what it will be like to live somewhere new or to do something you haven't done before. It's a choice about *what it will be like to be you*. When you choose to act, in hopes of bringing about a preferred outcome, you are choosing who you'll become, based on your preferences about what self, and what sort of person, you'd like to be. As such, transformative decisions concern some of the most important and personally meaningful choices you will ever make.

This makes transformative choice into a much higher-stakes choice than the relatively mundane choice of where to live or what job to take (except to the extent that such a choice could be transformative, depending on the details). You don't just want to know what something new will be like. You want to know what *you'll* be like, that is, you want to know what sort of self you are making yourself into. And in such a high-stakes case, knowing what your future could be like before you try to make it actual is even more important. The stakes are higher, for the nature of the experience involved, the experience of being who you are, is, epistemically speaking, the most intimate sort of experience possible.

A way of emphasizing the importance of introspective reflection in these sorts of cases is to say that high-stakes choices like this can be subject to an "authenticity norm." The authenticity norm concerns the way you make life choices in concert with your first-personal understanding of who you are and what you want from life. An agent who authentically understands herself first-personally grasps her defining nature and values from the inside, that is, she knows who she is under the subjective mode of presentation. (As Campbell (2015) points out, we might also need to authentically understand the perspectives of others. The therapist who treats the emotional pain of her patients is able to do so authentically if she is able to imaginatively grasp salient features of their emotional experience.)

Having a first-personal grasp on the subjective values of your possible futures allows you to make choices about your future authentically. Who you take yourself to be now and whom you are making yourself into is informed by your ability to imaginatively evolve your first-personal perspective into your different possible futures. Borrowing from the philosophy of mind, we might say that your grasp on the subjective nature of your possible future lived experiences allows for an authentic mode of presentation. The idea is that, for authentic understanding, you must understand, under the subjective mode of presentation, who you are now and how you'll evolve under change into your future self. In this way, you have a grasp on what your future could be like, and on who your future self could be, so if you choose to try to bring about this particular future, you choose it authentically.

The authenticity norm comes into relief when making life-making choices, for these are high-stakes choices that we usually want to make rationally *and* authentically. Authentic rational action seems to involve acting on the basis of one's deepest principles and values, where one rationally grasps one's own principled commitments under the subjective mode of presentation.[10] Such principled commitments include commitments to the kind of person you want to be and to what you care about most.

A paradigmatic case of a life-making choice is the choice to become a parent. This is a big, irreversible life choice that will probably have an outsized effect on the rest of your life. When you (carefully and deliberately) make a major life decision such as whether to become a parent, ordinarily, you think about who you are and what you want out of life, about your principles and values, and about your hopes and dreams. You look around you and see how your friends and relatives live their lives as parents—or as childfree types. You consider any available reliable testimony and empirical data about parenting and about the lives of the childfree. As you consider the relevance of the data, testimony, and any related anecdotal evidence to your value assessment, you compare yourself to the people that this evidence comes from. Finally, to determine your preferences about parenthood, you take your comparisons into account as you imaginatively consider and assess a future where you are a parent, caring for your child. To decide, you compare your expectations for your future as a parent to your expectations for a childfree life.

All of this might seem perfectly straightforward. But, as I have argued, it isn't—because the experience of becoming a parent is transformative.

4 The Transformative Choice to Become a Parent

To see how the life-making choice to become a parent is transformative, and to understand the particular challenges this raises, we need to set the decision context. First, assume that the decision maker is actually in a position to decide: that is, assume she will get pregnant and have a baby if she decides to have a child, that she has sufficient financial resources, and there are no other constraints that would prevent her from acting as she sees fit. Also assume that the child would be her first: that is, she has never experienced being a parent.

As I argued above (and in Paul 2014; 2015b), gestating, producing, and becoming attached to your child is a unique kind of experience, such that the experience of becoming a parent is epistemically transformative. The distinctive ground for the transformative nature of the experience is the epistemically distinctive attachment relation that is created between you and your child, along with its associated properties and the process that

[10] So authenticity, as I understand it, involves knowledge under the subjective mode of presentation. I'd like to thank Joshua Landy for discussion.

led to its creation. This creates the special, intense, and unique feelings of parental love, care, personal engagement, and responsibility that new parents experience. These feelings deepen, develop, and change throughout the extended experience of raising your child into adulthood.

The subjective value of this distinctive type of lived experience should be understood as encompassing the external, mind-independent fact that you are standing in an attachment relation to your child. It is grounded by your loving attachment to your child, your newfound sense of responsibility and joy, as well as the rest of the features of the intense, extended lived experience of being a parent, including longer term experiences associated with raising your child to adulthood. When you discover the nature of standing in this attachment relation through the experience of being psychologically attached to the child, you gain the capacity to have the beliefs, desires and other states that define the lived experience of parenthood. In this way, you gain the capacity to represent what it's like for you to be a parent, and can thus grasp the subjective value of being a parent.

In addition to being epistemically formative, the choice to become a parent is also (usually) life changing. Who and what you care about can change, often dramatically. So if you've never parented a child, in the ordinary case, becoming a parent is both epistemically and personally transformative. The attachment changes who you take yourself to be, in the sense that you define yourself, at least partly, as the parent of your child, and changes some of your core personal preferences. Many less fundamental preference changes follow on from those.

In this way, the experience of having one's first child is what I define as a transformative experience, and the choice to have the child is a transformative choice. As I argue in Paul 2014; 2015b, given the ordinary way we frame this choice, it is not rational. That is, on condition that we frame the choice in terms of what it would be like to become a parent, we cannot make the choice rationally.

The argument is not that there is no way to make the choice rationally. Rather, the argument is that the way many of us ordinarily *want* to make the choice, in terms of authentically grasping and assessing what it will be like to have a child *before we have a child*, is not rationally available to us.

So the problem is that, ordinarily, we want to be able to understand what it would be like to be a parent before we decide to do it, that is, we want to grasp the subjective values of our future lives as parents before we make the choice to have a child. After all, we are making a choice that will change the rest of our lives. It will have broad and significant effects on our careers, our loved ones, our financial and emotional situation, and on pretty much everything else in our lives. It's an enormous, irreversible life change, and carries with it huge responsibility and commitment. Lives depend upon it. Having a child isn't just a hobby that you spend a couple hours doing over the weekend. For many of us, it changes pretty much everything.

It's a much bigger commitment than renting an apartment for a year or taking a job in a new city. So of course, before we decide to do it, we want to know: will becoming a parent be like *this*, or will it be like *that*? Will I prefer *this* life, a life as a parent, to *that* life, a life as a childfree person? And so on. Very few people would buy a house or commit themselves to life in a new city, sight unseen, unless they had no choice. So too with parenting—we want to know what sort of life we are choosing before we undertake it. Since we cannot literally visit our possible future lives as parents before we choose (unlike the way we can visit a house, or visit a city), we try to acquaint ourselves with our possible futures in some other way. We attempt to mentally "visit" our possible futures using imaginative representation or cognitive modeling, in order to assess them and compare them, and to make an informed choice about what we want to do and who we want to become.[11]

Once we see the importance of representing our future possibilities in order to make the choice, the nature of the problem becomes clear. Before becoming a parent, the transformative nature of the choice means that we lack the capacity to knowledgeably represent ourselves as parents under the subjective mode of presentation. As a result, we cannot grasp the subjective value of what it's like to be a parent until we actually become parents. Put in technical terms, a person who has not parented lacks a value function for the outcomes involving parenting: she cannot represent the subjective values of those outcomes.[12]

Now, a natural idea at this point is to suggest that we turn to testimony, such as the advice we get from friends and relatives and the evidence we get from scientific experts. But as I discussed above, this sort of testimony won't give you the information you need to grasp your own subjective value for being a parent, for two important reasons.

The first reason is that the testimony is merely descriptive. (Compare the attempt to know what it's like to see red merely on the basis of someone else's description of what it's like to see red. If you've never seen red, such a description won't allow you to grasp the subjective value of seeing red.) Testimony from friends, relatives, and experts might be able to give you descriptive information about the transformative experience, and about the intrinsic value of the experience. (Harman 2015 and Dougherty et al. 2015

[11] And of course this imaginative act is *de re*, in the sense that you are imagining what the external environment will be like, in addition to your mental life in that environment. I'd like to thank Julia Staffel for discussion (and interesting objections) about the role of imagination in decision-making.

[12] If you lack a value function then you cannot simply represent the decision problem as standard uncertainty. (See Collins 2015 and Paul 2015a for relevant discussion.) In section 4 and section 5 of this paper, I discuss (in a less formal way) the possibility of creating a replacement value function using data from experts. The first problem with this strategy is that the necessary data may not exist. The second problem is that using this replacement may violate the authenticity norm. The third problem is that the values would be for your ex post self, not your ex ante self.

argue that testimony can inform us of intrinsic value.) But it won't give you what you need to grasp the subjective value. That's partly why people tell stories about how they knew what people had told them (about parenting, or going to war, or moving to a country with a very different culture, etc.) beforehand, but there was still a distinctive and extremely important sense in which that did not prepare them for what the experience was *really* like.

Can a person at least use testimony to find out the numerical utilities of her subjective values? So, for example, even if you cannot first-personally grasp the subjective value of becoming a parent before becoming a parent, can you at least use expert testimony to know the numerical range of utilities associated with the descriptions of possible outcomes for you, along with the credences you should attach to the different propositions?

No. The trouble is that the testimony, including the evidence we've got from science, doesn't give us enough of the knowledge we need.[13] This is our second reason for why testimony fails. The problems with incomplete evidence are not specific to the choice to have a child, though this choice provides an excellent case-study.

To start, anecdotal information from friends and relatives should always be regarded with caution. At best, it can provide some evidence of what those close to you think you should do, based on what happened to them in superficially similar situations. Scientific evidence is a better basis for rational assessments, but also has limitations, because, in the first instance, science deals in generalities concerning populations, not with specific, perfectly tailored recommendations for specific individuals in specific circumstances. (I am discussing evidence for choices made at the psychological and sociological level: other kinds of choices, such as choices made concerning biochemical outcomes at the physiological level, might be much more specific.)

That is, psychological and sociological evidence, in its current state, does not give you the individual-specific knowledge you want when making a high-stakes, life-making choice. What you want is knowledge that relates to your particular case, to your subjective values for your outcomes. But what you'll get from current psychological and sociological science is knowledge of a very general sort, concerning average effects, based on data gathered from a sample population. The problem is that the right sort of data, data that is fine tuned to your particular situation, is almost never available.[14]

[13] There are subtle complications here, in part with what we take the standard for sufficient knowledge to be, and in part with respect to the interpretation of statistical information and counterfactuals. See the sections on informed consent, the fundamental identification problem, and finkish preferences in Paul 2014. Related concerns about causation, decisions, and the interpretation of evidence are discussed in Cartwright 2011.

[14] That is, you want finetuned evidence and finetuned knowledge, because this is an incredibly high-stakes decision. Without evidence about your particular case, in this decision context, you lack the relevant knowledge needed for rational action. As Moss (Unpublished) puts it, in the high-stakes contexts of transformative choices, it is much harder for your probabilistic beliefs to count as knowledge (or, we might say that the weak contents of your beliefs count

The case of choosing to become a parent is an excellent example of a well-studied, widely explored life choice where, nevertheless, the science provides messy, unclear results, and gives only the grossest rough-cut estimate of the numerical utilities for any particular individual. (To get a sense of how unsatisfying it is to rely on a superficial understanding of current data to make the life-changing decision to become a parent, see Paul 2014, 124–140, and my reply to Sharadin 2015, below.)

The idea isn't that we can't use science to help us make decisions. Normally, we can. But that's because normally we rely on introspection about our subjective values to close the gap between the messy generalities of science and the specifics of our own personal situation.[15] We finetune the general statistical information we get from science, using introspective assessments to improve our knowledge about our individual situation. But transformative experience creates a special, distinctive problem: In cases of transformative decision-making, introspective finetuning is not available, because of the epistemic inaccessibility of your future subjective values.

Compare choosing between homes in two different cities, where you have no direct acquaintance with the house or the city. You don't even have photos. You can find out what the psychological data tells you about features people tend to care about in a home or neighborhood. You can read what others say about the area. Perhaps you can also find sociological data about where people of your income, race, class, and gender prefer to live. Now assume rationally justified introspective assessment is off-limits. To choose rationally, you can only consult the psychological and sociological evidence and testimony I just described, along with any data formulated for how individuals just like you should choose, to determine your personal preferences for the outcomes.

But what data is this? Where is this trove of results that the careful, non-introspective thinker is to consult? For example, where are you supposed to find a detailed, scientifically based recommendation that, given what we know about people like you, you'd prefer to live on Main Street in the large studio with glazed windows instead of on Franklin Street in the one bedroom with a small kitchen?

Such data doesn't exist, and you can't really wait around in the hopes that it will be created. You need to decide now. So, of course, you want to go and visit each place, in order to become acquainted with each house and neighborhood. Then you can introspectively assess what you prefer and combine that with what you know from psychology and sociology. You

as knowledge but the strong (finetuned) contents don't). You need knowledge, not mere belief, to support reasons for action.

[15] This may not be an especially good procedure even in contexts where transformative choice is not at issue, but it's what we do. Closing the gap successfully even in ordinary cases can be very hard, as cases involving the epistemic difficulties raised by informed consent make clear. For more on the special difficulties raised by combining transformative experience with informed consent, see the *Afterword* of Paul 2014.

visit in order to improve your ability to assess the possible outcomes, and if you are taking the evidence and testimony you gathered into account (as you should), you use it to introspectively finetune the rough psychological and sociological data to fit your particular case.

If nonspecific, general scientific information without introspection isn't good enough when choosing a house, it should be obvious that it's even less acceptable for the irreversible choice to become a parent. Finetuning matters even more when making a high-stakes personal decision.

Sarah Moss (Unpublished) argues that, for probabilistic contents of beliefs, you may treat such contents as reasons for some action if and only if those contents constitute the relevant knowledge for you. As she points out, in high-stakes cases, the standard for what can license your reasons to act is correspondingly high. This is relevant to our parenting example. The incomplete, rough and general scientific evidence about how people respond to being parents doesn't give us enough knowledge to license rational action in this high-stakes choice. I need first-personal knowledge about what it would be like for me to become a parent to close the gap between the messy, general, population-level scientific evidence and knowledge of my particular subjective values. Or I need sufficiently finetuned evidence drawn from a more complete science, evidence that can support probabilistic knowledge about my individual case.

The problems for transformative decision-making based on scientific evidence, thus far, have been practical.[16] But there is another sense in which science cannot give you the knowledge you need to rationally motivate a life-changing transformative decision. In cases of transformative decision-making, we have an in-principle problem with ex ante decisions made for ex post subjective values and preferences.[17] This problem is both a problem for decision theory in its own right, as it involves the possibility of incommensurable preferences across selves, as well as a problem involving authenticity, because authentic choice-making requires the right sort of knowledge about who you are making yourself into.

Another way we might think of the problem is in terms of motivation: without first-personal insight into the self I'll become, how can I be motivated to become that future self, if that self is incommensurable with who I

[16] Strictly speaking, the problem of the generalities of science isn't merely practical, because empirical evidence concerns average effects for members of populations, not individual effects. This (and related concerns, such as the fundamental identification problem) might make it in-principle impossible to get sufficiently finetuned evidence in these high-stakes cases, making it in-principle impossible to get the kind of knowledge we'd need to license action in transformative decision contexts. Here, it depends on just how high the stakes are and on just how specific one can be about the relevant population. See Paul 2014 and Cartwright 2011 for related discussion.

[17] See Pettigrew 2015 and Paul 2015a for further discussion.

am now?[18] I'll discuss this problem briefly in section 5, below, and it will surface in many of my replies in section 6.

5 Rational Decision-making Under Radical Change

The case of choosing to become a parent illustrates how, if you approach a life-making transformative decision intending to assess, understand, and then choose between the different ways your future could be like, you cannot make the choice rationally. This is because of the combination of epistemic and personal change involved in the transformative experience. The epistemic problem arises because the decision is to be made based on a subjectively informed assessment of your possible futures. The personal problem arises because the decision involves the possibility of undergoing an experience that changes you from the self you are now into a different, new self. Together, these problems create a situation where the nature of the experience and the way you'll respond to it involves the possibility of change from one self into an incommensurable, epistemically inaccessible self.[19]

The epistemic situation is analogous to problems with incommensurable theoretical paradigms and the rationality of discovery and theory change from one scientific paradigm to the next. In a life-changing choice, you, the chooser, cannot escape your current perspective. In a Kuhnian sense, you are trapped within the "normal" paradigm of your current self. When confronted with a transformative choice, you must decide whether to replace your current self and its perspective with a new self and that self's perspective. Yet, to grasp the nature of the new self you could become, you must undergo the transformative change, because the nature of that future self is epistemically inaccessible to you before the transformation.

There is no problem of strict personal identity here: we can assume both selves are metaphysically the same person. The trouble is that they are psychologically incommensurable with each other. And so a kind of existential crisis arises: due to the epistemic inaccessibility of the future self's perspective from the current self's perspective, the agent cannot know who she is making herself into. Moreover, due to the incommensurability of the preference change, she cannot adopt a principled decision rule to prefer one set of preferences over the other. Because she cannot step into a neutral first-personal perspective in order to evaluate and compare each possible

[18] Moss (Unpublished, 9.6) explores intuitive ways to understand what contents may count as your *motivating* or *personal* reasons for action. I think this relates to what I described above as the authenticity norm, where there is an intuitive sense in which your personal reasons for action must be contents that you can represent. (This might connect to questions explored by time-slice epistemologists. If at *t1* the agent cannot first-personally represent the self she'd be at *t2*, why should she be motivated to become that self?)

[19] See my exchange with Collins 2015 for related discussion.

successor first-personal perspective, she cannot formulate a higher-order rule that will adjudicate the decision for her.[20]

One way to represent this problem from a decision-theoretic perspective is to argue that, before your choice, if you cannot represent (or "see") the outcomes in the necessary way, you lack psychologically real, rationally assignable preferences about your post-transformation outcomes. We can describe this as a violation of a standard axiom of decision theory, the completeness axiom.[21] Completeness requires agents to have definite preferences for any gamble, such that the agent rationally prefers to take the gamble, or disprefers it, or is indifferent towards it. (We might think, loosely, of the different selves an agent could become as the different possible "prizes" she might win from a gamble.) On an account where we understand the agent's preferences to be psychologically real, an agent who cannot have rational preferences about her post-transformation selves lacks the needed attitudes, and so the axiom is violated. This is simply another way to put my point that an agent without a value function for transformative outcomes is an agent without a (standard) model for a rational decision.[22]

The possibility that the transformative change involves incommensurable self-perspectives and the replacement of one's current self with an epistemically inaccessible, incommensurable future self is what creates an in-principle problem for using testimony from others and for using future values determined by experts. (Note: this is different from the practical problem I discussed in section 4, the problem that, given the state of current psychology and sociology, relying on testimony and evidence won't get you sufficiently finetuned knowledge.)

The rabid fan of testimony will advise you to dispense with introspection. You are to make your decision entirely on the basis of what others tell you about their experiences, or about what scientists have discovered about others who have gone through similar experiences.

But ignoring introspection has serious costs. The first cost involves deciding without knowledge of the value of the lived experience. As John Collins argues in his contribution to this volume, an agent might be rationally neophobic, that is, she might be rationally averse to making a decision without grasping the subjective values of the outcomes she could bring about. In high-stakes cases, it's especially plausible that we'd be rationally averse to making life-changing decisions without knowing the subjective values of our possible futures. Such aversion also has connections to the desirability of authenticity and the role of motivation in personal decision-making.

[20] The source of this problem is the fact that the agent only learns what she needs to know, that is, she only has access to her post-transformative perspective, *after* the transformative experience. See Pettigrew 2015 and Paul 2015a for discussion.

[21] I thank Alan Hájek for suggesting this way of framing the discussion.

[22] For related discussion, see Collins 2015 and my reply below.

The second cost stems from the fact that you are to use data and testimony to determine your ex post values. The trouble with such testimony is that it is given by those who have undergone the experience, and so concerns the ex post self, the self that would result from undergoing the experience. That is, it's relevant to who you'll become as the result of the transformative experience, and purports to tell you how you can expect you'd assign values to the outcomes at that time. But from your ex ante perspective, that is, from the perspective of the self who makes the decision, given the transformative nature of the experience, your future self could be incommensurable with your current self. In other words, the subjective values given by testimony for the outcomes are values of your merely possible future selves, not your current self, the self who is making the decision. And the trouble is that, when you face a transformative choice, even if expert testimony can tell you what to choose, you still face an existential problem: Will you really be happier after the transformative change—or will you just be a different self?[23]

This matters, because you might find some of your possible future selves epistemically alien to who you are now. For example, you might be a career-driven childfree person who finds small children annoying, but if you went through the transformative experience of becoming a parent, you'd enjoy spending time around babies and look forward to your hours at the playground. As your current, childfree self, you might find such a future self repulsive—she is deeply epistemically alien to who you are now. If so, why should the subjective values of that merely possible future self be relevant to who you are now?[24] Put another way, even if we assume that scientific data and testimony could provide you with your future self's values, that is, with your ex post values after the transformation, this isn't enough. If, at *t1*, you want to choose consistent with your current self's preferences, and if she prefers to remain who she is, what matters for the rationality of your choice at *t1* are what you determine ex ante about the values for the self at *t2*, not what the ex post values are for the self at *t2*.

Normally, as I discussed in section 3, to assess our values for a future lived experience, we imaginatively project ourselves into the future possibility, and prospectively assess it to determine, ex ante, our ex post values. This works if there is no transformation of the self, and it might even work if you could assess the transformation from a "self-neutral" perspective.[25] But the combination of epistemic and personal transformation in transformative choice makes this impossible. You cannot imaginatively project yourself "through" the experience to assess your future self's first-personal

[23] Again, we can see this as a problem with completeness. It can also be seen as a problem with van Fraassen's Reflection Principle. In Paul 2014, I discuss the problem in the Afterword, in the section "Finkish Preferences."

[24] For related discussion, see Barnes 2015b, Briggs 2015 (and my reply below), and Paul 2015a.

[25] Compare Chang's (2015) suggestion that we adopt a "master utility" function in this situation.

perspective, for it is inaccessible to you, and thus you cannot prospectively model your transformed first-personal perspective in that possible outcome. As a result, you risk forming yourself into an alien self when you undergo a transformative experience. Relying solely on data or testimony to tell you your future values obscures this fact, because it merely tells you the values of your transformed self, not the values for who you are now. I discuss this cost further in my reply to Briggs, my reply to Dougherty, Horowitz and Sliwa, my reply to Harman, and my reply to Chang.

The problem of inaccessible, alien future selves also brings out how exploring transformative experience in the context of a psychologically rich decision theory or a formal epistemology gives us a richer perspective on the way that rationality and authenticity are intertwined with transformative choice. To act authentically, you must be adequately informed of the nature of the outcomes of your choice: authenticity can require knowledge under the subjective mode of presentation. Moreover, consistent with more traditional notions of authenticity, your values should be formed by who you are and what you know about yourself, and by your core principles and commitments. Simply adopting values given to you by external authorities is inauthentic. The same holds for moral values: you need to understand and grasp these values for yourself in order for them to be *your* moral values.

There *are* ways to rationally and authentically prefer to disprefer your current values, for example, when your higher-order preferences are to change your current higher-order preference structure. For such a decision to be authentic, you must understand that you are choosing to discover a new self, that is, you must reflectively decide to replace your current self with the new self that you will discover, knowing that you don't know what your future self will be like. (In an important sense, you prefer to annihilate your current self.) In *Transformative Experience* (2014), I argue that this preference structure involves a preference for discovery and revelation, and may be able to resolve the tension between rationality and authenticity in some transformative choices.

Once we see how epistemic and personal transformation work, it becomes apparent that some of life's biggest decisions are life-making choices. The deep philosophical problem is that these life-making choices involve experiences that teach us things we cannot first-personally know about from any other source but the experience itself. With many life-making choices, we only learn what we need to know after we've done it, and we change in the process of doing it.

The lesson I draw is that an approach to life that is both rational and authentic requires epistemic humility: life may be more about discovery and coming to terms with who we've made ourselves into than about carefully executing a plan of self-realization.

6 Discussion and Replies

6.1 Formal Epistemology and Decision Theory

John Collins (2015) argues that decision theory needs to make conceptual room for the rationality of neophobia, that is, for a rational aversion to the new and unfamiliar—just as we need to make room for psychological facts about agents' attitudes concerning risk or ambiguity.

He explores how, for a decision concerning an epistemically (but not personally) transformative experience, we can model the choice using a familiar proxy as a synthetic lottery, and represent an agent's preference for the synthetic lottery as a neophobic preference. We might then develop a decision rule for such preferences. I find his positive proposal and diagnosis of the source of neophobic preference structures creative and insightful, although I don't think that the deep source of neophobia is indeterminacy. But despite our differences, I completely agree that we need to make room in decision theory for neophobic and neophiliac preference structures, and find his decision rule for neophobic agents compelling.

Collins introduces his argument with a discussion of the metaphysics of utilities, arguing that I am committed to a species of nonconstructive realism for utilities. Constructive realists take an agent's utilities to be metaphysically constituted by her preferences. Nonconstructive realists deny this. Since I am not committed to nonconstructive realism, I'll start with a brief discussion of his claim and then move to a positive discussion of his substantive proposal.

I take preferences to be psychologically real, but I have no stronger claim about the metaphysics of utilities. On my view, an agent facing an epistemically transformative experience lacks psychologically real, rationally assignable utilities concerning the epistemically inaccessible outcomes because she cannot represent (or "see") the outcomes in the necessary way.[26] If utilities are understood in constructive realist terms, then on this view, psychologically real preferences don't exist either. The agent lacks the capacity to grasp or entertain the natures of the relevant outcomes, and thus lacks the desires and beliefs needed for her to have psychologically real preferences.

If an agent lacks the ability to have psychologically real preferences for a decision situation, we find ourselves without a model for rational decision in that situation. That is, we lack the capacity to represent and model the choice "in the usual and obvious way, as a gamble that might yield any one of a range of possible utility values, depending on how things turn out to be" (2015, 287).[27]

[26] Here I adopt a nice locution from Carr 2015.

[27] None of this rules out the possibility that the agent could have preferences based on confusion, or on false, incorrect, or otherwise inappropriate beliefs. Such preferences won't help us with rational decision-making in this context.

So the nature of the problem with epistemically transformative experience does not depend on whether one is a constructive or a nonconstructive realist, for an agent without rational preferences is also an agent without a rational decision model. But no matter: I agree with Collins's main thesis, that decision theory needs to make conceptual room for neophobia.[28]

Collins proposes that we explore the contours of the problem of the inaccessibility of the epistemically unknown outcomes using a synthetic lottery as a replacement model.

> (1) For any possible utility value x that the epistemically transformative experience may turn out to have for the agent, there is a possible outcome to the lottery that is both (a) experientially familiar to the agent and (b) has a utility that is (arbitrarily) close to x. (2) The chances of the various possible outcomes to the lottery are weighted so as to correspond to the agent's subjective probability distribution over the range of possible utilities that the epistemically transformative option *A* may turn out to have, whatever that subjective probability distribution happens to be. (290)

Perhaps an agent can determine her expected utilities using a synthetic lottery over the familiar space of options, then map that choice onto the space of epistemically transformative outcomes to determine her expected utilities for the transformative decision. Then, even if we can't grasp the characters of the unknown outcomes in order to determine their utilities and the expected utility of possible acts involving them, we can construct a proxy model with the same structure, based on mapping known utilities for known options to unknown utilities for ungraspable options in order to use it as a proxy for the model we'd have constructed if we were able to explore the space of options in the usual way.

This is an extremely interesting suggestion, and I discuss its implications in detail in Paul Unpublished. Of course, we'd need to know how to construct the mapping from the space represented by the proxy model to the space of the epistemically transformative situation, but let's assume that this is discoverable empirically or knowable via testimony from others.

As I argued above, the need for the model doesn't turn on whether one is a constructive or nonconstructive realist, for if one is a constructive realist the proposal is just a little more radical: instead of taking the values of the outcomes as synthetic replacements, take the entire model, the synthetic lottery itself, as a synthetic replacement for the epistemically inaccessible model of the decision. In constructive realist terms, we can think of the

[28] This small dispute is probably due to the fact that Collins was working from a draft copy of my book manuscript rather than the final version. In the draft copy I did not specify my preferred metaphysics of preferences and utilities.

synthetic lottery as providing us with synthetic preferences from which we construct synthetic utilities.

The deeper and more interesting issue here is whether there is a dimension of value that is not captured by the numerical values for utilities (whether or not they are constructed from preferences) that are assigned to decision outcomes via testimony. We can understand neophobia as arising from a situation where we cannot first-personally grasp or entertain the qualitative nature of the lived experiences described by the outcomes in the model. Even if someone else can tell me what the numerical values of my utilities are for outcomes involving durian-eating, there's something important about these outcomes that mere description or testimony can't communicate. Knowledge of numerical values for utilities alone isn't sufficient.

We might diagnose the source of the problem as one with our inability to form preferences, or to grasp or entertain the subjective values of these outcomes, as I do. (For more on these subjective values, see my discussion in section 2.) Or we might diagnose it as an inability to resolve an indeterminacy about the possible utilities I might experience.[29]

Collins argues that the best response for the defender of orthodox decision theory is to regard the decision problem with epistemically transformative outcomes as a problem involving a "basic and irresolvable" indeterminacy about utilities.[30] He writes:

> That's why the orthodox decision theorist's suggestion that we elicit her utility for the transformative outcome by the method of constructing a synthetic lottery need not always work. It's not possible to elicit a sharp determinate value for the utility of an outcome when it is just a fact that no such unique value exists. The synthetic lottery may yield some unique number, but so what? It's providing an answer to a different question. (2015, 287)

He suggests that the irresolvable indeterminacy could stem from the agent's inability to assess and compare the values of different outcomes, given her lack of acquaintance with them.

He then argues that, nevertheless, there is a rational basis for the neophobic agent to opt for a synthetic lottery over a lottery with unknown outcomes, drawing on Isaac Levi's work on the Allais problem. I find his proposal interesting, and it should certainly be included in the set of options we should consult when faced with a transformative decision problem.

[29] Collins, in conversation, points out that he and I might not be at odds here. If we both endorse constructive realism about utility then these would be two ways of saying the same thing: indeterminacy about utility would arise out of the inability to form psychologically real preferences.

[30] See Barnes 2014 for an excellent discussion and defense of fundamental metaphysical indeterminacy.

However, it is important to see that, whether we take the source of neophobia to be our inability to grasp subjective values, or we take it to be our inability to resolve indeterminacies about utilities and employ synthetic lotteries, in transformative decision contexts, the fundamental tension between rationality and authenticity remains.

In fact, Collins's proposal for modeling epistemically transformative decisions gives us a new and especially clear way to bring out this fundamental tension. We can see this by exploring how a synthetic model might be used to make a decision that is both epistemically *and personally* transformative.

Recall that the synthetic lottery works by substituting unfamiliar outcomes with familiar outcomes with the same utility. Now consider using a synthetic lottery for a transformative choice, like the choice to become a parent. If I'm neophobic, then even a perfectly constructed synthetic lottery will fail to yield an expected utility relevant to my decision problem.

In this case, I'm to use a synthetic lottery to decide whether to become a parent.

For example, let's say I've already had the transformative experience of becoming a doctor, and so I am familiar with the kind of epistemic and personal transformation involved in such a life change. Now I am considering becoming a parent, and I'm told that the possible outcomes of this new life change can be assigned utilities that mirror, in the right way, the values of the possible outcomes associated with my becoming a doctor. So I can use the synthetic lottery to determine that, if I maximized my utility by becoming a doctor, I will maximize my utility by becoming a parent.

The important detail in this case is that I'm using the synthetic lottery to decide whether to undergo an experience that could result in an alien future self, a self whose first-personal perspective is epistemically inaccessible to me now. What's familiar to me is the dramatic personal change involved with the experience of becoming a doctor. But for just that reason, I know that the utilities attached to the outcomes of becoming a parent carry with them the possibility that they are utilities for a self who is psychologically alien to who I am now. My choice is whether to make myself into this future self.

We might, then, understand neophobia as a perfectly rational aversion to the possibility of becoming an unknown self—and neophilia might be understood as a kind of self-hatred, that is, as a desire to replace one's known self with an unknown self. And so the central tension between rationality and authenticity remains.[31]

[31] I might use the synthetic lottery as a model for a decision between revelation and the status quo, where I am choosing to discover what it's like to become a new kind of person. See Paul 2014, Ch. 4, "The shock of the new." What the synthetic lottery can capture here is what the agent is familiar with, that is, what it's like to become a new kind of person, and whether this is desirable.

Like Collins, Jennifer Carr (2015) focuses on epistemically transformative experiences. She develops a model for epistemic transformation in epistemic decision theory. As Carr puts it, the question is:

> How can there be a decision theory for partial credence functions, when decisions hinge on possibilities the agent can't entertain? The problem is not uncertainty: it's not simply that the agent is unsure of the outcomes of her actions. Rather, the problem is limited conceptual resources: there are some possibilities that the agent can't "see," propositions she isn't in a position to entertain. (217)

An epistemic decision theory can be used to develop and model normative epistemic facts about agents, and a rational epistemic decision model can be understood as a model for epistemically rational belief and behavior.

Epistemic utilities (values) are understood in terms of what is epistemically desirable or in terms of epistemic goods, and credences are strengths of belief. The idea is to give epistemology an overtly decision-theoretic framing in order to discover and represent epistemological norms for rational agents, such as which epistemic states they should adopt given their evidence.

In Paul 2014 and 2015b, I formulate the questions about transformative decision-making neutrally, but they can be explored within more specific frameworks, like that of epistemic decision theory. The problem that transformative experience raises for epistemic decision theory concerns our epistemic guide for life-determining and life-changing choices, that is, our epistemic norms for how we should believe and act in order to maximize our epistemic utility when we make big life decisions.

Epistemically transformative experience, viewed through the lens of epistemic decision theory, generates a problem. How? Because the transformative nature of the experience rules out the possibility that the agent can subjectively grasp her utilities, since she cannot "see" these utilities or entertain the relevant propositions.

A solution, of course, is for an expert to tell the agent what her utilities are. However, as I discussed above, in most cases, such expert advice doesn't exist. And even if it does, simply knowing the numerical assignments for outcomes via testimony is not sufficient for an agent to fully grasp the nature of the outcomes, and thus is not sufficient for her to determine her expected subjective value. There's a dimension of the subjective value that can only be grasped by having the right sort of experience. It's an open question whether this arises from the subjective value s for outcome o not being representable by numerical value n, or whether it arises from the psychological fact that the agent cannot grasp or understand everything represented by n without the requisite experience.

The central tension between rationality and authenticity, then, can be illustrated in epistemic decision theoretic terms. The relevant norm of

epistemic rationality requires an agent to act in accordance with maximizing her expected epistemic utility. The relevant (cultural or practical) norm for authentic choice and action defines an agent as informed only if she can first-personally grasp, understand and subjectively value her outcomes. This norm of authenticity holds that agents should make life-defining choices as informed agents, which means an agent should knowingly make her life choices partly by understanding who she is, who she wants to become, and what her lived experience and the lived experiences of others could be like as the result of her choices.

In transformative decisions, if the agent's expected utility is only available to her via expert testimony, then to meet the epistemic norm the agent must make a decision about her future without a first-personal grasp on her future subjective values. To act (epistemically) rationally, the agent must violate the authenticity norm.[32]

As the clash between norms suggests, the most interesting version of the problem involves experience that is both epistemically and personally transformative. Such experience concerns big life decisions, and highlights what is really at stake in this discussion, for it highlights how an agent's ability to grasp the first-personal perspective of her future self is a central part of her ability to grasp who she is making herself into. The problem is that knowing the numerical utility for the self who results from a transformative personal experience is not sufficient for an agent to know who that self is, and whether that future self is desirable from her current point of view, for the personally transformative experience is *also* epistemically transformative. As a result, the fundamental tension between epistemic rationality and authenticity is highest in the cases that matter the most from a real world perspective.

Carr's approach to the issues involving epistemically transformative experience and epistemic decision theory focuses on how we are to understand the way agents with partial credence functions are to expand their views when they experience an epistemic transformation. How are we to understand the epistemic norms for an agent who discovers new propositions or can see new possibilities? She frames the epistemic transformation in terms of the agent's credence function changing domains from one set of propositions to a new, expanded set, and defends the view that credence functions with different domains are sometimes comparable.

On Carr's view, we can think of epistemic decision theory as including normative epistemic constraints, defined by the value of conceptual resources, on credence functions. In her paper, Carr focuses on accuracy-first epistemology, and explores different constraints under which credence functions with different domains are comparable.

[32] Interestingly, perhaps the best way for the agent to respond to her situation is to authentically grasp that she doesn't know her future subjective values and choose with respect to her desire for revelation. See my chapter 4, "The shock of the new," for discussion of this idea, as well Campbell 2015 and Paul 2015a.

I found her preferred proposal for modeling epistemic transformation, where we regard nonattitudes towards propositions as having the same utility as having maximally unopinionated attitudes to those propositions, to be an interesting and important way to understand some of the epistemological implications of epistemic transformation. The question I want to pose for her, however, asks how such an approach affects the way we should understand transformative decision-making that involves personal change.

In particular, how does the proposal pan out in a context of epistemic personal discovery, that is, a context where the agent discovers a new way to conceive of herself, and of the world in relation to herself? As I noted above, one of the central puzzles of transformative decision-making concerns decisions involving epistemically and personally transformative experiences. The puzzle arises when an agent's ability to see new propositions generates new, incommensurable personal preferences. How should we regard the utility of remaining maximally unopinionated in this context?

Moreover, what are the constraints on epistemic expansion when an agent extends her conceptual grasp on the world in a way that gives her a new, incommensurable representation of herself, or of the way the world is? What are the constraints on epistemic expansion when an agent changes (or an agent contracts) her conceptual grasp on herself and the world in a way that gives her a new, incommensurable representations of them? Such epistemic change, tied with personal change, might be more complex than mere contraction or mere expansion, if, as seems likely, some propositions are simply replaced by others.

Carr's proposal explores foundational questions about the epistemology of epistemic discovery and change. It would be good to know more about how it bears on more complicated cases involving change that is at once epistemic and personal.

In their engaging paper, Tom Dougherty, Sophie Horowitz, and Paulina Sliwa (2015) (hereafter, "DHS") defend the decision-theoretic homeland by raising a series of objections to my argument that the phenomenon of epistemically transformative experiences contributes to a new and distinctive problem for decision-making. Their central claim is that epistemically transformative experience "does no special work" because it is possible to estimate the intrinsic value of outcomes that involve epistemically transformative experiences even if one has not had that type of experience. They then explore ways of framing transformative decision-making as involving the kind of uncertainty that can be represented by imprecise credences. I found their discussion useful and informative in its exploration of some of the controversial features of the debate over transformative experience. However, while their arguments raise thoughtful objections that deserve attention, in the end, I reject their central claim, arguing that they have conflated subjective value with a different notion of value, intrinsic value, and I deny that the epistemic problems raised by transformative decision-making

are simply epistemic problems that can arise in cases where subjective experiences do not feature.

DHS target the following premise:

> (Premise 3) If an agent does not know what it is like to have an experience, and this experience is constitutive of a "phenomenal outcome," then she cannot rationally judge the subjective value of this outcome for her.

Premise 3 isn't a premise I actually defend: it is a premise culled from their characterization of my view. However, I shall accept it for the purposes of our discussion.

DHS reject Premise 3. They argue that distinguishing phenomenal character from value

> allows us to also draw an *epistemological distinction* between awareness of an experience's phenomenal character and awareness of its value. . . . Once drawn, this epistemological distinction should make us suspicious of Premise 3. From the fact that an experience is epistemically transformative, it only follows that the agent is not antecedently in a position to know what the experience would be like. This is consistent with the agent being able to rationally estimate the experience's value. (307)

However, distinguishing between phenomenal character and value won't do the work they want it to do. I agree that we can make an epistemological distinction between an agent's awareness of an experience's phenomenal character and her awareness of its subjective value. But this does not undermine the fact that, in the cases we are examining, experience is necessary for an agent to grasp and represent the subjective value of the outcome.

Recall that the subjective value of an outcome (a future experiential state of an agent) ontologically depends on the nature of the lived experience that constitutes it. We can certainly distinguish between an agent's being aware of, or grasping, the nature and character of the lived experience that constitutes the outcome, and the agent's being aware of, or grasping, the subjective value that depends on the nature and character of the lived experience. But the agent's inability to grasp and represent the propositions concerning the nature and character of the outcome, or, we might say, her inability to grasp these propositions under the right mode of presentation, implies that she cannot be aware of the subjective value of the outcome.

And, of course, subjective value is precisely the sort of value we are concerned with in this discussion. As I put it in Paul 2015a,[33] for many

[33] Another interesting distinction we can make involves the distinction between knowing the numerical value that an expert tells her to assign to the subjective value and the epistemic act of grasping or being aware of the subjective value.

life-changing decisions, the agent wants to assess her options by assessing the subjective value of her possible future lived experiences. Ideally, her assessment involves a determination of the subjective value of each possible outcome of her decision, that is, the subjective value of each possible lived experience, by imaginatively grasping what it would be like for her to live in those possible circumstances. This does not imply that subjective value merely concerns the character of one's internal mental life, or that it is somehow merely self-interested. Subjective value, instead, is concerned with the nature and character of an agent's lived experience, which can include her experience of her environment (and can include assessments of subjective values for other people as well). (For an expanded defense of the importance of subjective values, especially in contrast to hedonic values, see my reply to Kauppinen, below.)

But DHS claim that, even if the agent doesn't know what the possible lived experience will be like, she can still rationally estimate its value. How can this be true? If the agent is not antecedently in a position to know what the experience would be like, she cannot imaginatively grasp or otherwise represent what it would be like for her to have that lived experience, and so she cannot grasp the outcome's subjective value. This follows from the fact that she cannot grasp and represent the propositions concerning the nature and character of the lived experience.

What's gone wrong here? The problem is that DHS have subtly conflated *intrinsic* value, which they define as value that is had by an outcome "in virtue of its [that outcome's] intrinsic properties" with *subjective* value. And in fact, in the rest of their discussion, their arguments are entirely focused on how one can estimate the intrinsic value of an outcome (for example, by using testimonial and behavioral evidence). Intrinsic values, as they define them, can be communicated by testimony, description, and behavior: no experience needed.

But whether the agent can grasp the *intrinsic value* an outcome might have, a value it has solely in virtue of its intrinsic properties, is irrelevant to my argument that an agent cannot rationally determine her expected subjective value for that outcome.[34] Premise 3 is formulated as a claim about *subjective* value. Indeed, all of my arguments concerning transformative experience and the rationality of choice and decision-making are focused on the agent's decisions as decisions framed in terms of the subjective value of outcomes and the expected subjective value of acts. This does not imply that intrinsic values aren't to be included in a global assessment of expected utility. It's just not the type of value that matters most in this context, because it's not the type of value involved in the contents that we are seeking to know when making this life-making decision.[35]

[34] In any case, we need to stay away from affectless, experience-free characterizations of value.
[35] See Moss Unpublished for related discussion, and an argument that transformative decisions can fail a probabilistic knowledge norm.

Indeed, one way to see that subjective value is importantly different from intrinsic value is to draw out how, for many life-changing decisions, extrinsic properties must be part of what is assessed. For example, it doesn't seem to be intrinsically more valuable to be a person who can see than to be a person who is blind. But one might argue that, given the way our society privileges those who can see, it is *extrinsically* disvaluable to be blind. If so, the extrinsic disvalue will likely be determined by the nature of one's lived experience as a blind person in our society, and should be assessed by the agent when she makes her global determination of the subjective value of the lived experience.[36]

So, one reason why the phenomenon of epistemically transformative experiences does special work is that it causes an agent's subjective value function to be partial. This also helps to clarify why attempting to represent transformative decision-making in terms of imprecise credences is insufficient: the problem, at least in the first instance, is with subjective *values*, not credences. (That said, DHS are entirely correct in their point that the sparseness and messiness of the evidence also leads to problems with credences. It's just that this isn't the only problem, or even the most destructive problem, for transformative decision-making.)

But there is a second important role for epistemically transformative experience: its role in the problem of transformative decisions involving life-changing experiences and the incommensurable preferences they create. The epistemic inaccessibility of the nature and character of her future lived experiences means that the agent facing a transformative decision is epistemically isolated from her future. She cannot look forward and see what it will be like to live in her future circumstances, or indeed, to see who she'll become. This means, first, she cannot neutrally examine her future, incommensurable preferences and compare them to her current preferences in order to develop higher-order preferences for how she wants her life to evolve. And second, because she cannot first-personally know who she is making herself into, she faces a distinctively existential problem, a kind of crisis of rational self-control and self-development.

As DHS point out, "if future preferences are rationally significant for present choices, then this means that one would have to either concede that decision theory is not fully comprehensive as a theory of practical reason or to find a way to extend decision theory so it provides guidance about how to act in light of preference-shift" (319).

Grant that future preferences are indeed rationally and personally significant. Then, in cases of transformative decisions involving life-changing experiences, an agent's current preferences might shift so that she finds herself in new scenarios with new, incommensurable preferences. But, when she faces this decision, the epistemically transformative nature of the

[36] For related discussion about the role of social conditions in transformation and the nature of lived experience, see Barnes 2015a.

life-changing experience blocks her from using her first-personal epistemic capacities and knowledge to mentally evolve herself forward and assess these new counterfactual scenarios from her first-personal perspective before she acts. So she lacks the capacity to develop a decision-theoretic guide for how to act.

As I described above, in section 5, the epistemic inaccessibility of the future possibility of radical self-change creates a first-personal, existential analogue of the familiar theoretical issues involving scientific discovery and conceptual revolution. We can describe the problem as a first-personal version of a Kuhnian paradigm shift, that is, where the agent undergoes a first-personal conceptual and preferential paradigm shift as the result of a transformative experience. The combination of epistemic and personal transformation leads to disaster for decision theory as a guide for action. As Bas van Fraassen (1999) puts it:

> What is the theoretical problem here for decision theory? Imagine that we contemplate a decision in favor of a certain option, of which the outcome is, by our present lights, ourselves speaking and thinking nonsense, while faring materially much better. Does that make sense? It seems that this would be a true abdication of reason, and not just because we classify it as a bad outcome, an outcome with a low value. Rather, we must doubt that we are coherently framing a decision for ourselves here. For how can we tell that what we *now* see as material welfare in that future will be cognized as such *then*? And if it is not, what about a future in which we are by our *present* lights well off, and by our lights *then* miserable or suffering a great loss? . . . Turn back now to the person totally inside a certain scientific world picture which is becoming burdened with more and more blatant anomalies, severe calculational difficulties, failing predictions, epicycle-laden explanations, and so forth. An alternative appears, some people are beginning to talk about a strange new theory which makes absolutely no sense, and violates the most basic commonsensical expectations of what nature can be like. What is still classified as a satisfactory outcome? To solve the problems of course; taking some absurdity seriously certainly does not count as a solution, and even if it did one would have to be an imbecile to expect it to be vindicated by future experiments. If that person stops a moment to envisage himself converted to the strange new ideas, he sees himself in imagination stooping to irrationalism, he hears himself babbling with (*c'est le bouquet!*) an air of having explained the inexplicable. (75)

The conceptual problem for revolution in theory undermines the idea that the individual proceeds rationally as she makes her first-personal life-changing decisions just as much as it undermines the idea that science proceeds rationally through theoretical and conceptual revolution. Because of the epistemic inaccessibility of the nature of her future, an agent cannot step back and neutrally or selflessly compare her current epistemic situation and her current preferences to her future epistemic situation and her future preferences. She cannot step outside of her conceptual framework or her "preferential" framework in order to develop a consistent guide for the radical shift in perspective and preference that transformative experience implies. Echoing a phrase of van Fraassen's, there is no first-personal view of the self that is invariant under such transformations.

In closing, while I've been critical of DHS's argument, I want to emphasize that their objections are well-taken, and there is much in their discussion that I find suggestive and interesting. In particular, their discussion in section 3 of their paper raises a new and interesting puzzle concerning the rationality of decisions that might have temporally distant transformative implications.[37] They also raise important questions in section 4.1 of their paper about when it can be authentic to use the testimony of experts in place of one's own first-personal judgments for life-defining choices such as becoming a parent. These are subtle issues about what authenticity can license that deserve further exploration.[38]

6.2 Social Choice, Social Justice, and Social Identity

Rachael Briggs (2015) explores ways of making sense of wellbeing in contexts of personal transformation. She is interested in working out the structure of intrapersonal comparisons in cases of transformative choice, where the decision is whether to undergo a personally transformative experience.

While the puzzles involved in interpersonal comparisons have been well explored in social choice theory, the problems with intrapersonal comparisons have been neglected, and she sheds new light on the nature of intrapersonal comparisons by assessing them against interpersonal theories of comparisons between different individuals. Her paper gives us an excellent assessment of the terrain for interpersonal comparisons of utility, and gives us an important new way to think about these problems by showing us how to use this map to explore related problems with intrapersonal utility changes, including places where interpersonal utility comparisons and intrapersonal utility comparisons diverge.

In her paper, Briggs sets aside the complication with epistemically transformative experiences in order to focus on the distinctive issues concerning

[37] Moss (Unpublished) argues for a solution to this problem.

[38] For related discussion about authenticity and expertise in transformative decisions, see Collins 2015, Campbell 2015, Paul 2015a, and Paul Unpublished.

personally transformative preference change. She explores the questions around how we should understand wellbeing in cases of transformative personal experience and transformative personal choices, under the assumption that we have full information about the nature of these transformative preference changes across worlds and times.

Briggs points out that intrapersonal utility comparisons involve the notion of prudence, where prudent choices are those that the agent believes will maximize her overall wellbeing, rather than her wellbeing at just one (short segment of) time. So how we decide to manage intrapersonal utility comparisons is intimately related to how we think about prudence. We might even think of the central complication in personally transformative choice, the problem of how my current self is to regard decisions about what future self to become, as a distinctive kind of prudential complication. Should I remain the self that I am? Or should I become a new self? This raises a further question: how do our decisions about prudential wellbeing map onto first-personal decisions at a time, where those decisions affect our preferences for self-change?

Such questions are versions of the question of whether I should act conservatively or liberally with respect to allowing myself to change. As Briggs points out, "without intrapersonal utility comparisons, preference satisfaction theories are ill-placed to explain why and how I should defer to my future preferences" (202). And once we see the need for intrapersonal utility comparisons, questions arise about how to adjudicate the preferences of my current self versus my merely possible future selves, especially if I currently find those possible future selves repugnant or alien. As Briggs argues, unless our concept of rationality is disappointingly thin, we must find a way to take these future preferences into account when we make decisions. But the question of transformative preference change across selves is exactly this: *how* are we do so, and what is the role of authenticity, intrinsic subjective value, and self-knowledge in all of this?

The issue arises in Briggs's discussion of the rigidification strategy for intrapersonal utility comparisons across worlds in cases of transformative choice, and for intrapersonal utility comparisons across times in cases of personally transformative change (= personally transformative experience independent of whether the change was chosen).

The rigidification strategy is a strategy that selects a particular preference ordering as privileged. Briggs explores a hybrid strategy as a way to manage cases of personally transformative choice and personally transformative change. If we use this strategy, in cases of personally transformative *choice*, we should rigidify to the actual world, that is, we should assess the choice in light of the person's actual preferences, rather than her counterfactual preferences. This allows us to say that what is good for the parent is that she has a child—because she actually has a child. Similarly, what is good for the Deaf adult is that he does not have cochlear implants—because he never received cochlear implants. The idea here is that these agents are

better off in their actual scenarios rather than their counterfactual scenarios, because what is good for them is based on their actual preferences. Their counterfactual preferences are irrelevant.[39]

In cases of personally transformative *change*, on the other hand, we rigidify eternally, that is, we rigidify across times instead of rigidifying to the now. So I'm describing the strategy as "hybrid" because we privilege our local world, but we don't privilege our local time, in the sense that my actual preferences are privileged, but my preferences now are not privileged. "[W]e interpret my preferences as preferences about what happens at *t*, so that what is good for me is to get peanut butter when I have peanut butter cravings, and Vegemite when I have Vegemite cravings" (214).

As Briggs notes, the hybrid strategy has mixed results for an assessment of wellbeing. "Rigidifying allows the preference satisfaction theorist to give intuitively correct answers in a variety of cases involving transformative choice. In cases involving transformative experience [personally transformative change], the rigidifying strategy is less promising, since it requires us to arbitrarily favor an agent's preferences at one time over her preferences at all other times" (p. 213).[40] The point here is that there is a natural sense in which an assessment of a person's wellbeing can depend on how they are in the actual world but might be independent of how they might have been. But we don't have a non-arbitrary way to say that a person's wellbeing depends on how they are at one time rather than how they are at another time, and in contexts of transformative change we are presented with the need to perform precisely this sort of assessment of wellbeing.

I found Briggs's discussion very illuminating: it highlights the deeper structure of the way we must think about intrapersonal comparisons and raises a cluster of interesting questions. However, I have concerns about the rigidification strategy for transformative choice. There is a natural sense in which we want to privilege the actual world in some cases. Recall the discussion from section 5, of the childfree person who, ex ante, finds the possibility of being a happy parent repulsive. In such a case the rigidifying strategy yields the answer that it's best for her to be as she is, even if the scientific evidence and testimony tell her she'd be a happy parent.

But why accept the Panglossian premise that my actual situation *always* has an advantage over my counterfactual situation? Briggs does not defend the rigidification strategy for all cases of transformative choice, she is merely concerned with showing that in some cases, it provides an intuitively satisfying answer. I agree that in the particular cases she describes the answer seems satisfying. But I am unsatisfied with the justification for the rigidification to the actual world, because I don't see how to apply it in a principled way.

[39] For an interesting, related argument, see Barnes 2015b.

[40] In correspondence, Briggs notes that she doesn't endorse the mixed strategy, although she thinks it has a lot going for it.

The question is particularly pressing in the context of the solution to temporal variation in cases of personally transformative change. Consider the case of choosing to have a child. I think Briggs would agree with me that it is parochial to think that satisfying my preference for being childfree at the time when I prefer being childfree is somehow better than satisfying my preference for being a parent when, at a later time, I prefer being a parent. But then why should I privilege my actual, childfree preferences over the preferences of a counterfactual, pro-parental self? Isn't it equally parochial to think that satisfying my preference for being childfree in a world where I prefer being childfree is somehow better than satisfying my preference for being a parent in a world where I prefer being a parent? If so, the rigidification strategy is the wrong one to pursue in this case.

What settles which preferences we should privilege? Which preferences matter most? And why should actuality play a special role in the privileging? My conclusion, drawn from Briggs's rich and interesting discussion, is that in transformative contexts there may be no non-arbitrary way for a preference satisfaction theorist to adjudicate between sets of preferences to give an account of wellbeing and prudence.

Elizabeth Barnes (2015a) develops the idea that whether an experience is transformative can depend on social conditions. She argues that we can distinguish degrees of transformation, and that social conditions can affect one's transformation in a way that has implications for social justice.

Barnes grants that whether a person undergoes a transformation when she has an experience depends on her previous experiences and on what she is like, but she argues that it also depends on the nature of the environment that she is in. Features of an agent's environment can contribute to her transformation because these features can be causally relevant to her psychological response.

I agree with much of Barnes's excellent paper: I will confine myself to drawing out three implications of her central points.

First, Barnes's examples involving disability show how properties of the environment can affect the nature of one's lived experience in deep and far-reaching ways. Consider someone who, as a fully grown adult, is under four feet tall. Given the "standard" sizing of everything from cabinet height to airplane seats to steps, she will stand in relations to her environment that will have a negative effect on her lived experience.[41] This brings out an especially clear way in which the subjective value of an outcome for a person depends on much more than her internal, intrinsic properties. Subjective values concern lived experience of outcomes, or *what it's like to live in this*, understood as an assessment of what it will be like for a person (or for others affected by her decision) to live in the circumstances of some possible outcome.[42] It can be a matter of social justice to improve

[41] For related points see Barnes Forthcoming.
[42] Also see Campbell 2015 and Paul 2015a.

the subjective values of lived experience, and by extension, to construct conditions that facilitate certain types of transformative experiences.

Second, the connection to social conditions and social justice ties questions about transformative experience and decision-making to interesting issues in law and public policy.

In many legal, behavioral economic, and social policy contexts, value must be assessed in terms of monetary costs and benefits. Serious problems arise when the values of some outcomes cannot be quantified, that is, when these outcomes cannot be assigned a value.[43] If there is no way to assess the utility of those outcomes, then there is no way to calculate an adequate monetary representation of the value (or a monetary quasi-equivalent of that value) that can be used for the relevant cost/benefit analysis. But as a practical matter of fact, in policy contexts, monetary representations of the outcomes are essential in order to be able to assess, manage, understand, and regulate the scenarios they concern.

Examples of nonquantifiability in contexts involving radical transformation abound. Consider values for lived experience outcomes concerning veterans returning from war trauma, or for victims of terrorism, financial collapse, or significant personal injury, or for individuals making decisions in medical and legal cases. If a subjective value cannot be assigned to the outcome in any straightforward way, unless we are to pretend the problem doesn't exist (and thus ignore the needs of citizens we have a responsibility towards) we need some way to construct a proxy value. This is a matter of social justice.

Recent work by Cass Sunstein is intended to address this problem. He identifies the problem and develops the foundations for what he describes as "breakeven" analysis.[44] Breakeven analysis is designed to show how to partially manage decision models for cases where the subjective values of the outcomes cannot be determined.

> The problem of nonquantifiability is a recurrent one in both public policy and ordinary life. Much of the time, we cannot quantify the benefits of potential courses of action, or the costs, or both, and we must nonetheless decided whether and how to proceed. Under existing Executive Orders, agencies are generally required to quantify both benefits and costs, and (to the extent permitted by law) to show that the former justify the latter. But agencies are also permitted to consider apparently nonquantifiable factors, such as human dignity and fairness, and also to consider factors that are not quantifiable because of the limits of

[43] I am indebted to Cass Sunstein for discussion of these ideas.

[44] A related paper (Vermeule 2013) discusses how problems like this can arise in legal contexts. The paper discusses the way judges and regulatory bodies lack appropriate models for recognizing and regulating the problem, and argues that a more rational decision process would recognize the knowledge gaps.

> existing knowledge. When quantification is impossible, agencies should engage in "breakeven analysis," by which they explore how high the nonquantifiable benefits would have to be in order for the benefits to justify the costs. (Sunstein 2014, 1369)

Social and legal theory needs to explicitly recognize the possibility of gaps in value assignments for these types of outcomes in order to adequately diagnose problems, correctly interpret cases, and design models that can manage the unknowns as effectively as possible. Extending Barnes's point about social justice, recognizing the way that transformative experience can be the source of nonquantifiability and understanding the structure of the epistemic and personal change involved in transformative experience is essential for the proper identification and management of these social conditions, and seems to be part of what's needed for a just society.

Finally, the importance of properties of the environment to the nature of an individual's transformative experience highlights the complexity of using empirical data to predict how an individual will respond to a given experience, for the prediction must take the individual's environment into account. How an individual responds to a transformative experience may depend as much on her environment, and on the particular combination of her physical properties with her micro-environment, as it does on her intrinsic psychological states.

The importance of environment and context to our ability to make accurate predictions for an individual must not be underestimated. This issue arises even when our data is drawn from research that meets the highest standard, for example, when we are doing evidence-based medicine, where many predictions are made on the basis of randomized clinical trials. In a real-life environment the accuracy of a prediction for an individual can be significantly affected by the shift from controlled experimental contexts to messy, real-life situations. Well tested, highly verified models that work beautifully in controlled settings can crash dramatically in real-life contexts, often because the properties of a particular environment affect individual responses in ways the models are not able to predict (Cartwright 2011).

Rachel McKinnon (2015) argues that the decision to change one's gender is a transformative decision, since transitioning from one gender to another bears all the hallmarks of other types of transformative experiences. I agree that transitioning from one gender to another is a transformative experience, and that the decision to transition is a transformative decision. If so, then before you transition, you cannot know what it will be like to have your new gender, and so you cannot assign a value to your future lived experience with that gender. Moreover, you are constructing a new self with your choice, which will change your core personal preferences. As a result, you cannot rationally choose to transition, if your choice is based on what it will be like for you to have transitioned.

Does this imply that we cannot make such decisions rationally? No. As McKinnon argues, you might find the status quo intolerable. That is, you might find life with your current gender to have a high negative subjective value, and as a result you place a high value on change. As I'd describe the situation, you disprefer the status quo, and as a result you prefer to change your preferences.

McKinnon argues that we need a model for rationally choosing to transition.[45] I agree. In Paul 2014, I argue that one way to make a rational choice to have a transformative experience is to prefer to discover the new preferences you'll form as the result of having that experience. I describe this as choosing revelation, that is, through action, you choose to reveal to yourself who you become. This, then, is one model for rational, transformative choice: just as you might rationally choose to have a child based on the preference to discover a new self, yourself as a parent, you might rationally choose to transition based on the preference to discover yourself with a new gender.

McKinnon also shows how epistemically transformative experience connects with the literature on feminist standpoint epistemology and situated knowledge. Understanding the nature of epistemically transformative experience connects to issues concerning epistemic trust and epistemic humility in work on oppression and intersectional identity, with further applications in political theory.[46]

Ryan Kemp (2015) explores the rationality of radical personal transformation in the context of debates over the rationality of moral norms and moral self-transformation. As he points out, these debates face a version of the problem with transformative decision-making: how can a person rationally choose to transform her current self into a radically new self?

I find Kemp's discussion interesting and provocative, and was particularly interested in the connections he draws between the work of Sartre and Kierkegaard and contemporary philosophical issues concerning self-transformation. It may well be true that radical self transformation derives from contingent facts and that it involves a leap of faith. As Barnes (2015a) points out, such transformation may also depend on properties of the social environment, and as McKinnon (2015) argues, the individual may feel that she has no choice but to undergo transformation.

I was unconvinced, however, by Kemp's argument that radical self-transformation cannot be covered by a choice based on revelation, that is, by a choice to transform oneself based on one's preference to discover a new self. Kemp argues that cases of radical self-transformation are cases where

[45] While I don't provide a model in Paul 2015b, I do develop a model in Paul 2014. The dates notwithstanding, Paul 2014 was not published when McKinnon's article was accepted for publication while Paul 2015b was widely available online from January 2013. Thus McKinnon is justified in not engaging with the (2014) book.

[46] For discussion of the problems that transformative experience raises for accounts of democratic ideals that rely on cognitive empathy, see Stanley 2015, 102–105.

> a person explicitly decides to make a change with the precise intent of uprooting her central preferences. To put the point a bit more colorfully, transformative experiences involve a decision to risk normative death in order to experience something new; radical self-transformation involves a decision to embrace normative death at the outset. (Kemp 2015, 395)

If I understand Kemp correctly, he wants to distinguish between transformative decisions where a person makes a decision under conditions of uncertainty as to whether a consequence of her act is self-replacement, and transformative decisions where a person is certain that a consequence of her act is self-replacement. This is an interesting distinction that may well have normative implications in the moral domain. However, one can rationally choose self-replacement based on a preference to replace one's self with a new self in both types of conditions. If the choice to φ is transformative, one can rationally choose to φ if one's choice is made based on the higher-order preference to discover what it's like to become a new self. Such an act meets the normative standard for rationality even if it creates problems for other types of norms.

Muhammad Velji (2015) discusses the question of accommodation for those with religious preferences, such as accommodation of those who prefer to veil or accommodation of those who prefer to keep kosher. He argues that we should not refuse to accommodate religious preferences on the grounds that such preferences result from the choice to become religious. His interesting argument turns on the sort of epistemic and personal transformation involved in the slow process of religious transformation through pious enagagement and practice.

Velji argues that we should not see the choice to train oneself in religious piety as a fully informed choice, for the self-development involved in the transformation of oneself into a pious believer can change a person into a new self with new preferences, but the preferences of this future self are epistemically inaccessible before her religious transformation is undertaken. If the prospectively religious believer cannot know whether she will require religious accommodation, such as accommodation for veiling, until she achieves a certain level of piety, and once she reaches that level of piety, veiling is part of her religious identity, then her choice to believe is not informed in a way that undermines her right to religious accommodation.

The central idea is that, while the choice to be pious is indeed a choice, the choice to pursue religious piety should not be regarded as a choice analogous to choosing a particular lifestyle such as choosing to drive a fancy car or choosing to develop one's physical prowess through skiing. Rather, it's a choice that can transform a person. If so, veiling and other religious practices should be accommodated, for they are practices that are constitutive of one's self-identity, not mere lifestyle choices, and the

nature of one's religious self-identity (especially, the nature and extent of the accommodation needed) is only discovered after the path to piety has been embarked upon.

6.3 Subjective Value and Happiness

In his engaging paper, Antti Kauppinen (2015) argues that (i) "nonexperiential" values are far more important for our big life choices than "experiential" values, and (ii) choice based on valuing revelation is not normatively acceptable.

Kauppinen's primary target is my notion of subjective value. As I define subjective value, it is the value of lived experience. This fact seems to have been misunderstood by some of my critics, so I welcome the opportunity to clarify the idea. In particular, Kauppinen, along with some of the other ethics-oriented papers in this volume, such as those by Chang, Dougherty, Horowitz, and Sliwa, and Harman, take subjective value to be something it is not. Kauppinen thinks it is the value of mere subjective feel, and rejects it as a suitable value for making life choices. But lived experience is much more than mere subjective feel.[47]

While the idea that we value lived experience is very natural and intuitive, it *is* unfamiliar from the perspective of contemporary practical ethics, especially since much of "analytic" philosophy, with the exception of the philosophy of mind, has traditionally regarded phenomenology, and talk of experience more generally, with fear and trembling. Moreover, my arguments for the necessity of experience in generating the epistemic capacity to imagine new types of future lived experiences exploit classic examples from the philosophy of mind, but those examples were traditionally used in arguments concerning mere qualitative feel. So it might seem that I am arguing that big life choices should be made based merely on what sort of phenomenology they create, and not on more important bases, such as the objective value of those outcomes. This misrepresents my project in a significant and far-reaching way.

Clarification is in order. As I'll explain below, my argument is that experience is necessary for our ability to represent and imagine the nature and character of outcomes that involve much more than mere subjective feel.[48] Such outcomes include those involving love, betrayal, fear, friendship, loyalty, personal sacrifice, etc., or what I describe as "rich, complex, experience-involving" states of the world, and are the sorts of outcomes that Kauppinen and others are very rightly concerned with.

Let's start with whether the concern is merely about our inability to imagine mere subjective feels. As John Campbell (2015) points out:

[47] See Campbell 2015 and Paul 2015a for further discussion.

[48] Barring special machines that can create brain states that would duplicate the brain states we'd get from having the experience, etc.

> Although current philosophy of mind has long recognized the importance of imaginative understanding, it's been given a remarkably restricted role: providing one with knowledge of the qualia, the purely internal characteristics of someone's mental life.

But there is no need to restrict imaginative understanding to such a confined role, and my arguments do not do so.[49] Instead, I emphasize the central importance of imaginative representation in prospective planning, anticipation, and decision-making, and show how enlarging the role for imagination enlarges both the interest and the scope of transformative decision-making.

That is, in many situations, including situations where we are struggling to assess possible life choices, we want to be able to imaginatively represent new types of outcomes in which we experience new events and live in the world in new ways, including ways we'd experience ourselves in such outcomes. My argument is that well-known arguments from the philosophy of mind about the need for experience to represent possible qualitative states should be extended to our ability to represent these new outcomes. In particular, the role of experience is just as important in imaginatively representing the nature of possible states in which we experience and live in the world in new ways as it is when imaginatively representing fairly simple qualitative states. And, importantly, possible states in which we experience new events and new ways of living in the world include some of the rich, complex, experience-involving states of the world that concern us most when we make big life decisions.

Why is experience necessary for a grasp on the nature of these rich, complex, experience-involving states? As follows: To grasp the (salient) nature of a complex, many-featured state of the world, one must grasp the nature of the (salient) features that compose or constitute this state.[50] In the complex states of interest, such as outcomes that involve love, betrayal, friendship, aesthetic beauty, sacrifice, etc., some of the salient features that partly compose or constitute the states are fundamentally experiential: they are sensory, such as what it's like to feel pain, or they concern internal experience, such as what it's like to feel certain basic emotions or what it's like to hear beautiful music, or they include our experience of our more complex physiological and psychological responses to various events or properties in the world.

Other salient features that may partly compose or constitute the states are not primarily experiential, but our ability to grasp their natures depends on our experience, such as our ability to grasp what it's like for another

[49] For further discussion see Paul 2015a.

[50] This is a necessary condition. Sometimes we can have a grasp on the natures of the parts, yet the nature of the whole continues to elude us. Also, I'm using the term "salient" to gloss the possibility that there might be features that make a largely irrelevant contribution to the nature of the state. These are not features we need to be concerned with.

person to feel pain or to see color. (For example, we have to know what it's like to feel pain to empathize with the pain of another.)

Of course, there may be other features that partly compose or constitute these complex states that are not experiential, nor is our grasp of their natures dependent upon our grasp of something experiential. We can call such features *wholly non-experiential*. But unless the salient nature of a state is wholly composed of wholly non-experiential features, to fully grasp its salient nature, we will need experience.

And finally, the states we care about when we make big decisions, such as states of the world that concern the fear, anger, friendship, love, aesthetic beauty, sacrifice, etc., of ourselves and others, are not wholly composed of wholly non-experiential features. They are not "experience-free" states, and their salient natures are grounded, at least partly, by the natures of their experiential features. Love and friendship, for example, involve experiences and feelings towards others and responses to others, as does the attachment between parent and child. Most of the complex states involving new events and ways of living in the world that we want to assess in life-changing situations are constructed from all three types of features, and in most, experience is highly salient.

Now we can see the role of the argument from experience: the rich, complex, and meaningful states of the world that concern us in many big life decisions are not experience-free states, they are experience-involving states. In order to imaginatively represent the nature of these experience-involving states, it is necessary for a person to have had the right type of experience. Without having had the right sort of experience she cannot represent the nature of some of the most important features that compose the state, and thus she cannot represent the (salient) nature of the state itself.

So the idea is that experience of the relevant kind is needed for one to have the epistemic capacity to grasp the nature of the experience-involving states involved in big life changes, for it gives us the imaginative capacity to first-personally represent or model our possible future selves (and the possible future selves of others) in those states. And note that subjective value is intended to be understood *de re*, in the sense that they concern what it's like (for oneself and for others) to live in these possible circumstances, or as I sometimes describe it, "what it's like to live an outcome" in the world.

As I indicated above, I connect the nature of these experience-involving states with subjective value, describing it as "the value of lived experience" or the value of what it's like to live in possible outcomes. The idea is that only by grasping the nature of experience-involving states do we have the capacity to represent and assign them subjective value.

This is why arguments against subjective value as grounded solely by mere subjective feel miss the mark. Arguments based on "experience machine" worries that suggest that experience isn't important are also misguided. My argument is, first, that subjective values of the important types

of lived experiences, understood *de re*, are among the values that matter to us, especially when making life-defining transformative choices. Second, that experience is *necessary* to grasp these important subjective values. I am not arguing that experience is sufficient to grasp these subjective values, nor that some sort of thin, purely internal, affectless phenomenal character is sufficient to ground subjective value.

You might think that objective value is ultimately what we should care about. I do not object to this, given that, for experience-involving states, their objective value often depends in substantive ways on their subjective value. (See my discussion of Ruth Chang's paper in section 6.4.) One might also wish to assess the intrinsic value of these states, as Harman (2015) and Dougherty et al. (2015) do. Again, I have no objection to this, as long as we are clear that, somehow, when assigning objective values to outcomes, we will need to assign them subjective values. This is because in transformative decision contexts, subjective value is of central importance to the decision maker.

Kauppinen's discussion might seem somewhat orthogonal to all this. He says that he is concerned to assign values that he describes as "non-experiential." But Kauppinen also wants to assign values to states that include "integrity, commitment, friendship, meaning, or achievement," just as I do. Where is the significant difference? Mostly, I suspect, in how we are drawing distinctions. In particular, Kauppinen argues that experiential value is merely (prudential) hedonic value, and as a result thinks it is of little importance. But since my experience-based subjective values are not merely hedonic values, I am inclined to value many of the same features of the world as he is.

For these reasons, I agree with Kauppinen that prudential goods like achievement, friendship, and self-respect are the sorts of things we should assign value to when we assess the expected value of acts involved in major life decisions. It's just that valuing these goods for transformative decision-making involves assigning them subjective value, and that's the type of value we need to focus on in this context.

Attempting to value outcomes involving friendship, love, and self-respect without including lived experiences in these states leads to a weird sort of zombification of what we are supposed to value.[51] The zombie equivalent of love is all right, I suppose, but it's not the kind of thing that's suitable for prudential value. We want to value real love, and real love involves conscious experiences and emotional relationships and many other sorts of things that subjective value is designed to capture. (Again, this does not mean that love is merely experiential or is merely a subjective feel.)

In correspondence, Kauppinen grants that achievement, friendship, and the like necessarily involve experiences, but maintains that an important

[51] Mark Johnston (2006) makes related criticisms of what he terms the "Wallpaper View," where sensation is treated as a mere add-on to perceptual judgment.

part of their value is grounded in their non-experiential features. On this much, we agree. Where we disagree concerns what we have to know or be able to represent. On my view, the subjective values of lived experience outcomes are among the values that matter most in these decision contexts, and first-personally grasping such values requires experience-based knowledge, or at least, an experience-generated epistemic capacity. On Kauppinen's view, experience-based knowledge or capacities are not required for grasping the values of lived experience outcomes, that is, he thinks experience is not needed to first-personally grasp the aspects of lived experience outcomes that are relevant to their value. What our debate brings out is a deep difference in our views about the epistemic and conceptual requirements for assigning value to outcomes involving lived experience, and raises further interesting questions about how to understand the metaphysical structure of lived experiences and the grounding of the relevant values.

Kauppinen also criticizes the adoption of revelation, or discovery, as a subjective value. I see his concern, and indeed, I do not think that choosing based on revelation alone is an especially successful way to resolve the problems with transformative decision-making. Simply choosing a particular life path to discover what that path will be like is not at all how we normally want to proceed, especially given the modern focus on planning one's future and thinking in terms of narrative goals. So I agree with Kauppinen that it can be normatively unacceptable to choose a life path for revelatory reasons alone. But the norm that's violated is not an epistemic or a rational norm, at least, not as long as you are choosing consistent with your rational preferences. Rather, it's a norm of authenticity, because we normally prefer to know the nature of the life path, at least to some significant degree, that we are choosing. And the main problem with transformative decision-making, as I've been at pains to point out, is that the norms of authenticity conflict with the norms of rationality, and there is no easy or obvious way to resolve this conflict.

6.4 Decision-making in Contemporary Ethics

Elizabeth Harman (2015) explores the relationship of transformative choice to her groundbreaking work on "I'll be glad I did it" reasoning. She discusses two cases that we both find to be of great interest, the case of a woman choosing to have a child and the case of a parent choosing to implant their deaf child with a cochlear implant. Harman argues that experience is not necessary for people to rationally make decisions in cases like these, and argues that reliance on reliable testimony, as opposed to reliance on faulty "I'll be glad I did it" reasoning, will allow us to make rational decisions in cases of transformative choice.

In reply, I argue that experience of the right sort is necessary if we choose based on our assessments of the expected subjective value of becoming a parent. I will discuss the structure of "I'll be glad I did it" reasoning and

show how the structure of transformative decision-making and the values it concerns are importantly different from those involved in faulty "I'll be glad I did it" reasoning. (Exposing the difference between these structures is also important for my discussion of Howard 2015, below.)

After I distinguish the structure of transformative decision-making from "I'll be glad I did it" reasoning, I'll explain why testimony can't provide the relevant information to the agent who lacks a subjective value function. I'll close by exposing how an ambiguity in Harman's claim that it is "better" for a deaf child to have a cochlear implant illustrates the way that deep questions concerning this sort of real-life choice concern epistemic and personal transformation and the value of subjective lived experience, not merely testimony and faulty "I'll be glad I did it" reasoning.

To set the stage for her defense of the view that people can rationally make transformative choices by relying on testimony, Harman argues that becoming a parent is not epistemically transformative. Her argument is that (i) she had parenting-like experiences with her much younger sister and the child of a close friend, so she knew what it would be like to become a parent before she had a child, and (ii) she knew before she became a parent that she would experience joy and other sorts of emotions when she had her child.

Harman's personal experience does not refute the argument that one has to stand in a parent-child attachment relation to a child to know what that distinctive type of experience (parenting experience) is like. Why? Because, by her own admission, Harman had the experience of being a parent before she physically produced her own child, through experiences with her younger sister and with her friend's baby. She notes that she had feelings for her sister that were "parental in their nature" and when describing her feelings for her friend's baby says "I experienced my own love for her baby, which was unlike any feelings I had ever had (as an adult) for a baby" (326). If we take this description of her experiences at face value, Harman *alloparented* her friend's child (and her younger sister): that is, she loved and cared for children that were not her own.

Alloparents are individuals who are not the biological parents of a child, but who stand in a parent-type relation to that child. So Harman had experience with being a parent, just not of being a parent of a child she'd physically created. But my argument does not require that one must be a biological parent to know what it's like to be a parent. My argument is just that the type of loving attachment to a child that one experiences as a parent is a distinctive experiential kind. If you've had experiences that are instances of that kind, you can know what it's like. Just like you can know what it's like to taste Vegemite once you've tasted Marmite (both are yeast extract spreads),[52] if you parent other children before you parent your own, you can know what it's like to be a parent.

[52] Well, many Australians and English would deny that Marmite and Vegemite taste anything like the same. And New Zealanders have their own kind of Marmite. I get it. But bear with me for the example: most Americans can't taste the difference.

While my argument is philosophical, these questions are being explored in the psychological literature as well. Preliminary empirical results from the work of Josiah Nunziato and Fiery Cushman indicate that many people report having transformative experiences when they experience significant life events such as becoming a parent. In particular, people report that the experience changed them in ways that they could not foresee. In addition, further results suggest that people are over-confident about their ability to anticipate the changes that will occur in their beliefs, preferences, desires, and values as a result of a transformative experience.[53]

There is another problem with Harman's inference from her own case. The experience of being a parent is a very broad type of experience, one that every parent who forms an attachment to her child has acquaintance with. But within the parental experience-type, there may be further distinctions to be made, for there are many distinctive types of parenting experiences: parenting an extremely gifted child, parenting a severely physically disabled child, parenting a terminally ill child, parenting a mentally ill child, etc. Each of these subtypes of parenting experience may be distinctive and different enough from the others to require experience of that particular subtype in order to properly assess their subjective value. (This is a reason why having a second child can be transformative.)

Given her description, Harman was *lucky* with the type of parenting she had experience with: that is, her account suggests that her child was relevantly similar to the children she alloparented, and that she responded to her child in a way that was relevantly similar to the way she'd responded to the other children. But if something had been different, for example, if the child she physically produced had been distinctively different from the other children she'd alloparented (e.g., by being terminally ill), her alloparenting experience might not have been similar enough in the relevant way for her to grasp a significant part of the subjective value of that outcome. This brings out a mistake in her description of an argument she attributes to me.[54] My argument for the epistemically transformative nature of parenting experience is not based on the claim "that one cannot know ahead of time, regarding any specific possible experience of pregnancy and parenthood that one may have, what *that experience* would be like." Rather, the argument is that one needs experience of the relevant *type* of parenthood experience in order to be able to represent experiences of that type in order to grasp their subjective values.

The second problem with Harman's argument against parenting as an epistemically transformative experience concerns the distinction between knowledge-that and knowledge derived from experience.[55] There is no

[53] I'm indebted to discussion with Fiery Cushman here.

[54] Harman constructs her own version of the argument, so I am focusing on (and quoting) the version of the view she describes. The arguments I actually develop in Paul 2014 and Paul 2015b are not quite the same.

[55] For a related objection to my view, see Krishnamurthy 2015.

question that there is a lot of knowledge that you can have before you decide to have a child, such as knowledge that it will probably be tiring, knowledge that you are likely to love your child, and knowledge that you could have less time for favorite hobbies. This was never in dispute. But Harman seems to think knowing-that it is likely that one will have these experiences, via moral testimony and via observing others, is enough to know what it's like to be a parent. In particular, she seems to think knowing that one is likely to feel the types of emotions that many associate with parenthood is enough to assign *being a parent* subjective value. "There is definitely a kind of joy I had never experienced until now. But I knew there would be" (326).

However, knowing *that* you will (probably) feel new emotions isn't knowing what it's like to feel them, just like knowing that you'll see red isn't knowing what it's like to see red. Her arguments for the role of testimony in gaining knowledge suggests that, like Dougherty et al. (2015), Harman is implicitly replacing assessments of subjective value with other types of value, such as intrinsic value, which is not the sort of value my arguments for choosing to become a parent concern. So her claims about the knowledge that we can have before a transformative experience do not constitute an argument against my view that people cannot rationally choose to become parents based on their assessment of the expected subjective value of that act. Moreover, she makes general, problematic assumptions about what we can know about our own case from knowing anecdotes and statistics about other people.[56] I will return to the question about the types of value judgments we are making in my discussion of cochlear implants, below.

In the second section of her paper, Harman discusses reasoning involving decision-making.[57] One part of her discussion concerns a distinctive kind of reasoning, involving "I'll be glad I did it" considerations, where a person reasons from the fact that she is attached to the outcome she actually chose to the claim that she should have chosen that outcome. I agree with Harman that this sort of reasoning is faulty.

But there is a second part of her discussion in section two that deserves a more critical assessment: her discussion of transformative decision-making, where a person makes a decision to undergo an experience that transforms her core personal preferences. Transformative decisions include decisions to become a parent and decisions to radically change one's sensory capacities, such as the decision to get a cochlear implant (if one is congenitally deaf). The cochlear implant case is complicated by the fact that, usually, parents must make the transformative decision for their very young child. In

[56] See my reply to Dougherty, Horowitz and Sliwa for further discussion about intrinsic versus subjective value. For a case study illustrating the dangers with naïve interpretations of empirical results, see my reply to Sharadin 2015. Finally, for detailed discussion about problematic inferences from anecdotes and statistical results to one's own case see Pettigrew 2015, Paul 2015a, Moss Unpublished, and section 3–section 5 above.

[57] Thanks are due to Matt Kotzen for very helpful discussion.

Paul 2014, I argue that a central problem for such decisions concerns the incommensurability of the individual's preferences before and after the decision, and the epistemic inaccessibility of the first-personal perspective of the future, transformed self from the current, untransformed self's first-personal perspective. (I discuss the general structure of this problem in section 5, above.)

Harman claims that we can rely on the testimony of satisfied parents or other "experts" as a guide for transformative decision-making. However, as I argue in reply to Dougherty et al. (2015) and others, and as I discuss in section 5, above, the agent cannot use such testimony to know whether she is making a decision that will satisfy her current preferences, or whether it simply satisfies the future preferences of a future, alien self.

Is, as Harman's discussion suggests, the real source of the problem with using testimony merely that agents can be susceptible to the distorted sort of reasoning based on attachment we see in some of the "I'll be glad I did it" examples? Is it correct to think that, once we set this sort of reasoning aside, we can use testimony to resolve cases of transformative choice, such as choosing to become a parent or choosing a cochlear implant?

No. The faulty "I'll be glad I did it" reasoning Harman discusses is not what is creating the deep problem for rational choice in cases of transformative decision-making. What's going on in cases of transformative decision-making (which include real-life cases of transformative decision-making, such as the choice to receive a cochlear implant) is a different problem.

I'll show this by comparing the structure of Harman's cases to the structure of cases of transformative decision-making that create the deep problem for rational choice.

Harman's first case involves an exam. In this case, at *t1*, you decide to skip a film and study for your exam instead. While you might enjoy seeing the film, you have a higher-order preference to skip the film and study instead, a preference to do well on the exam rather than see the film. At *t2*, you are glad you studied at *t1*. You were right to study for the exam at *t1*, because your preferences, including your higher-order preferences, are consistent across *t1* and *t2*.

Harman's second case involves a woman who had a baby as a teenager. In this case, at *t1*, she chooses to have the baby. At *t2*, the woman is glad she has the baby. Harman says:

> A woman who became a parent as a teen might say, truly, "I should not have had a child as a teen. But I love my son and I'm so glad I did, because otherwise I wouldn't have had him. That I love him and am glad to have had him—that I would not wish to change anything for myself—in no way makes me think that teen parenthood is a good choice for anyone to make." (2015, 335–336)

In Harman's second case, the woman's first-order preferences conflict with her higher-order preferences. At *t1*, she chooses to have the baby. At *t2*, she prefers to have the baby because she has formed a reasonable attachment to him. But her higher-order preference to not be a teenaged mother remains consistent across *t1* and *t2*. (We could also tell a story where her first-order preferences changed from *t1* to *t2*, perhaps because she had the baby against her will. At *t1* she did not prefer to have the baby, though at *t2* she prefers to have the baby. Still, her higher preference is consistent across *t1* and *t2*.) Because her higher-order preference to not be a teenaged mother remains, it is still correct, at least with respect to this higher-order preference, that she should not have had the baby at *t1*.

What's the structure of transformative decisions such as the choice to give one's congenitally deaf infant a cochlear implant or the choice of a mature adult to have a child? These cases of transformative decisions are importantly different from the exam case and the case of the teenaged mother. In these cases, the agents' preferences, including their high-order preferences, are transformed as the result of the choice. Cases involving transformative decisions involve incommensurable preference changes across time, not higher-order preference change accompanied by cross-temporally consistent higher-order preferences.

Let's consider the choice of an adult, Anne, to have a child. At *t1*, she does not prefer to have a child. Moreover, all of her higher-order preferences are consistent with this: she prefers not to have a child relative to her other preferences, she prefers to prefer not having a child, etc. Yet, the expert tells her that at *t2*, if she has a baby, she'll have preferences that are satisfied then.

Harman is correct to argue that a bad way for Anne to reason about the expert's testimony is to think that the fact that she'll "be glad" to have the baby at *t2* is a reason she should decide to have a baby at *t1*. For if Anne has a higher-order preference to remain childless that remains consistent across *t1* and *t2*, her choice violates this preference, despite a higher-order preference to have the child that is created at *t2* by the existence of the baby.

But Harman suggests that, once we set aside faulty "I'll be glad I did it" reasoning, we can simply accept expert testimony as a guide to action in transformative decision-making.[58] This is a mistake. Faulty "I'll be glad I did it" reasoning is not what creates the deep problem for transformative decision-making. This is because the source of the problem with faulty "I'll be glad I did it" reasoning, as Harman describes it, involves a distortion due to attachment, with a consequent first order preference change.

[58] Note that this could require a person to replace her introspective assessments with expert testimony. I discussed problems with this strategy in section 3 and section 4 above. Deciding based solely on expert testimony about one's future preferences would also conflict with the rigidification strategy discussed by Briggs 2015.

Such distortion can occur even when the agent has consistent higher-order preferences across the change, as in the teenaged mother case.

The deep problem for transformative decision-making is different: it comes from transformations that create higher-order preferences that are *inconsistent* with those of the agent prior to the change, and a transformed first-personal perspective that is inaccessible to the agent prior to the change. That is, it involves incommensurable core personal preferences, including higher-order preferences, across the transformative epistemic change of the agent. Testimony won't allow us to evade this problem. (See my discussion of Howard 2015 for a related point.)

Go back to Anne, who is choosing whether to have a child. At *t1*, she does not prefer to have a child. Moreover, all of her higher-order preferences are consistent with this first order preference: she prefers not to have a child relative to her other preferences, she prefers to prefer not having a child, etc.

How should Anne regard the expert's testimony? The trouble is that, in the transformative case, the radical transformation she'll undergo when becoming a parent means that her preferences at *t2*, including her higher-order preferences, will be incommensurable with her preferences at *t1*. So if Anne decides to have the baby, she violates all of her preferences at *t1*. Moreover, her transformed first-personal perspective at *t2* is inaccessible to her at *t1*. And so the expert testimony that her preferences will be satisfied at *t2* cannot guide her rational action at *t1*. (A related way to see the difficulty is to explain that Anne cannot assess the relevance of the expert advice to her current self, because the expert advice merely concerns the preferences of her possibly transformed self, a self that Anne at *t1* regards as psychologically alien.) Adding insult to injury, if Anne decides not have the baby, she violates all of the preferences she'd have had at *t2* if she had had the baby.

The problem with incommensurable, inaccessible future perspectives also infects the structure of cases involving cochlear implants. Such cases are best thought of as forced choices, because parents must decide whether to implant their child when the child is very young. The comparison of interest is between very different, distant future outcomes: an outcome with a (significantly older) Deaf child, and an outcome with a (significantly older) child who can hear in a species-typical way.

Consider a parent who must decide whether to implant her congenitally deaf infant. Let us assume, first, that the parent does not have a higher-order preference to have a child who can hear, nor does she have a higher-order preference to have a Deaf child.

Assume that the parent chooses to reject the implant at *t1*. At *t2*, the Deaf child is glad to be Deaf. Is it correct to say that the defense of the decision to reject the implant at *t1* must involve faulty "I'll be glad I did it" reasoning? No. Unlike the teenaged mother, this parent does not have a

higher-order preference across *t1* and at *t2* conflicting with her higher-order preference to reject the implant at *t1*.

Now complicate the case. Perhaps, at *t1*, an expert tells the parent that, if she implants her child at *t1*, that at *t2*, the child will be glad to have the implant. But if the parent refuses the implant at *t1*, at *t2*, the child will be glad to be Deaf. What's the right assessment for the parent to make?

The different preferences at *t2* stem from the fact that receiving a cochlear implant is transformative for the child. If he receives the implant, at *t2*, he'll have one set of preferences, P_H, formed in part by his life as a person who can hear. If he does not receive the implant, at *t2*, he'll have a different set of preferences, P_D, formed in part by his life as a Deaf person. So if the child is implanted, at *t2*, the parent and the child have preferences (P_H), including higher-order preferences, that the child is implanted. If the child is not implanted, at *t2*, the parent and the child have preferences (P_D), including higher-order preferences, that the child is Deaf. Preferences P_H are incommensurable with preferences P_D.

In this case, if there is no preference had by the parent at *t1* prohibiting implantation, choosing to implant is rationally permissible. Likewise, if there is no preference had by the parent at *t1* prohibiting the refusal to implant, refusing to implant is rationally permissible.

We can complicate the case further. A parent might have preferences, including a higher-order preference, to refuse the implant. The expert might tell the parent that, were the child to be implanted, at *t2* the parent and the child would have preferences P_H, including higher-order preferences, that the child is implanted. Does this imply that the parent should choose to implant? No. If the experience of becoming a person who can hear (and of being the parent of that person) is transformative, then the parent's preferences at *t1* are simply incommensurable with her preferences at *t2*. And as a result of the radical preference change that could occur, the parent at *t1* can regard the self she'd be at *t2* as alien to her current self at *t1*. (And she might regard who the child would become at *t2* as alien to who the child is now, at *t1*.) So testimony will not evade the problem for rational decision-making for a parent faced by this sort of transformative choice.

Is it nevertheless "better," as Harman suggests, for the child to receive the implant than to forgo it? In what sense could it be better? One way it could be better is that the parent's (and child's) preferences could be better satisfied in one outcome than they are in another. But as we have seen, preferences P_H are satisfied in the outcome where the child is implanted, and preferences P_D are satisfied in the outcome where the child is not implanted. Moreover, preferences P_H are incommensurable with preferences P_D. So the sense of "better" with regard to better satisfying one's preferences does not apply.[59] Is there another sense of "better" in play here? In particular, is Harman suggesting that being able to hear in a species-typical way is

[59] For related discussion see Barnes Forthcoming and Howard 2015.

somehow *intrinsically more valuable* than being Deaf? Is this what is supposed to solve the problem for the parent faced with the transformative decision to implant her deaf infant? Her defense of the role of testimony, and her approving citation to Dougherty et al. 2015, suggest this possibility, but she does not explicitly endorse it in this paper.

I categorically reject the thesis that being able to hear in a species-typical way is somehow intrinsically more valuable than being Deaf. If Harman is arguing that, even in a world where cochlear implant technology is perfect, parents should give their deaf children cochlear implants because being able to hear is intrinsically more valuable than being Deaf, we need an explicit and compelling argument for that view.[60] And once we move to cases in the actual world, what deserves further scrutiny are the possibilities that (a) being Deaf is extrinsically less valuable than being able to hear in a species-typical way, given the way that many societies are organized, and (b) being Deaf is extrinsically more valuable than being implanted, given technological and other facts about cochlear implants.[61] Such questions about extrinsic value (and the role of the subjective value of lived experience in contributing to extrinsic value) are difficult but highly salient in this context.[62]

In sum: the deep questions surrounding big life decisions and the transformative choices they involve concern epistemic and personal transformation and the subjective value of lived experience. Harman's arguments that we can make these decisions rationally simply by relying on testimony about the intrinsic value of outcomes conflates subjective value with intrinsic value, misdiagnoses the source of the problems, and fails to recognize the deep problems with preference change and alienation when relying on expert testimony to make life-changing transformative decisions.

Dana Howard (2015) discusses the relationship between "I'll be glad I did it" reasoning and transformative decision-making. She argues, contra

[60] See Harman 2009 for related discussion.

[61] See Barnes 2015b for related points about transformative experience and social conditions.

[62] In Paul 2014, I argue that parents without the experience of being Deaf cannot assess the subjective value for their child of being Deaf. I also argue that parents without the experience of species-typical hearing cannot assess the subjective value for their child of being able to hear in a species-typical way. I then argue that, as a result, we should not expect parents to be able to have rationally defensible preferences concerning these subjective values when they make the decision whether to implant. Moreover, in real-life cases, especially given the limitations of current technology, a pressing and important concern for parents with a deaf child involves their future ability to fully communicate and engage with their child, which brings with it the fear of alienation from one's own child. Often, parents prefer to keep their child in the same community as they are, in hopes of maximizing the child's chances of a happy and successful childhood and subsequent preparation for adult life. Thus, Deaf parents may refuse the implant, judging that their child is best off as a member of the Deaf community, and hearing parents may choose to implant, judging that their child is best off as a member of the hearing community. Consideration of facts like these could be part of what leads Deaf parents to form a higher-order preference to have a Deaf child, and for hearing parents to form a higher-order preference to have a hearing child.

Harman (2009), that "the fact that one will be glad one did it never offers up a conclusive reason to believe that one should do it" (358).

In her argument, Howard distinguishes between preference change based on adaptive preference formation, understood here as stemming from a diminished set of life options, and preference change based on other sorts of reasoning. She argues, rightly, that when a disabled person has a preference to be disabled, that we should not assume that that the disabled person's preferences stem from adaptive reasoning. To show that adaptive reasoning is the source of the preference we must first show that the life options of the disabled person are in fact diminished.[63]

She then gives a careful and rigorous diagnosis of what goes wrong in paradigmatic "I'll be glad I did it" reasoning. As she points out, it is morally impermissible for parents choose to disable their children or to permit them to remain disabled based on reasoning about future preferences that stem from adaptive preference formation. She distinguishes such cases from the case of cochlear implantation, holding that the question of what is morally permissible in cases like these is still open. I agree with Howard's conclusions about moral permissibility and faulty "I'll be glad I did it" reasoning.

However, as should be clear from my discussion of Harman 2015, above, the basic structure in "I'll be glad I did it" reasoning can differ dramatically from the structure of a case involving transformative decision-making. As a result, it is not clear how Howard's conclusions apply to transformative decision-making, especially to the problems concerning the inaccessibility of the subjective value of lived experiences in cases of transformative change. (In Paul 2014, I discuss this problem in the context of decisions involving cochlear implantation.)

Recall that, when making a transformative choice, the act the agent performs can lead to events that transform her self-perspective. This is not a mere change in preferences; it is a transformation of one's epistemic capacities and a replacement of some core personal preferences. In transformative change, the self that results from the transformative choice can have preferences that are incommensurable with those of the earlier self. Moreover, the preferences of the transformed self can include the preference to be that transformed self.

Although Howard is correct that both "I'll be glad I did it" reasoning and transformative choice involve changes in a later self caused by an earlier self, the problems of transformative choice are not created by adaptive preference formation.[64] Rather, the source (and the magnitude) of the problem with transformative decision-making arises from the inability of the self to endorse epistemically inaccessible preference change across

[63] See Barnes Forthcoming for related discussion.

[64] Problems with transformative decision-making remain even when there is no constriction whatsoever in the life options of the person who undergoes (or who refuses to undergo) the transformative experience.

contexts of radical self-change. Using Howard's notion of "endorsement," the problem with transformative choice is that, unless you can endorse the preferences of your future self *before you change your current self*, your future self's preferences cannot justify your choice at *t1*.[65] (This is why, contra Howard's assertion [2015, 369], relying on testimony about your future self's preferences isn't an easy way out of the problem of rationally deciding to become a parent. You lack the ability to endorse the testimony.)

So transformative change involves incommensurable, epistemically inaccessible preference change from one self to another. We cannot distinguish acceptable transformative preference change from unacceptable transformative preference change merely by distinguishing adaptive reasoning or "sour grapes" preference formation from other types of preference formation.[66] Nor can we solve the problems of transformative choice simply by making a more careful comparison of the options for the agent. As a result, while Howard has made significant progress in developing our understanding of "I'll be glad I did it" reasoning, her solution to that problem fails to engage with the deeper problem of transformative choice that cochlear implant cases and other sorts of life-changing decisions raise.[67]

Ruth Chang (2015) develops the connection between transformative experience and contemporary debates in ethics about reasons, self-constitution, normative character, and objective values. She describes my view of transformative choice as "event-based" transformative choice and contrasts it to "choice-based" transformative choice, distinguishing between the different types of transformation involved and their implications for practical decision-making. The transformation in choice-based transformative choice involves changes in one's normative character (the change could be minor or major), whereas event-based transformative choice involves major changes in one's epistemic capacities and personal preferences.

In her paper, Chang (i) explores the relationships between event-based transformative choice and choice-based transformative choice. She argues that event-based transformative choice poses no threat to decision theory because (ii) experiences like having a baby are not epistemically transformative, (iii) objective value rather than subjective value is the value of interest for transformative decision-making, and (iv) the possibility of radical personal transformation can be solved by standard approaches to rational choice.

---

[65] Similarly, if you are choosing for another person, such as a child, if what you chose at *t1* formed her preferences at *t2*, the fact that her preferences at *t2* are such that she prefers to have those preferences not justify your actions at *t1*. This relates to the discussion of choice ex ante versus choice ex post in my section 5.

[66] The classic text for this is Elster 1983.

[67] Howard, in discussion, emphasizes that her central project is to explore how deference to the testimony of others could play a role in morally and practically justifying our decisions. This feature of her project dovetails nicely with the concerns I raise in Paul 2014 about informed consent, disability, and testimony.

In reply to (i) I give a model for choice-based transformative choice, highlighting its distinguishing feature, mental commitment as a basis for character formation. I show how choice-based transformation differs from event-based transformation, and how it can be embedded into the structure of event-based transformation. I point out that the radical epistemic and personal transformation of event-based transformation can undermine the rationality of making mental commitments involved in choice-based transformative choice. In reply to (ii), I'll discuss a problematic assumption and direct readers to relevant literature. I'll reply to (iii) by explaining that subjective value is part of what grounds objective value in the cases of interest.[68] In reply to (iv), I'll explain how radical personal transformation creates problems for rational choice.

6.4.1 Transformative Choice

Chang distinguishes between what she describes as "event-based" transformative choice and "choice-based" transformative choice. In order to maximize the possibility for productive discussion, I will engage with Chang on her own terms, and assume that my view of transformative choice is close enough to the view that she describes as "event-based transformative choice" to make meaningful comparisons.

In event-based transformative choice, you choose to perform act *A*. On the simplest version, the choice to perform *A* is rational if performing *A* maximizes your expected subjective value. We can frame this in causal terms, as Chang's description suggests: the choice to perform *A* can lead to a transformative outcome *O* that is causally downstream from the choice. For simplicity, I'll assume *A* determinately causes *O*.

It is important to be clear about the causal structure here. On this view, some of the causal outcomes of making a transformative choice to perform *A* are events that transform you, and some are events of your transformation *O* (assuming you are in fact transformed). Given that causation is transitive, this simply amounts to saying that the choice causes your transformation, either directly or by being among the events in the causal chain leading to your transformation. It is certainly the case that in this sense, if you choose to act, you choose to transform yourself. In my introductory remarks, above, I characterized such transformative choices as "life-making" choices. As I put it in *Transformative Experience*:

> when facing [transformative] big life choices, the main thing we are choosing is whether to discover a new way of living: life as a parent, or life as a hearing person, or life as a neurosurgeon, and so forth; that is, we choose to become the kind of person—without knowing what that will be

[68] To forestall confusion: to say "*x* is grounded in *y*" does not entail that *x* is entirely grounded in *y*. The language is similar to causal language: saying "*c* causes *e*" does not entail that *c* is the *only* cause of *e*.

> like—that these experiences will make us into. (Paul 2014, 123)

Chang raises a series of interesting concerns about the rationality of self-formation and the problems for practical decision-making, and contrasts event-based transformative choice with choice-based transformative choice. To understand the contrast, we must first understand how choice-based transformative choice is supposed to be different from event-based transformative choice.

Chang describes choice-based transformative choice as the kind of choice where "you change who you are by the very making of a choice, not by some experience or event downstream from your choice" (2015, 239). How is this different from saying that performing *A* causes *O*?

The idea is that the choice-based transformative choice is a choice made *before* the choice that is the performance of act *A* leading to outcome *O*. Chang describes it this way: "When we choose in a thick sense, that is, by committing to an alternative, we create reasons for ourselves to choose it—our commitment is that in virtue of which we have a reason to do something" (242).

When you commit to an alternative, you choose to prefer that alternative over the others. After you commit to an alternative by (mentally) preferring one act-outcome (*A*–*O*) sequence over other possible sequences, you then choose again. This time, you choose to act in a way that is most likely to bring about that outcome, that is, you choose to perform *A*. To understand Chang's sentence quoted above, then, we need to see that only the first use of "choose" in her sentence involves the "thick sense" of choice. This sort of choice is a mental decision to prefer an alternative, such as deciding to pursue a course of action. The second use of "choose" in the quote involves the "thin sense" of choice, a choice that is merely a performance of an act, such as physically performing act *A*.

This gives us an interpretation of the structure for choice-based transformative choice as a structure that is temporally prior to the causal structure involved in event-based transformation. Choice-based transformative choice involves a mental commitment to one act-outcome sequence over another, where mental commitments create reasons. The idea is that this structure is embedded in the early part of the causal sequence that ultimately leads to the choice to perform act *A*, and *A* in turn leads to outcome *O*.

Let's flesh it out a bit more with reference to our paradigmatic example, choosing to have a child.

Start with an important distinction, drawn from the distinction between thick and thin choice, between committing to an outcome in the sense of *mentally deciding to endorse* performing act *A* that leads to outcome *O*, and committing to an outcome in the sense of *actually performing* act *A* in order to bring about *O*. We need to be clear about which type of commitment we are talking about, so call the first type "mental commitment" and the

second type of commitment "performative commitment." Let's take the outcome *O* to be the effect that is the final product: the transformed person. In our example, *O* is the outcome of becoming a parent and forming a parent-child attachment.

In choice-based transformation, first, you mentally commit to an act-outcome sequence from *A* to *O*. For example, you mentally commit to having a child, that is, you mentally commit to performing the act of having a child with the outcome of becoming a parent and forming a parent-child attachment. This mental commitment, by hypothesis, creates a will-based reason (R): so you now have a will-based reason to bring about *O* via *A*. In our example, you now have a will-based reason to have a child and become a parent with a parent-child attachment.

This new will-based reason constitutes a new normative character for you. This is what Chang wants to highlight when she says "So by choosing, we can create new reasons for ourselves, thereby transforming 'who we are.'" In Chang's sense, we change who we are by changing our normative character. Now that you have a reason to have a child, you have a new normative character, one which reflects your desire to become a parent. This is the structure that is distinctive of choice-based transformative choice.

The rest of the scenario for our paradigmatic case of choosing to have a child just involves event-based transformative choice.[69] Once you have a will-based reason *R* for bringing about *O* via *A*, you then make a performative commitment to bringing about *O* by choosing to *A*. So you then choose to perform act *A*, and bring about *O*. Continuing the example, because you have a will-based reason to have a child, you choose to have a child so that you will become a parent and stand in a parent-child attachment relation.

Once the structure of such transformative choices is worked out, we can see that choice-based transformative choice and event-based transformative choice involve two different types of change we can undergo when making big life choices. As a result, they raise different kinds of issues to address when we understand the possibilities for how we might construct ourselves via our choices.

Choice-based transformative choice concerns the rationality of practical decision-making given the way we understand and justify the mental commitments that create our reasons and normative character.[70] In choice-based transformation, your mental commitment creates new reasons that constitute a new normative character. The change in character need not

[69] We can also continue to add preliminary "choice-based transformative choice" structure in various scenarios: perhaps you mentally commit to mentally committing to the act-outcome sequence from *A* to *O*. Then you create a will-based reason *R** that is a reason for your will-based reason *R*. One interesting question concerns the way to understand and rationally justify this sort of regressive structure for will-based reasons. This question raises problems for debates in practical ethics over the nature of self-constitution and reasons.

[70] "We can now see how in choice-based transformative choices your choice can be both what transforms you and that in virtue of which you are transformed. In deciding whether to have a child, by hypothesis, the given reasons are on a par. You have the normative power

be a radical change. Your choice might be a small one, and so your new normative character might be pretty similar to your old one. Still, in this sense, your choice (your mental commitment) makes you who you are.

Event-based transformative choice concerns the possibility of radical epistemic change, the lack of a subjective value function, and the problems for decision-making when radical epistemic change is accompanied by radical personal change involving inconsistent preferences. In event-based transformation, you transform yourself by performing an act that causes a radical change in your epistemic capacities and your core personal preferences. In this much more dramatic sense your choice (your performative commitment) forms who you are.

While the types of transformation and self-formation are different in each kind of choice, the possibility of epistemic and personal transformation does raise a problem for choice-based transformative choice. In particular, if hard choices can result in outcomes that are epistemically and personally transformative, this can make the alternative act-outcome sequences "noncomparable" (in Chang's sense) for an agent who mentally commits to one *A*–*O* sequence over another. This brings out how the possibility of transformative experience and transformative decision-making can have implications for debates about morality, self-constitution, and rationality.[71]

Return to the choice to have a child. If being a parent is epistemically and personally transformative, then the agent cannot grasp the subjective value of that outcome before she becomes a parent. This means that, at the mental commitment stage, the agent cannot evaluate the subjective value of the outcome of having a child. As a result, the value of the outcome of having a child is "noncomparable," that is, the agent cannot compare it to the subjective value of the outcome of remaining childless. This case undermines Chang's general thesis about how to understand mental commitment and choice-based transformative choice as a form of rational decision-making, since values for alternatives must be accessible and evaluable in order to guide reason and choice.

In other parts of her paper, Chang argues that epistemic and personal transformation create no problems for rational decision-making. I reject

---

to commit to one of the options or one of its features. You might commit to forming a parent-child attachment. That commitment just is choosing to have a child in the thick sense. That commitment then creates new will-based reasons for you to have a child, that is, your commitment is that in virtue of which you now have a new will-based reason to have a child. Your new will-based reason then interacts with your other, given, reasons and guides your choice in the thin sense. You may now have most all things considered reasons to choose to have a child. Your new will-based reason transforms you because it is a reason that determines your normative character. You are now the sort of person who has most all things considered reasons to have a child. Before the choice you were the sort of person for whom the reasons for having a child and remaining child were on a par. By choosing, you change the reasons that determine your normative character" (Chang 2015, 275).

[71] In addition to the connection to Chang's work, there are connections to Korsgaard 2009 and Kierkegaard 2006. Also see Barnes 2015b and Kemp 2015.

these arguments. Below, I'll discuss some interesting issues that come up in the discussion.

6.4.2 *Are Experiences Like Having a Baby Epistemically Transformative?*

Chang argues that the subjective values of the outcomes of transformative choices can be known, at least well enough, before the transformative experience occurs.

> Maybe the experience of having a child falls under types that include the experience of being in a family, passing a kidney stone, having a pet, and so on. Since you've had experiences that fall under the same types before, you will know something about what it's like to have a child. (249)

The suggestion that passing a kidney stone (while your sister is visiting? as your cat looks on?) teaches you what it's like to give birth is, quite frankly, bizarre.[72] While Chang's other objections to the possibility of epistemic transformation are many and far-ranging, they also miss the mark, largely because the views she attacks bear only a passing resemblance to my own views. As space is limited, readers interested in critical discussion of epistemic transformation should consult Barnes 2015b, Campbell 2015 and Paul 2015a, as well as Collins 2015, Dougherty et al. 2015, Kauppinen 2015 and my replies to these papers. Sharadin (2015, section 2), does an excellent job of characterizing the idea that it is the distinctive nature of the lived experience of having a child that is relevant to my argument concerning the rationality of choosing to have a child based on what it will be like to become a parent, not merely the experience of changing diapers, feeling tired, etc.

6.4.3 *Is Subjective Value the Type of Value of Interest?*

Chang argues that subjective value is not a suitable ground for the values of the relevant decision outcomes. Now, I've been at pains to point out above in my reply to Kauppinen that subjective value is not mere subjective feel. As I put it in Paul 2014, it's the value of *lived experience*, and as such, should be able to ground, at least partly, the values involved in many big life decisions. The subjective value of a lived experience is not merely a matter of the phenomenal character of the internal characteristics of one's

[72] Chang seems to radically underestimate the real-life epistemic difficulties here. Preliminary empirical results in psychology from the work of Josiah Nunziato and Fiery Cushman indicate that many people report having transformative experiences when they experience significant life events such as becoming a parent. In particular, they report that the experience changed them in ways that they could not foresee. In addition, further results suggest that people are over-confident about their ability to anticipate the changes that will occur in their beliefs, preferences, desires, and values as a result of a transformative experience. (I thank Fiery Cushman for discussion.) For an interesting and relevant example of the kinds of complications and massive difficulties involved in gaining and regaining sight see Sacks 1993.

inner life. It's a richer value, a value that includes what it's like to live in a particular set of circumstances, to live one's life in a particular way, or to "live an outcome." (For further discussion of how subjective value extends past one's internal mental life, see Campbell 2015 and Paul 2015a.)

Chang thinks that what really matters for transformative choices are what she describes as "objective values" and "objective goods," not subjective values. She uses the example of Mike May, who was blind but regained his sight through an operation:

> What matters in May's choice about whether to see again are not only the objective and subjective values of the *experience* of seeing but also the objective *goods* (which we can characterize in terms of events) he will have in his life if he is sighted. Indeed, it makes sense to think that it was not the experience of seeing that primarily transformed him but other events, like communing with his wife over a beautiful sunset, responding to visual feedback from his children, and learning new skills that gave him greater opportunities that did the transforming work. (Chang 2015, 262)

As an objection to the need for assessments of subjective value this is puzzling. For of course goods like "communing with his wife over a beautiful sunset" are the ones that matter. But a good like this depends at least partly on the nature of May's lived experience, and the subjective values of the experiences Chang describes are precisely what May cannot evaluate before he has his operation.[73] As I'd put it, for May, a grasp on the subjective value of regaining his sight is necessary to assess whether an outcome like communing with his wife over a beautiful sunset is objectively good—and if it is objectively good, just how good it is.

Some of Chang's arguments about our grasp of the relevant values concern testimony, and her objections reflect those that are raised by other contributors to this volume. I discuss the relationship between experience and subjective value in my reply to Kauppinen 2015, and discuss the limitations of testimony in determining subjective value in my discussion with Dougherty et al. (2015), above, with Harman (2015), Howard (2015), and in my exchange with Richard Pettigrew (2015; 2015a).

6.4.4 *Does the Possibility of Radical Personal Transformation Challenge Decision Theory?*

Chang suggests that decision theory can manage the cases involving personally transformative experiences that Edna Ullmann-Margalit (2006) and I

[73] We are assuming, by hypothesis, that even though May once had sight, before having his operation to restore his sight, he lacked the capacity to represent events involving seeing things like sunsets.

describe. She claims that the problem can be solved simply by predicting and assessing future preferences or by employing "a master utility function" that could order consecutive sets of preferences with respect to any given choice. But, of course, this misdescribes the situation. Our examples of personal transformation crucially concern *discontinuous* preferences, including discontinuous higher-order preferences. These are cases where no master utility function or prospective resolution is rationally available.[74]

It is worth setting the context with some familiar examples. Jon Elster considers a case where Ulysses knows his future self will be temporarily irrational. In such a case, Elster argues, it is rational for Ulysses to bind his future (irrational) self (Elster 1979). Derek Parfit's young Russian nobleman values wealth redistribution, but knows that when he is old, he will value keeping his money instead. By assumption, the preferences of the young Russian nobleman are discontinuous with the preferences of the old Russian nobleman. Parfit suggests that the solution is for the young Russian nobleman to bind his future self, just as Ulysses did, even though the nobleman's future self is not irrational.[75]

If we have rational grounds for privileging the preferences of the current self over the preferences of the future self, perhaps because the current self and the future self share higher-order preferences (or there is some independent reason that the future self's preferences are rationally disqualified), then the current self can rationally choose to bind the future self consistent with her current higher-order preferences.

However, this solution is untenable in transformative contexts such as choosing to have a child: it cannot provide a rational guide for life's transformative decisions.[76] In transformative decisions like choosing to have a child, certain core preferences of the current self are radically discontinuous with preferences of the future self, and there is no independent basis for disqualifying these preferences of the future self. Moreover, because of the epistemic inaccessibility of the preferences of the future self, the current self cannot (imaginatively) prospectively assess the lived experience of the future self to form a higher-order preference that would be consistent across the selves at different times.[77]

There is another implication in the neighborhood. If morality is bound up with rationality, and the decision for one's future self concerns a morally transformative experience, then the problem for rational decision-making, especially for rational decisions concerning the construction of one's future

[74] Moreover, in my cases, we lack epistemic access to our future preferences. For discussion of the issue, see Briggs 2015 and Pettigrew 2015 and my replies.
[75] Korsgaard (2009) challenges this.
[76] Unless, as I argue, the current self chooses solely on the basis of preferring to discover new preferences, that is, on the basis of the preference to replace the current self with a new self. (Or, presumably, to end the existence of the current self, independently of the question of replacement.)
[77] See Briggs 2015 for a different response.

self, becomes a problem for moral decision-making and the construction of one's future moral perspective.

Korsgaard (2009) argues, in circumstances where core preferences will be radically transformed, your current self must either take your future self's preferences into account, or regard your future self as irrational. In my cases of transformative choice, you cannot take your future self's preferences into account. First, you lack first-personal epistemic access to your future self's preferences. And second, you cannot simply prefer to prefer your future self's preferences, because they conflict with your current preferences, including your higher-order preferences. Third, being told by an expert what your future self would prefer you to do now is not sufficient for you to prefer that future self's preferences, because without first-personal imaginative understanding of this future self, you are entitled to regard her as irrational (or alien).

Thus, the problem remains: how can it be rational to choose to have a child? Or, as one might put it, how can it be rational for you to make yourself into someone you regard as irrational?[78]

6.5 Empirical Research and Choosing to Have a Child

There is a certain sort of easy reply to the problem of transformative decision-making. It is to say that a person should simply replace her introspection with scientific evidence that will tell her what to expect.[79] Nathaniel Sharadin (2015) argues for a version of this reply, arguing that you can use currently available empirical research to predict what it would be like for you to become a parent, and this will solve the problems with the transformative decision to become a parent. His reply fails. Why?

First, there are practical difficulties. As I discuss in section 4, above, contemporary psychological work is simply not yet advanced enough for us to use it to make sufficiently accurate predictions in our own case about how we'd respond to having a child. Once introspection is set aside, without suitably precise empirical information about my own particular, individual response to having a child, I'm left adrift. All I have is highly general, incomplete, empirical information that I'm somehow supposed to interpret in a way that applies to my own particular case. In a high-stakes case like this, where I am making one of the most important, irreversible, momentous, and personal decisions of my life, currently available psychological data cannot provide a satisfying replacement for introspection.

Second, even if we do have sufficiently complete, detailed empirical results, there are deeply philosophical interpretive difficulties with the testimony grounding the evidence. Almost without exception, one of the most serious problems is that the empirical results measure the preferences and

[78] In Paul 2014, I propose a solution involving revelation.

[79] We saw a version of this reply in Harman 2015, where the suggestion is that we should simply rely on "reliable testimony."

the satisfaction of *already transformed* individuals.[80] As I have discussed in detail in section 5 above, the epistemic perspective and preferences of the agent after she is transformed may be discontinuous with her epistemic perspective and preferences when she is making the decision, raising questions about how she is to understand and interpret the evidence.

Sharadin (2015) contests both claims, but his argument focuses primarily on the first difficulty: the use of current psychological and sociological data to make the transformative decision to become a parent. He grants that introspection about the nature of the experience can fail, but thinks a person can simply replace introspection with current empirical research to discover what it would be like for her to have a child.[81]

> [P]rospective parents cannot rationally decide to have a child by reflecting on the phenomenal character of that experience: it's in principle epistemically inaccessible to them. But this does not mean that prospective parents cannot rationally decide to have a child by reflecting on what it is like to have a child. It just means they have to take a somewhat circuitous route: prospective parents must reflect on the non-phenomenal features of the experience, on what they themselves are like, and on the principles that link how they are to how the experience is likely to affect them. (450)

According to Sharadin, contemporary social and psychological science gives us these linking principles, and a person can use them to "reasonably expect" the valence of what it will be like for her to become a parent. From this, he concludes that a person can indeed rationally choose to become a parent based on what it will be like for her.

There are two serious problems with his argument.

The first problem is that Sharadin claims that I deny the existence of such linking principles. But I do not deny the existence of such linking principles. I am happy to grant that there might exist such psychological "laws" for individuals. It's discovering them that's the problem.

So let's assume that linking principles exist. The second, much more important problem with Sharadin's argument is that he thinks it is manifestly obvious that we know what many of these linking principles are, and that an individual can and should use them to determine what the valence of her outcomes will be.

[80] There are also problems involving average effects, the fundamental identification problem, and reference class worries. These problems are significantly more serious when introspection is unavailable. See Paul 2014 and 2015a.

[81] I found Sharadin's discussion in section 2 of his paper, of the distinctive nature of what it's like to have a child, to be thoughtful and interesting, and his discussion of the role of swamping and the relevance of distinctive subjective values in the argument for epistemic transformation is right on target. Section 3 is where the problems begin.

> [T]here manifestly *are* such linking principles, and we know what some of them are. For just one example, depression on the part of either parent, but especially maternal depression, is linked to both affective and behavioral disorders on the part of children (Lovejoy et al. 2000; Tan and Ray 2005). And parents of affectively or behaviorally disordered children report significantly higher rates of stress and lower levels of subjective well-being—as good a measure as any of the valence of the phenomenal character of their experience of what it is like to have a child (Tan and Ray 2005, 77). (Sharadin 2015, 449)

He goes on to argue that

> This should come as no surprise at all: what people are like helps determine how things turn out for them. And, thanks to years of psycho- and sociological research, we can often safely predict how things will turn out for an agent given enough psycho- or sociological information about them. Of course, the situation is no different when it comes to ourselves than it is in the case of others. Or at least, it is not relevantly different. Just as I can know that, given that some agent is depressed, the phenomenal character of her experience of having a child is unlikely to be positive, I can know of myself that, given I am depressed, the phenomenal character is unlikely to be positive. And so, ceteris paribus, I can safely predict that it would be unwise, just now at least, for me to have that experience.[82] (449)

Unfortunately, however, Sharadin has misunderstood the implications of the work he is citing, for Tan and Ray (2005) and Lovejoy et al. (2000) don't give us anything like what an agent would need to in order to reasonably predict, in her own individual case, what it would be like for her to have a child. (By extension, they do not give us the linking principles he thinks they provide.) This is no surprise: as I noted in Paul 2015b, the relevant psychological and sociological science is not yet complete enough for an individual to discover her own personal linking principle or even a reasonable approximation thereof.

A closer look at the papers will be instructive: it will demonstrate the perils of relying on naïve interpretations of empirical work in an attempt to shrug off the implications of transformative experience for decision-making. In particular, a closer look will bring out first, how easy it is for nonexperts to misunderstand what empirical work actually tells us, and second, just how difficult it can be to take empirical results and apply them to your own case.

[82] He acknowledges "there might be countervailing reasons . . . to expect it will be positive" but "The point is just that such expectations are sometimes warranted" (449).

The Tan and Ray paper is a small study done in Malaysia. The authors have a matched sample of depressed and non-depressed children, and ask whether parents of depressed children have a harder time. The answer seems to be yes: "Parents of depressed children reported higher parenting stress and were more likely to perceive their children as 'difficult'" (Tan and Ray 2005, 76). But note: this study does not give you information about your chances of having a child that will be depressed. Moreover,

> [u]nivariate analysis demonstrated a relationship between children's depression and maternal depression but not with paternal depression. This became non-significant with multivariate analysis, implying that maternal depressive symptoms may have been due to caring for a depressed, thus 'difficult' child. (76)

In other words, a prospective parent could worry that depressed mothers have depressed children. But while the data show a correlation between the two, simple further analysis with a few controls made the association disappear. This suggests that it might be the child that makes the mother depressed, not the other way around.

So a closer look at the Tan and Ray paper provides no support for the claim that it provides knowledge of linking principles a person could use to predict her outcomes of what it would be like, if she were depressed, for her to become a parent.

The Lovejoy et al. 2000 paper is a much more substantial review. It is a widely cited meta-analysis seeking "to assess the strength of the association between depression and parenting behavior" (561). It's a typical and responsible example of its type, covering more than twenty years of peer reviewed studies of different kinds on the relationship between maternal depression and parenting behavior.

Again, note that this review does not speak at all to the question of who is likely to become depressed should they have a child. Lovejoy et al only look at studies of people who are *already* mothers and, moreover, at mothers who are *already* depressed.

With this in mind, consider the findings in Lovejoy et al. 2000. For the sake of argument, let us simply assume that, somehow, as a prospective parent, you already know that you will be depressed after you have your baby. This would be extremely difficult to establish in practice in most cases, and making this assumption rather misses the point of the entire discussion of transformative experience. But as the empirical evidence Sharadin cites would be otherwise completely irrelevant, it seems charitable to grant the assumption. What then can we "safely predict" from the data presented?

Lovejoy et al's review asks, given that you are a depressed mother, how are you likely to behave toward your child? Do depressive mothers act more negatively towards their children, are they dissociated from them, or are they positive towards them? Lovejoy et al suggest there is indeed an

observed relationship between maternal depression and parenting behavior. But it is not a simple one.

> The association between depression and parenting was manifest most strongly for negative maternal behavior and was evident to a somewhat lesser degree in disengagement from the child. The association between depression and positive maternal behavior was relatively weak, albeit significant. (561)

This result is summarized in Table 3 of the paper (Lovejoy et al. 2000, 579). They find a positive and statistically significant association between maternal depression and negative behavior. They also find a positive and significant association between maternal depression and disengaged behavior. And finally they find a positive and significant association between maternal depression and positive behavior! (The effect sizes vary. The association with negative behavior is the largest. The association with positive behavior is about two and a half times smaller, but still positive. It is also significant according to conventional standards.)

If it seems strange to you that all three behaviors could be positively associated with maternal depression, bear in mind this is a meta-analysis of many studies. Positive and negative behaviors are not measured on a single scale with positive on one end, negative on the other, and dissociated in the middle. Instead they are measured separately through the observation of many behaviors.

From a first-personal point of view, this means that—given that we have granted that you know with certainty that you will be depressed after you have your baby—each of the measured behavioral outcomes remains a live possibility for you. It is a possibility just in the important but weaker sense that you might be an "outlier" with respect to a single general tendency in the population, but also in the specific sense that the review finds all three patterns are in fact observed as significant general tendencies. This means that acting on the basis of the best evidence you have, you, as a depressed person, are left largely unsure about what sort of behavior you would be likely to show to your child, if you had one. While you can certainly attach credences to different possible outcomes based solely on this very general information, it is hardly specific enough to count as giving you evidence (much less knowledge) about your own possible outcomes, let alone a workable "linking principle" for a decision at the individual level.

So how, then, is our prospective decision-maker to "safely predict" how she is likely to behave? At the outset, the authors issue a general caution:

> It is important to note, however, that the association between depression and child adjustment problems may not be causal. Child behavior problems could, for example, contribute to the development of maternal depression. It is also possible that a third variable is causally related to both

> maternal depression and child adjustment problems. . . . [T]here are a number of associated features of depression that may account for the relationship between maternal depression and childhood difficulties. These include substance abuse, personality disorders, and marital discord. (562)

This is an important caveat, and the fact that it is a routine one in summaries of observational studies should not tempt us to simply set it aside. As we move on to the details, we see, as is typical of reviews of this sort, that the authors make a careful and cautious interpretation of the observed patterns:

> Our statistical analyses of 46 observational studies demonstrated a moderate association between maternal depression and parenting behavior in the domain of negative behavior, a small to moderate effect in the domain of disengaged behavior, and a small effect for positive interactions. Thus, depression appeared to be associated most strongly with irritability and hostility toward the child, to be associated to a somewhat lesser degree with disengagement from the child, and to have a relatively weak association with rates of play and other active and pleasant social interactions. However, there was marked variability in the effect sizes obtained in each domain of parenting behavior, only some of which could be explained by the moderators we included in our analyses. (583)

So although some general patterns are evident, there is more than one significant tendency. What we should conclude from this is that, even with twenty years of data, things are a long way from being fully explained. Moreover, the available covariates are both relatively few in number and relatively limited in their explanatory power. The authors also express some (again, responsible and quite typical) concern about how different behaviors were coded and assessed across studies. They note for example that

> negative coercive behaviors and positive behaviors were coded fairly evenly across the studies, regardless of child age; however, disengaged behavior was assessed primarily in mothers of young children. . . . Because the behaviors coded in the original studies were more varied for the mothers of very young children, it lends the appearance that the vestigial parenting problems associated with depression are more pervasive for infants and toddlers. We believe this is more than an artifact of the coding system and reflects the dependency of infants and very young children on their

> caretakers to initiate interaction, and maintain contact that is coordinated with the child's affect and behavior. (584)

Heterogeneity in the measurement instrument, and its possible correlation with outcomes of interest (e.g., the age of the child), make interpretation even more difficult. So, for example, the authors note that:

> Although certainly not conclusive, the obtained pattern of age differences argues against strong child effects in the development of depression and parenting difficulties.

And:

> Although our findings do not resolve questions of causality in the relationship between depression and parenting, they also do not suggest that the majority of parenting difficulties of depressed mothers originate from individual differences in child behavior. (585)

Because the authors are specifically interested in the effects of depression—considered not simply as a folk category but as a well-defined clinical condition—they also note a potential theoretical issue arising from the findings:

> With respect to the pattern of effect sizes, our analyses demonstrated that the largest effects occurred for negative/coercive behavior; however, involvement, sensitivity, and pleasant social interactions would logically seem to be most sensitive to depressive symptoms. Relying solely on the diagnostic criteria for depression, one would predict that depressed mothers would have the most difficulty in the domain of engagement. . . . Thus, the defining characteristics of depression would lead us to expect the effect sizes to be largest for disengaged and positive behavior and smallest for negative/coercive behavior. In contrast, we found the effect size to be largest for negative behavior and smallest for positive behavior. This finding, which is not consistent with predictions based on the symptoms of depression, suggests that some of the parenting problems observed among depressed mothers may be associated with negative affectivity, and that these same parenting difficulties may occur among women with other emotional problems and general psychological distress. (587)

Here we see a pattern of reasoning—again, very common in research of this kind—where the results prompt a reevaluation of the original theory rather than simply establishing a "finding" with a clear recommendation attached.

The authors suggest that instead of thinking in terms of depression,

> conceptualizing depressive parenting problems in terms of disturbances in positive and negative affect leads to predictions somewhat more consistent with the findings of our meta-analysis and provides a useful theoretical model from which to interpret our results. (587)

Note that the concern *is not to develop or pinpoint advice to particular sorts of individuals*, let alone for individuals in any fine-grained sense. Instead the concern is to try to square the observed pattern of results with a general claim about the nature of an association *across a population*.

The paper closes with a call for further research, and more caution:

> In summary, our analyses suggest that the strength of the association between depression and parenting behavior varies as a function of the type of behavior observed. Effects were strongest for negative/coercive behavior and least strong for positive interactions. However, there was significant variability in effect sizes within each category of behavior and many of the effect sizes for individual studies did not differ from zero. Although our findings suggest that age of child, socioeconomic status, and timing of depression account for some of the differences between studies, significant heterogeneity in effect sizes remained. . . . Explication of the variability of effects remains a critical task for understanding the developmental risks associated with maternal depression. . . . The results of our analyses are consistent with the hypothesis that parenting behaviors are a component of the risk associated with living with a depressed mother. . . . However, further research is needed to more fully understand which children of depressed mothers are most likely to be exposed to inadequate parenting. (588)

This should make it abundantly clear that the research in the papers cited by Sharadin has nothing to say to the prospective parent worrying about whether having a child will make them depressed. This is because they only study the behavioral consequences, positive and negative, for people who are already parents and already depressed. And, as should also be clear at this point, even a depressed individual will not find any "linking principles" that will support inferences about what it will be like for *her* to become a parent.

However, in a deeper sense, the reviews do cast light on the role of research like this in making potentially transformative decisions, and this is why I have discussed them at such length. We find at least three kinds of illumination. First, the careful analysis and discussion on the part of the authors of these studies shows just how difficult it can be to make a clear causal analysis in these cases, and how reluctant responsible researchers are

to go beyond the data to make simple recommendations to specific people about how they should choose based on the evidence.

To be sure, there are some cases where things are easier. If you and your partner are tested for Huntington's Disease, then you will learn some very specific and relevant facts about your own future and the chance that any future child of yours will have the disease. But many or most relevant psychological aspects of your possible future as a parent are not as causally specific or empirically testable as this.

Second, even very good, responsible sociological and psychological research—of the sort carried out and reported in Lovejoy et al—can take an extremely long time to reach its conclusions, relative to the decision windows of prospective parents. The questions raised in the Lovejoy et al review could well prompt a further twenty or thirty years of study that might yield a round of real but slow progress. Science advances. But for many who have to choose now, or soon, it does not advance nearly fast enough. And so even if we, collectively, eventually arrive at a well-established and relevant body of knowledge, being able to "safely predict" outcomes of interest is, for a huge range of cases, a very long way off. There is no point in grandly referring to "years of psycho- and sociological research" that either does not give you the information you need now, or that might provide it to you in a more appropriately fine-grained fashion a few decades after your death.

Third, as I said at the beginning, although it is a model of responsible meta-analysis, Lovejoy et al's article is in fact strictly irrelevant to the person facing a potentially transformative choice. Sharadin might object that the cited papers were merely "just one example" meant to serve as a sort of placeholder for a vast body of well-established, properly-validated, causally impeccable, internally consistent social-scientific findings of direct relevance to individuals facing the particular choices they are interested in. I do not think this body of knowledge exists in this form.

In an interesting way, Sharadin's discussion mirrors the epistemic attitude of most people who face possibly transformative decisions in their own lives. Such people often have strong intuitions about what they want; they have a reasonable belief that research findings should have a role to play in helping them choose; and they have a conviction that there must be a rational course of action that combines their feelings with "the data."

But where the research, especially in these areas, is careful, complex, controversial, slow-moving, and concerned mostly with tendencies at the level of whole groups, ordinary people need to make individual-level decisions, and they need to make them *now*. And so they sit at their computer and cast their line into a sea of scientific research. They reel in a few studies and read the abstracts. They fail to grasp what the research can actually establish, they glide over the caveats inserted by the authors, and they convince themselves that with these papers in hand they can—ceteris paribus!—"safely predict" their own futures. The result is a parody of

rational choice, a pretense of rationally justified decision-making rather than a clear-eyed step into an acknowledged unknown.

L. A. Paul
E-mail : lapaul@unc.edu

References:

Barnes, Elizabeth. 2014. "Fundamental Indeterminacy." *Analytic Philosophy* 55 (4): 339–362. http://dx.doi.org/10.1111/phib.12049.

Barnes, Elizabeth. 2015a. "Social Identities and Transformative Experience." *Res Philosophica* 92 (2): 171–187. http://dx.doi.org/10.11612/resphil.2015.92.2.3.

Barnes, Elizabeth. 2015b. "What You Can't Expect When You Don't Want to be Expecting." *Philosophy and Phenomenological Research* 91 (3).

Barnes, Elizabeth. Forthcoming. *The Minority Body*. Oxford: Oxford University Press.

Briggs, Rachel. 2015. "Transformative Experience and Interpersonal Utility Comparisons." *Res Philosophica* 92 (2): 189–216. http://dx.doi.org/10.11612/resphil.2015.92.2.7.

Campbell, John. 2015. "L. A. Paul's *Transformative Experience*." *Philosophy and Phenomenological Research* 91 (3).

Carr, Jennifer. 2015. "Epistemic Expansions." *Res Philosophica* 92 (2): 217–236. http://dx.doi.org/10.11612/resphil.2015.92.2.4.

Cartwright, Nancy. 2011. "The Art of Medicine: A Philosopher's View of the Long Road from RCTs to Effectiveness." *The Lancet* 377: 1400–1401. http://dx.doi.org/10.1016/S0140-6736(11)60563-1.

Chang, Ruth. 2015. "Transformative Choices." *Res Philosophica* 92 (2): 237–282. http://dx.doi.org/10.11612/resphil.2015.92.2.14.

Collins, John. 2015. "Neophobia." *Res Philosophica* 92 (2): 283–300. http://dx.doi.org/10.11612/resphil.2015.92.2.6.

Dougherty, Tom, Sophie Horwitz, and Paulina Sliwa. 2015. "Expecting the Unexpected." *Res Philosophica* 92 (2): 301–321. http://dx.doi.org/10.11612/resphil.2015.92.2.5.

Elster, Jon. 1979. *Ulysses and the Sirens*. Cambridge: Cambridge University Press.

Elster, Jon. 1983. *Sour Grapes: Studies in the Subversion of Rationality*. Cambridge: Cambridge University Press.

Harman, Elizabeth. 2009. "'I'll Be Glad I Did It' Reasoning and the Significance of Future Desires." *Philosophical Perspectives* 23: 177–199. http://dx.doi.org/10.1111/j.1520-8583.2009.00166.x.

Harman, Elizabeth. 2015. "Transformative Experience and Reliance on Moral Testimony." *Res Philosophica* 92 (2): 323–339. http://dx.doi.org/10.11612/resphil.2015.92.2.8.

Howard, Dana Sarah. 2015. "Transforming Others: On the Limits of "You'll Be Glad I Did It" Reasoning." *Res Philosophica* 92 (2): 341–370. http://dx.doi.org/10.11612/resphil.2015.92.2.9.

Johnston, Mark. 2006. *Perceptual Experience*. Edited by T. S. Gendler and John Hawthorne. Oxford: Oxford University Press.

Kauppinen, Antti. 2015. "What's So Great about Experience?" *Res Philosophica* 92 (2): 371–388. http://dx.doi.org/10.11612/resphil.2015.92.2.10.

Acknowledgements I'd like to thank Nomy Arpaly, Elizabeth Barnes, Paul Bloom, Rachael Briggs, Jennifer Carr, John Collins, Fiery Cushman, Elizabeth Harman, Kieran Healy, Dana Howard, Jonathan Jacobs, Antti Kauppinen, Matt Kotzen, Enoch Lambert, Rachel McKinnon, Sarah Moss, Josiah Nunziato, Cass Sunstein, Julia Staffel, and Greg Wheeler for discussion. This research is part of the Experience Project, and was supported with a grant from the John Templeton Foundation. The opinions expressed in this publication are those of the author's and do not necessarily reflect the views of the John Templeton Foundation.

Kemp, Ryan. 2015. "The Self-Transformation Puzzle: On the Possibility of Radical Self-Transformation." *Res Philosophica* 92 (2): 389–417. http://dx.doi.org/10.11612/resphil.2015.92.2.11.

Kierkegaard, Søren. 2006. *Fear and Trembling*. Edited by C. Stephen Evans and Sylvia Walsh. Cambridge: Cambridge University Press.

Korsgaard, Christine. 2009. *Self-Constitution: Agency, Identity, and Integrity*. Oxford: Oxford University Press.

Krishnamurthy. 2015. "We Can Make Rational Decisions to Have a Child: On the Grounds for Rejecting L. A. Paul's Arguments." In *Permissible Progeny*, edited by Richard Vernon, Sarah Hannan, and Samantha Brennan. Oxford University Press.

Lewis, David. 1989. "Dispositional Theories of Value." Supplementary Volume. *Proceedings of the Aristotelian Society* 63: 113–137.

Lovejoy, M. Christine, Patricia A. Graczyk, Elizabeth O'Hare, and George Neuman. 2000. "Maternal Depression and Parenting Behavior: A Meta-analytic Review." *Clinical Psychology Review* 20 (5): 561–592. http://dx.doi.org/10.1016/S0272-7358(98)00100-7.

McKinnon, Rachel. 2015. "Trans*formative Experiences." *Res Philosophica* 92 (2): 419–440. http://dx.doi.org/10.11612/resphil.2015.92.2.12.

Moss, Sarah. Unpublished. "Probabilistic Knowledge."

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

Paul, L. A. 2015a. "Transformative Experience: Replies to Pettigrew, Barnes, and Campbell." *Philosophy and Phenomenological Research* 91 (3).

Paul, L. A. 2015b. "What You Can't Expect When You're Expecting." *Res Philosophica* 92 (2): 149–170. http://dx.doi.org/10.11612/resphil.2015.92.2.1.

Paul, L. A. Unpublished. "Preference Capture."

Pettigrew, Richard. 2015. "Transformative Experience and Decision Theory." *Philosophy and Phenomenological Research* 91 (3).

Sacks, Oliver. 1993. "To See and Not See." *The New Yorker*.

Sharadin, Nathaniel. 2015. "How You Can Reasonably Form Expectations When You're Expecting." *Res Philosophica* 92 (2): 441–452. http://dx.doi.org/10.11612/resphil.2015.92.2.2.

Stanley, Jason. 2015. *How Propaganda Works*. Princeton, NJ: Princeton University Press.

Sunstein, Cass. 2014. "The Limits of Quantification." *California Law Review* 106 (2): 1369–1422.

Tan, Susan and Joseph Ray. 2005. "Depression in the Young, Parental Depression and Parenting Stress." *Australasian Psychiatry* 13 (1): 76–79. http://dx.doi.org/10.1080/j.1440-1665.2004.02155.x.

Ullmann-Margalit, Edna. 2006. "Big Decisions: Opting, Converting, Drifting." *Royal Institute of Philosophy Supplement* 58: 157–172. http://dx.doi.org/10.1017/S1358246106058085.

Van Fraassen, Bas. 1999. "How is Scientific Revolution/Conversion Possible?" *Proceedings of the American Catholic Philosophical Association* 73: 63–80. http://dx.doi.org/10.5840/acpaproc19997313.

Velji, Muhammad. 2015. "Change Your Look, Change Your Luck: Religious Self-Transformation and Brute Luck Egalitarianism." *Res Philosophica* 92 (2): 453–471. http://dx.doi.org/10.11612/resphil.2015.92.2.13.

Vermeule, Adrian. 2013. "Rationally Arbitrary Decisions (in Administrative Law)." *Harvard Public Law Working Paper No. 13–24*. http://dx.doi.org/10.2139/ssrn.2239155.

Zelcer, Mark. 2015. "Conscientious Objection and the Transformative Nature of War." *Journal of Military Ethics* 14 (2): 118–122. http://dx.doi.org/10.1080/15027570.2015.1070035.