Philosophy and
Phenomenological Research

Philosophy and Phenomenological Research
Vol. XCI No. 3, November 2015
doi: 10.1111/phpr.12249

Précis of *Transformative Experience*

L.A. PAUL
University of North Carolina, Chapel Hill

As we live our lives, we repeatedly make decisions that affect our future circumstances and shape the sort of person we will become. Some of these are major, life-changing decisions. In such cases, we stand at a personal crossroads and must choose our direction. If we make these sorts of life-changing decisions about our futures rationally, can we also make them authentically?

In *Transformative Experience* I argue that, under the most natural and ordinary construal of decisions like this, we cannot. My argument draws on debates in philosophy of mind about how experience is necessary for us to have certain epistemic capacities and cognitive abilities. It also draws on debates about the intrinsic value of subjective color experience, and the importance of the first-personal perspective in understanding the self and its possibilities. I use familiar examples from these classic philosophical debates to raise new questions about experience, its value, and its role in prospectively assessing our first personal futures. Using formal tools drawn from decision theory, causal modeling, and cognitive science, I assess first personal decision making and self-construction in contexts of what I call "transformative decision-making", a well-defined and—it turns out—very common choice situation in everyday life. In the Afterword, I discuss how the argument has formal applications in and substantive relevance to counterfactual semantics, formal epistemology, and the philosophy of statistics, social science, cognitive science and psychology.

The tension between rationality and authenticity arises when we consider decision-making from the first-personal perspective in contexts of radical epistemic and personal change. A natural way to make major life choices, such as whether to start a family or to pursue a particular career, is to assess our options by imaginatively projecting ourselves forward into different possible futures. But for choices involving dramatically new, life-changing experiences, we are often confronted by the brute fact that before we undergo the experience, we know very little about what these future

outcomes will be like from our own first-personal perspective. Our imaginative and other epistemic capacities are correspondingly limited, with serious implications for decision-making. If we are to make life choices in a way we naturally and intuitively want to—by considering what we care about, and imagining the results of our choice for our future selves and future lived experiences—we only learn what we really need to know after we have already committed ourselves. If we try to escape the dilemma by avoiding the new experience, we have still made a choice.

1. Epistemic and Personal Transformation

Central to the argument is the notion of a transformative experience. As I develop it, a transformative experience is a kind of experience that is both radically new to the agent and changes her in a deep and fundamental way; there are experiences such as becoming a parent, discovering a new faith, emigrating to a new country, or fighting in a war. Such experiences can be both epistemically and personally transformative.

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. A *transformative experience*, then, is an experience that is both epistemically and personally transformative. Transformative choices and transformative decisions are choices and decisions that centrally involve transformative experiences.

2. Transformative Experience and Rational Decisions

The main problem with transformative decisions is that our standard decision models break down when we lack epistemic access to the subjective values for our possible outcomes. We can't grasp these outcomes in the relevant way, the way we'd need to in order to knowledgably assess them. As a result, in cases of transformative choice, the rationality of an approach to life where we think of ourselves as authoritatively controlling our choices by imaginatively projecting ourselves forward and considering possible subjective futures is undermined by our cognitive and epistemic limitations. If we attempt to fix the problem by adjusting our decision-theoretic models and eliminating the role for imagination and first personal assessment, the authenticity of our decision-making is undermined.

My target is the ordinary and plausible assumption that, when making big life choices, the ideal rational agent acts authentically by reflecting upon

how she wants to realize her future, and perhaps to realize herself as a certain kind of person, before she makes her choice. On this approach, you, as the agent, review your options and do a kind of cognitive modeling from the subjective perspective. You imaginatively project different possible futures for yourself, futures that stem from different possible choices you could make. When you are considering your options, roughly, you evaluate each option by mentally modeling what the outcome would be like, should you decide to choose that option. When you assess each outcome, you assign it a subjective value, and then you compare all the different values when you make your choice.[1]

Of course, when you decide, you also take into account any outside testimony and empirical evidence that bears on the question of what to do, but in the end, you evaluate the options by weighing the evidence and considering the expected value of each act from your own perspective. This process of simulation or imaginative acquaintance fits with how normative decision theory is supposed to provide a guide for how agents, if they are making rational decisions about their future, should proceed.

I argue that big life choices often concern transformative experiences, compromising our ability to assign subjective values to the radically new outcomes of these choices.[2] This in turn compromises our ability to use our preferred decision models to make these choices rationally. If you can't first-personally grasp your possible futures, including the future perspectives you'll have as the possible outcomes of your acts, you can't model and assess them for their subjective value. An important issue embedded in this argument concerns the personally transformative nature of the epistemically transformative experience: because you change dramatically, your preferences concerning the new outcomes can also change dramatically. If an experience irreversibly changes who you are, choosing to undergo it might make you care about very different things than you care for now. Who you are and what you care about may change when you strike out into the unknown. As a result, having the new experience can dramatically change how your post-experience self values the outcomes, including your valuing of your

---

1 Subjective values, as I understand them, are experientially grounded values attaching to lived experiences. 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. (There are other types of values, of course, such as moral and political values, that also come into play when we make big decisions.) Subjective values can be based on more than merely qualitative or sensory phenomenology: they may also include values arising from nonsensory phenomenological content. They are intended to include contentful features of rich, developed experiences that embed a range of mental states like beliefs, emotions, and desires. See Paul (2015) for further discussion.

2 As Jennifer Carr (2015) puts it, you can't metaphorically "see" your outcomes. One way you might find yourself in this situation is when you lack certain phenomenal concepts.

higher-order values, creating a problem for how you are to adjudicate between these different sets of preferences.

A complication thus arises: If, before you make the transformative choice, the dramatic future changes in yourself are first-personally inaccessible to you, *then from within your first personal perspective*, *you cannot "foresee" the ways your future self will change or foresee how your high order values will evolve*. You cannot first-personally foresee or understand who you'll become. In this way, the transformative tie between the epistemic and the personal forces you to face a first-personal version of a Kuhnian paradigm shift. With transformative decision-making, you face a problem of self-discovery and self-knowledge in a context of radical self change paired with radical epistemic change. Echoing a phrase of Bas van Fraassen's, there may be no first-personal view of the self that is invariant under these epistemic transformations.[3] This results in a new kind of existential problem for a model of rational choice based on maximizing one's expected utility and raises interpretive difficulties for relying on post-hoc testimony.

3. Becoming a Vampire

I illustrate the situation with vampires. Imagine that you have a one-time-only chance to become a vampire. With one swift, painless bite, you'll be permanently transformed into an elegant and fabulous creature of the night. As a member of the Undead, your life will be completely different. You'll experience a range of intense new sense experiences, you'll gain immortal strength, speed and power, and you'll look fantastic in everything you wear. You'll also need to drink animal blood (but not human blood) and avoid sunlight. Suppose that all of your friends, people whose interests, views and lives were similar to yours, have already decided to become vampires. And all of them tell you that they love it. They describe their new lives with unbridled enthusiasm, and encourage you to become a vampire too. They say things like: "I'd never go back, even if I could. Life has meaning and a sense of purpose now that it never had when I was human. It's amazing. But I can't really explain it to you, a mere human—you have to become a vampire yourself to know what it is like."

So, the question is, would you do it? And the trouble is, how could you possibly make an informed choice? For, after all, you cannot know what it is like to become a vampire until you become one, since the experience of becoming a vampire is transformative. That is, it is an experience that is radically new, such that you have to have it to know what it will be like for you, and when you undergo it, it changes your core personal preferences. In this situation, you can't possibly know what it would be like before you

[3] Bas van Fraassen, (1999).

try it. And you can't possibly know what you'd be missing if you didn't. So you can't rationally choose to do it, but nor can you rationally choose to avoid it, if you want to choose based on what you think it would be like to be a vampire.

The vampire case is structurally parallel to a version of Frank Jackson's case of Mary growing up in a black and white room, but where Mary is an ordinary person like you or me, rather than someone who knows all of complete science at the end of inquiry. The parallel here concerns whether Mary knows what she needs to know when she decides to leave her black and white room, if her choice is based on what she thinks seeing color will be like. (She should leave her room if seeing color will be like *this*, but she shouldn't leave it if seeing color will be like *that*.) In this situation, she cannot perform the sort of cognitive modeling that she needs to be able to perform in order to assign values to her possible outcomes, and thus to calculate her expected value for leaving her room. She lacks the ability to imaginatively acquaint herself with the future event, what it will be like for her to see color, in a way that can provide a guide for how she should choose.

We can see, pretty readily, how the puzzle arises in fictional cases like choosing whether to become a vampire or Mary's choice to leave her room. But there are real life cases of transformative choice. Such cases involve choices like a congenitally blind adult choosing to have a retinal operation or a person choosing to have her first child. In these cases, you also can't know what it will be like to have the characterizing experience before you have it, and if you choose to have it, it will change you significantly and irreversibly.

4. Choosing to have a Child

I elaborate the idea by developing the real-life example of the choice to have one's first child. Having a child often results in the transformative experience of gestating, producing, and becoming attached to your own child. At least in the ordinary case, if you are a 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 dramatic

hormonal changes and enter other new physiological states, all of which help to create the physical realizer for the intensely emotional phenomenology 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 experience associated with having one's first child. Raising a child is then a temporally extended process that extends, deepens, and complicates this relationship.

This unique type of experience often transforms people in the personal sense, and in the process, changes one's preferences. If the salient details of the nature of the transformative experience of producing and becoming cognitively and emotionally attached to your first child are epistemically inaccessible to you before you undergo the experience, then you cannot, from your first personal perspective, forecast the first-personal nature of the preference changes you may undergo, at least not in the relevant way. If so, the choice to have a child asks you to make a decision where you must choose between earlier and later selves at different times, with different sets of preferences, but where the earlier self lacks crucial information about the preferences and perspectives of the possible later selves, and thus cannot foresee, in the relevant first-personal sense, the self she is making herself into.

Once we see how epistemic and personal transformation work, it becomes apparent that many of life's biggest decisions can involve choices to have experiences that teach us things we cannot first-personally know about from any other source but the experience itself. With many big life 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 is more about discovery, and coming to terms with who we've made ourselves into via our choices, than about carefully executing a plan for self-realization.[4]

References

Carr, J. 2015. "Epistemic Expansions." *Res Philosophica*, Vol. 92, No. 2, 217–236.

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

Paul, L.A. 2015. *"Transformative Experience: Discussion and Replies.*" Res Philosophica, Vol. 92, No. 2.

van Fraassen, B. 1999. "How is Scientific Revolution/Conversion Possible?", *Proceedings of the American Catholic Philosophical Association* 73, 63–80.

[4] I thank Tyler Doggett, Kieran Healy and Enoch Lambert for comments.

Philosophy and
Phenomenological Research

Philosophy and Phenomenological Research
Vol. XCI No. 3, November 2015
doi: 10.1111/phpr.12240

Transformative Experience and Decision Theory*

RICHARD PETTIGREW
University of Bristol

I have never eaten Vegemite—should I try it? I currently have no children—should I apply to adopt a child? In each case, one might imagine, whichever choice I make, I can make it rationally by appealing to the principles of decision theory. Not always, says L. A. Paul. In *Transformative Experience*, Paul issues two challenges to decision theory based upon examples such as these (Paul, 2014). I will show how we might reformulate decision theory in the face of these challenges. Then I will consider the philosophical questions that remain after the challenges have been accommodated.

1. Deliberative and justificatory decision theory

The subject matter of decision theory is decision problems. We model a decision problem in which an agent must choose between a range of acts as follows:

(i) *Acts* $\mathcal{A}$ is a set of propositions each of which describes a different possible act that our agent might perform and states that she does in fact perform that act.

(ii) *Preferences* $\preceq$ is a preference ordering on the set of acts $\mathcal{A}$.

(iii) *States* $\mathcal{S}$ is a set of propositions each of which describes a different possible state of the world; they are mutually exclusive and exhaustive.

* I'm extremely grateful to Samir Okasha, Ben Levinstein, Jason Konek, Greg Wheeler, and Anya Farennikova for very helpful discussion of the proposals mooted in this paper; and, of course, to Laurie Paul for many wonderful, fascinating conversations. I was supported by an ERC Starting Researcher grant 'Epistemic Utility Theory: Foundations and Applications' during my work on this paper.

(iv) *Utilities u* is a function that takes a conjunction of the form $A \wedge S$, where A is in $\mathcal{A}$ and S is in $\mathcal{S}$, and returns the utility that the agent would obtain were that conjunction to hold: that is, the utility she would obtain if she were to perform that act in that state of the world. Conjunctions of the form $A \wedge S$ are called *outcomes relative to* $\mathcal{A}$ *and* $\mathcal{S}$.

(v) *Credences p* is a subjective probability function that gives (at least) the agent's credence in a state S in $\mathcal{S}$ under the subjunctive supposition of a proposition A in $\mathcal{A}$: this is the agent's credence that the world is in state S under the subjunctive supposition that she performs act A. This is often written $p(S||A)$.

(vi) *Expectations* We define the *expected utility of A relative to p and u* as follows: $E_p(u(A)) = \sum_{S \in \mathcal{S}} p(S||A)u(A \wedge S)$.

For many decision theorists, the only attitudes of the agent that have substantial psychological reality are their preferences. While these decision theorists accept that a rational agent has a utility function and a credence function, they contend that this says no more than that those functions together *represent* the agent's preference ordering in the sense that the ordering of the acts by their expected utility relative to those functions matches the ordering of the acts given by the agent's preference ordering: that is, $E_p(u(A)) \leq E_p(u(B)) \Longleftrightarrow A \preceq B$. They maintain that rationality imposes conditions on the agent's preference ordering and they show that any ordering that satisfies those conditions is representable by a unique probabilistic credence function and a utility function that is unique up to affine transformation. Thus, for these decision theorists, decision theory is concerned only with the rationality requirements that govern preferences, and the ways in which we can represent agents who satisfy those requirements. It is not concerned with how we might deliberate about which preferences to have nor with how we might justify those preferences. On this view, I can justify choosing an act by noting that I prefer it to all others. But, if you ask me to go further and justify those preferences, there is nothing I can say. I cannot appeal to my credences and utilities. I have those credences and utilities in virtue of having the preferences I have; so they cannot justify those preferences. We might call this the *preference-first conception of decision theory*.

However, these justificatory and deliberative tasks are clearly important. When you ask me to justify my preferences, I do not fall silent in the way predicted by the preference-first conception. Instead, I appeal to my credences and my utilities and the expected utility values for acts relative to them. That is, I treat credences and utilities as psychologically real and capable of justifying the preferences I have. Thus, there is an alternative account of decision

theory on which at least part of its job is to say how I can appeal to credences and utilities to justify my preferences or to set those preferences in the first place. This sort of decision theory can be used by an agent deliberating about a choice she must make; and it can be used by an agent after she has made her choice at the point at which she needs to justify it. When she is deliberating, she attends to her credences and her utilities, she uses them to calculate her expected utilities, and she chooses an act with maximal expected utility. When she is justifying her choice, she demonstrates that it maximises expected utility relative to her credences and utilities. We might call this the *deliberative conception of decision theory*. It is to this conception that L. A. Paul addresses her challenges. These challenges do not affect the preference-first conception: that is, they do not provide counterexamples to the conditions that rationality is taken to impose on an agent's preference ordering.

2. Epistemically transformative experience

The first challenge arises from the existence of epistemically transformative experiences (ETEs). Recall: ETEs are those that teach you something about the phenomenal character of a kind of experience that can only be learned by having an experience of that kind.

Suppose I face a choice between a range of alternative acts; and suppose that one of those acts has a possible outcome that involves an ETE. For instance, suppose I have never tried Vegemite. And suppose I must choose whether or not to accept a bet on a coin flip that gains me a piece of toast spread with Vegemite if the coin lands heads, and loses me £1 if it lands tails. Now, if I am to use decision theory to help me deliberate about what to do, or to justify my decision once it is made, I must at least have access to my credences over the states under the subjunctive supposition of the acts and to the utilities I assign to the outcomes of the available acts. After all, I must use them to calculate the expected utilities of the available acts. The problem is that, on one plausible formulation of the decision problem I face, I do not have access to these utilities. On this formulation, $\mathcal{A} = \{Bet, Don't\ bet\}$ and $\mathcal{S} = \{Heads, Tails\}$. Now, while I know the utility I assign to the outcome *Don't bet* (it is just the status quo) and the outcome $Bet \wedge Tails$ (it is just the disutility of losing £1), I don't have access to the utility I assign to $Bet \wedge Heads$. On this latter outcome, I gain a piece of toast spread with Vegemite, so a large part of what will determine the utility I assign to it is the phenomenal character of the experience of eating Vegemite; and this is something to which I lack access at the time the decision must be made, since the experience is an ETE.[1]

[1] In fact, Paul allows that an agent might make such a decision rationally if, for instance, she has a strong desire to have a new experience. I will be concerned with the cases in which there are not these other motivating desires.

This is Paul's first challenge to the deliberative conception of decision theory. A natural response is to reformulate the decision problem in question by reframing the uncertainty about the utility function as uncertainty about the world. I will call this the *redescription strategy*. Thus, instead of taking the set of possible states of the world to be $\mathcal{S}$, we instead take it to be $\mathcal{S}'$, a fine-graining of $\mathcal{S}$, where the fine-grained possible states of the world specify not only how the world is, but also what my utility function is over the outcomes relative to $\mathcal{A}$ and $\mathcal{S}$. Thus, let $u_1, \ldots, u_n$ be the utility functions I might have over the outcomes relative to $\mathcal{A}$ and $\mathcal{S}$. Then we reformulate the decision problem by taking the set of possible states of the world to be:

$$\mathcal{S}' = \{S \wedge \textit{My utility function is } u_i : S \in \mathcal{S} \text{ and } i = 1, \ldots, n\}$$

And we define the utility of the outcomes relative to $\mathcal{A}$ and $\mathcal{S}'$ as follows: for each i=1, ... , n,

$$u(A \wedge S \wedge \textit{My utility function is } u_i) = u_i(A \wedge S)$$

Doing this solves the original problem of epistemically inaccessible utilities. But it requires me to have credences over a new set of states of the world under the subjunctive supposition of the various available acts being performed. For instance, I must have a credence in (*Tails* $\wedge$ *My utility function is* u_i) under the subjunctive supposition of *Bet*. Is this a problem? I think not. Firstly, the utility hypotheses simply specify numerically the utility the agent obtains at each outcome. So you don't need to know the possible phenomenal characters of the experiences that you will have at each outcome in order to know what the possible utility hypotheses are in order to include them in $\mathcal{S}'$. You simply need to know the possible values of your utility function—and you do know that, since possible utility values are all real numbers. Secondly, once you've set out the various possible utility hypotheses, they are simply empirical hypotheses about which you can accumulate evidence in the usual way.

This, then, is the redescription strategy: faced with uncertainty about the utilities I assign to possible outcomes of available acts, I simply fine-grain the possible states of the world to include the various possible utility hypotheses about which I am uncertain; my utility function over the possible outcomes of the acts relative to these states of the world is then accessible to me and I can quantify my uncertainty about my utility function over the original outcomes using my credences over the new states (under the subjunctive supposition of having performed one of the various available acts). Does this answer Paul's challenge? Paul thinks not.

Paul's concern is that, when I am uncertain of my utility function, the redescription strategy exhorts me to make decisions based on my credences

over hypotheses about my own utility function; but, when an ETE is involved, these credences must be based in turn on statistical evidence that summarises what *other people* report about *their* utility functions; by assumption, none of the evidence can be about *my* utility function. Why is this a problem? We are accustomed to using statistical evidence based on facts about other people in order to make decisions about ourselves: I choose to exercise because statistical evidence based on data from other people suggests that it will improve my health. Why is it different when the statistical evidence bears on the hypotheses about my own utility function? Paul's concern is that, by basing our credences over the utility hypotheses solely on data about others and not about me, I threaten the *authenticity* of my decision.

> [W]e also want to choose authentically, that is, we want to choose in a way that is true to ourselves, in a way that involves our self as a reflective, deliberating person, choosing after assessing our preferences from our first-personal point of view and then living with the results. (Paul, 2014, 128)

Thus, Paul's worry stems from existentialist concerns. Elsewhere, she compares making decisions based on statistical evidence that summarises the experiences of others to making decisions based on the dictates of a putative morality whose values you do not fully internalise as your own, or based on the dictates of a group of which you are a member but with whom you share few values. The latter cases are, of course, the existentialist's archetypes of inauthentic action. And the existentialist's concern is that they alienate you from your decisions.

I contend that these cases are not analogous. While it may be true that I will be alienated from my decision if I simply defer to the dictates of morality or my social group and studiously ignore my own utilities, it is not true in the cases of decisions made on the basis of credences that quantify uncertainty about my own utility function. The reason is that, although in the latter case my evidence does not include facts about my own utility function, it nonetheless provides evidence that supports credences in propositions concerning my own utility function. Indeed, my purpose in collecting statistical evidence based on the utilities of others is precisely to try to overcome the epistemic barrier to knowing my own utilities in the case of ETEs.

Suppose I read a survey that reports that, amongst a large randomly selected sample from a population of which I am a member, 70% of respondents assigned a high utility to being a parent, while 30% assigned a low utility. In the absence of other evidence, I might reasonably assign a credence of 70% that *I* will assign high utility to being a parent. But notice what would happen were I then to learn that all members of this sample population had drastically different utilities from mine in every other sphere.

In this situation, I would reasonably abandon the credences I had assigned, because I would have learned that the utilities assigned by this population were not a good indication of my own. What this shows is that, when I use statistical evidence to set my credences about my own utilities and then use these credences to make a decision, I am not simply deferring to the majority opinion in a way that renders my decision inauthentic. Rather, I am using the opinions of others as evidence about my own utility function. Put another way: I attend to the opinions of others not because I wish to follow the majority decision, but rather because I want to find out about myself.

Thus, while facts about my utilities may not be contained in the evidence, my utilities are nonetheless a key ingredient in my deliberation in a way that is very different from the case in which I simply defer completely to morality or to the mores of my social group. There is a difference between being certain of my utilities and ignoring them in favour of acting in accordance with moral laws or group decisions, on the one hand, and being uncertain of my utilities and using all available evidence in order to predict them and then acting in accordance with my best possible predictions about them, on the other. One way to see the difference is to note that, when I make a decision based purely on the demands of morality, the utility function that I use in the expected utility calculation is not my own—it is rather an objective value function that encodes the demands of morality. Thus, when I choose, I do not choose the action that I expect to have highest utility by my lights; I choose the act that I expect to have highest utility by the lights of objective morality. That is alienating. The same is true if I use the utility function of my social group when I make my decision, rather than my own utility function. Again, that is alienating. On the other hand, when I choose between acts that might give rise to ETEs, I choose the act that I expect to have highest utility by *my* lights, even though I'm uncertain about what those lights are. In this case, it seems to me, the decision need not be alienating.

3. Personally transformative experience

Paul's second challenge arises from the existence of personally transformative experiences (PTEs). Recall: PTEs are those that lead you to change what you value and to what extent.

Suppose I face a choice between a range of acts; and suppose that one of those acts has a possible outcome that involves a PTE. For instance, suppose I must choose whether or not to become a parent for the first time. Thus, in this decision problem, $\mathcal{A} = \{Adopt, Don't\ adopt\}$. And let's assume that there is no uncertainty in the world—if I choose to adopt, I will adopt; if I choose not to, I won't. Thus, on a natural formulation of the decision problem, there is just one possible state of the world in $\mathcal{S}$. Thus,

the outcomes relative to $\mathcal{A}$ and $\mathcal{S}$ are just *Adopt* and *Don't adopt*. In the former, I become a parent; in the latter, I don't. And let us say, very crudely and ignoring the fact that becoming a parent may be an ETE as well as a PTE, that I know that, if I become a parent, I will come to value time spent with friends less and time spent with family more. What's more, I know that the outcome *Adopt* involves a lot more time with family and a lot less with friends, while this is reversed in *Don't adopt*. Thus, my current utility for *Adopt* is lower than for *Don't adopt*; and this is true of my future utilities as well in the outcome in which I don't adopt. But, if I do adopt, then my future utilities will be the reverse of my current utilities: I will value the family time entailed by *Adopt* more than the time with friends entailed by *Don't adopt*. How, then, am I to make this choice? This is Paul's second challenge to decision theory: it is ambiguous in cases in which my utilities change over time.[2]

A natural response is this. First, introduce the notion of a *local utility function*: my local utility function at a given time is the function that measures how much I value outcomes at that time. Next, let us demand that, as in the previous section, the possible states of the world are fine-grained enough to include a specification of the relevant facts about my utilities: in this case, where my utilities may change over time, that will include a specification of my local utility function for each time during my life in that state of the world. Now, let us fix attention on a particular outcome that results from conjoining an act with a state of the world that is fine-grained in that way: thus, this outcome will include a specification A of the act that is performed and a specification S of how the world is (call the conjunction $A \wedge S$ its *worldly component*); and it will include a specification of my local utility functions (call this its *utility component*). Let us simplify and assume that, in this outcome, there are just finitely many moments in my life, $t_1, \ldots, t_n$. And let $lu_{t_1}, \ldots, lu_{t_n}$ be my local utility functions at those moments. Thus, the outcome in question is:

$$A \wedge S \wedge \bigwedge_{i=1}^{n} \textit{My local utility at } t_i \textit{ is } lu_{t_i}$$

Then we might say that the utility function to which I appeal when I make a decision at one of those moments (say t_i) is my *global utility function at* t_i (which we write u_{t_i}), where the utility that u_{t_i} assigns to the outcome just described in some way aggregates the local utilities assigned to the worldly component of that outcome ($A \wedge S$) by each of the local

[2] Again, Paul allows that such a decision may be made rationally if the agent has some other motivating desire, such as a desire to have an heir. Again, I will be restricting attention to the case in which there is no such other motivating desire.

utility functions $lu_{t_1}, \ldots, lu_{t_n}$. The natural means of aggregation in this case is the weighted sum. Thus, my global utility at t_i for the outcome described above is determined by a series of weights $\beta_{t_1}, \ldots, \beta_{t_n} \geq 0$. These specify how much each of my local utilities for the worldly portion of that outcome contribute to the global utility for the whole outcome, which is the utility I use when I make a decision or justify it once made. So, for each j=1, . . . , n, the weight β_{t_j} specifies the extent to which I take into account the values I had or have or will have at time t_j. Thus:

$$u_{t_i}(A \wedge S \wedge \bigwedge_{i=1}^{n} \textit{My local utility at } t_i \textit{ is } lu_{t_i}) = \sum_{i=1}^{n} \beta_{t_i} lu_{t_i}(A \wedge S)$$

This solves the formal problem. That is, it describes a formal framework that can be used to make decisions in cases in which (local) utilities might change. Yet, in doing so, it does little more than provide a framework in which we can pose the difficult philosophical questions precisely. These questions mainly concern the constraints that rationality places on the weights $\beta_{t_1}, \ldots, \beta_{t_n}$ that an agent uses.

The most important such question, it seems to me, is whether an agent should assign greater weightings to her other local utility functions the more similar they are to her present local utility function. On the one hand, this would permit her to heavily discount the opinion of a future self that she knows will have shifted significantly from her current self in its moral opinions. For instance, when deciding how to vote, it would release her current left-wing self from the obligation to place much weight on the values of her possible right-wing future self. It also answers another possible existentialist concern: if an agent does not set her weights in this way, then she will end up deferring in large part to value judgments that are not currently her own, thereby rendering her decision inauthentic. Having said that, it is unclear to me whether the existentialist motivation for authenticity militates against deference to one's own future value judgements or merely deference to the value judgements of others or to those of an objective morality. On the other hand, such a weighting can lead to a certain conservatism as well as an unappealing chauvinism or parochialism about one's current values, especially if one thinks that a future local utility function, while very different from one's own, is nonetheless a permissible local utility function to have. If I assign little weight to my future local utility function as a parent because it lies far from my current utility function as a non-parent, and I choose not to adopt on that basis, something has gone wrong. It may be thought that second-order utilities can help solve this problem, but it is worth noting that my second-order utilities can change as well, and we will then need third-order utilities to adjudicate cases in which those might change; and so on. Also, in the case of

choosing to become a parent, for instance, it seems that my second-order utilities will change with my first-order utilities: currently, I value valuing friends above family; and, if I adopt, I will value valuing family above friends. Thus, even if we use second-order utilities to assign weights, we'll be left with the same chauvinism about current values that we wished to escape.

Thus, we have seen how deliberative decision theory may be reformulated in order to avoid Paul's challenges, at least formally. However, those challenges do raise profound philosophical questions about the status of decisions we make using that theory. In particular, they raise a number of important questions concerning the extent to which such decision making is compatible with the claims of existentialism.

References

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

Philosophy and
Phenomenological Research

Philosophy and Phenomenological Research
Vol. XCI No. 3, November 2015
doi: 10.1111/phpr.12242

What You Can Expect When You Don't Want to be Expecting

ELIZABETH BARNES
University of Virgina

Transformative Experience is a rich, insightful, compelling book. LA Paul persuasively argues that our standard way of thinking about major life choices (and some minor ones too) is inadequate, because it fails to take into account the subjective phenomenal values of lived experiences. When deciding whether to do something, we need to assess how good the outcome will be for us. But Paul argues that in many such cases, we simply don't have enough information to do this. And that's because we don't have information about the subjective phenomenal value of the experience we're considering—that is, we don't know *what it's like* (for us) to have that experience. This means our decision is inherently under-informed. We can't decide how to assign values to possible outcomes (undergoing the experience or failing to undergo the experience) because we don't have a complete picture of what those values really are.

I find much of what Paul argues in the book completely persuasive. What I want to argue here is that some of her points are not as widely generalizable as she takes them to be. Specifically, I'm going to argue that there are plenty of cases in which we don't know what an experience is like, but we nevertheless can rationally choose to avoid that experience based on projected outcomes. And that's because we can rationally choose based on the belief that *whatever* that experience is like, we're fairly sure it's something we don't want.

1. Some Brief Autobiography

I have never wanted kids. Neither has my partner. And it isn't that we don't want kids in a 'well, despite the appeal of having kids, all things considered it is probably best for our careers and our overall wellbeing that we stay

child-free' kind of way. We don't want kids in the sense that we've never had the slightest desire to have them, and other people's strong desire to have them is somewhat mystifying to us.

It's not that we don't like kids. We're fond of our friends' kids and love our nieces very much. But spending time with kids has always felt a bit like going to the circus: entertaining—in a loud, boisterous sort of way—but not something we'd seriously consider joining ourselves.

Given my and my partner's preferences, deciding not to have kids seemed pretty rational. Indeed, it seemed like one of the most rational, clearest choices we've ever made. So it would be a surprising result, to say the least, if it turned out not to have been rational at all.

2. Choosing Whether To Have Children

If Paul is right, however, our choice not to have children *wasn't* rational. Because we don't know what it's like to have kids, we can't rationally choose to abstain from doing so. Or, more carefully, we can't rationally choose to abstain from doing so in the standard way, which is by thinking about our potential kid-having future, thinking about our potential not-kid-having-future, and then deciding which we prefer. We can't make such a choice rationally, Paul argues, because the experience of having a child is *transformative*.

Paul distinguishes between two types of transformative experiences. *Epistemically* transformative experiences give you access to new sorts of phenomenological information that was previously unavailable to you. *Personally* transformative experiences are those that fundamentally alter your beliefs, preferences, or sense of self. Epistemically transformative experiences needn't also be personally transformative. Trying Vegemite for the first time is epistemically transformative—there is just no way to know *what it's like* to taste Vegemite until you taste it—but for most people trying Vegemite for the first time isn't something that profoundly shapes who they are as a person. But many of our major life experiences, Paul argues, are both personally and epistemically transformative.

Having a child is one such experience, according to Paul. It's epistemically transformative because you can't know what it's like to have a child until you have one. And it's personally transformative because having a child reshapes your preferences, your desires, and even your own sense of who you are in radical ways.[1] And because of this, Paul argues, we can't rationally choose whether to have a child (or at least can't do so by projecting child-having and non-child-having outcomes and comparing them).

---

[1] According to Paul, 'If an experience changes you enough to substantially change your point of view, thus substantially revising your core preferences or revising how you experience being yourself, it is a personally transformative experience.' (p. 16)

Paul bases her argument on what she labels the subjective values of experiences. According to Paul, subjective values are:

> [V]alues of experiences... 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 arise from nonsensory phenomenological features of experiences, especially rich, developed experiences that embed a range of mental states, including beliefs, emotions, and desires. (p. 12)

These subjective values are an important part of how good (or bad) an experience is for us. And this is what creates the problem for the standard model of rational decision-making in the case of transformative experience. In deciding whether or not to undergo an experience, we need to assign values to the relevant outcomes—and so we need to have a reasonably informed sense of how good or bad an outcome will be for us. But a crucial aspect of how good or bad an outcome will be for us is its subjective value. Yet, Paul argues, in the case of transformative experiences, this is something we cannot know (or even predict with accuracy):

> Subjective values, grounded by what it is like to have lived experiences, are first-personal values... Given this, you must have had the right kind of experience to know a subjective value, because you must know what an experience of that type is like to know its [subjective] value—for example, you must experience color before you can know the subjective value of what it's like to see color. (p. 13-14)

Deciding whether to have a child is, in part, deciding whether to undergo a transformative experience. According to Paul:

> When you face a transformative choice, that is, a choice of whether to undergo an epistemically and personally transformative experience, you face a certain kind of ignorance: ignorance about what it will be like to undergo the experience and ignorance about how the experience will change you. Thus, you face a certain kind of ignorance about what your future will be like. (p. 31-2)

And this ignorance—ignorance of the very the things we care about the most—renders a fully rational decision impossible. You can't make a rational choice simply because you're dramatically—and inherently—under-informed.

So much for deciding to have a child. Why should ignorance of the subjective value of having a child prevent us from rationally deciding *not* to have a child? If we've been child-free up to now, we know what it's like to

be child-free. And while we don't know exactly what it will be like to continue to be child-free—simply insofar as no one can really know what the future will be like—we can at least make an educated guess.

The problem, Paul argues, arises from the transformative nature of having a child, particularly the personal transformativeness. Because having a child both gives you new information and (in some cases radically) alters your preferences and your sense of self in a fundamental way, those who remain child-free don't know what they're missing. As a result, in choosing to remain child-free they can't adequately compare a child-free future with a child-having future. Paul writes:

> When we face a choice like this, we can't know what our lives will be like until we've undergone the new experience, but if we don't undergo the experience, we won't know what we are missing. And, further, many of these new and unknown experiences are life-changing or dramatically personally transformative. So not only must you make the choice without knowing what it will be like if you choose to have the new experience, but the choice is big, and you know it is big. You know that undergoing the experience will change what it is like for you to live your life, and perhaps even change what it is like to *be* you, deeply and fundamentally. (p. 3)

To rationally choose whether to have a child, according to Paul, you need to be able to compare the child-having future with the non-child-having future—which is exactly what you can't do if you don't know the subjective values of the child-having future. And so, Paul argues:

> you cannot rationally choose to have the experience, nor can you rationally choose to avoid it, to the extent that your choice is based on your assessments of what the experience would be like and what this would imply about your future lived experience. (p. 18-19)

Importantly, Paul doesn't take the upshot of this to be that rational choices about having children (or other transformative experiences) are impossible. She just thinks you can't make rational choices *based on comparing projected subjective outcomes*—which is, of course, what we often take ourselves to be doing in this situation, and what standard models of decision theory represent us as doing. We could rationally choose whether to have a child for objective reasons that have nothing to do with the subjective values of the experience. We could choose to have a child if we needed an heir, for example, or we could choose not to have a child if we knew we couldn't afford it.

But Paul also argues that we can make rational choices in cases of transformative experience simply by choosing whether we want a new experience (whether we want 'revelation'). In the child-having case, we don't

choose whether to have a child based on the projected values of having a child compared to not having one. Instead, we choose based on whether we want a new experience—whether we want our lives to stay roughly the same or whether we want to find out who we'd become as parents:

> In general, then, the 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. If you choose to undergo a transformative experience and its outcomes, you choose the experience for the sake of discovery itself. (p. 120)

But this is, Paul readily admits, not the decision procedure most of us follow, nor the type of choice most of us take ourselves to be making. Her proposed solution is, in this sense, radically revisionary. And if she is right, it means that choices like my and my partner's—the choices of the resolutely child-free to remain so, based on the simple thought that we just don't want kids—aren't rational, at least if they're made (as they so often are) on the basis of projected subjective values for outcomes. But, as I'm going to argue, I don't think Paul is right about this.

3. Swimming With Sharks

Notably, Paul doesn't think that *all* projected-value decisions involving transformative experiences are irrational. We probably don't know what it's like to get eaten by a shark, she grants, but we can still rationally try to avoid it. And, more importantly, we can rationally choose to avoid it for reasons stronger than wanting to avoid revelation.

The case of getting eaten by a shark no doubt introduces some complications, though. Ostensibly, getting eaten by a shark is lethal. It might, in general, be rational to preserve your life even if you don't know what it's like to die, and Paul's view can accommodate this (since Paul can plausibly grant that you needn't know the subjective values of continuing to exist to prefer that outcome to its alternative). So let's consider, rather than getting eaten by a shark, getting your leg chewed off at the knee by a shark. That's not lethal (let's assume you have access to medical care that will stop the bleeding) but it's something most of us want to avoid, even if we don't know what it's like.

Paul thinks—I'm assuming, based on the discussion of the shark case—that you can rationally choose to avoid getting your leg chewed off by a shark, even if you don't know what it's like to get your leg chewed off by a shark. And that's because you've got a decent amount of evidence which suggests that what it's like to get your leg chewed off by a shark is a what it's like to be avoided. You are pretty sure that it would be painful, for one

thing, even if you don't know in exactly what way, and so insofar as you don't like painful things you have good reason to suppose that you won't like this particular painful thing. Moreover, there aren't lots of surfers proclaiming the transformative magic of shark attacks. Nor do your friends and family try to persuade you that, strange as it may seem to you now, you will really will enjoy getting attacked by a shark once you try it.

And so, Paul argues, in the case of getting attacked by a shark, you can engage in a kind of projective forecasting, even though you don't know what it's like to be attacked by a shark. And that's because you have good reason to suppose that *whatever* it's like to be attacked by a shark is something you want to avoid. That is, you have good reason to think you will prefer not getting attacked by a shark to being attacked by a shark, even though you don't have access to the phenomenal value of getting attacked by a shark. And so it's rational, Paul argues, to avoid swimming in shark-infested waters, even if you don't have full information about the phenomenological values this could lead to.

Paul maintains, however, that this concession is a minor one. Indeed, she states that:

> I am assuming, here and throughout, that cases like the shark-eating case are outside of the scope of this discussion. (p. 32, note 39)

But in what follows, I argue that many choices involving potentially transformative experiences have more in common with the shark case than Paul admits.

4. Thought Police And Assimilation

When considering whether to get attacked by a shark, you're considering whether to undergo an experience that you're pretty sure, given your current preferences, you'd rather not undergo. It's also the case that you don't know of anyone who reports back from getting attacked by a shark and says it was amazing. Other experiences, however, meet the first condition without meeting the latter condition. That is, some people *do* report back and say those experiences are amazing, but you're still pretty sure your current preferences make it such that you'd rather not undergo them. What I want to argue is that we can rationally choose to avoid this type of experience.[2]

[2] For an interestingly different way of arguing for a very similar conclusion, see Sharadin, Nate (forthcoming) 'How You can Reasonably Form Expectations When You're Expecting'. *Res Philosophica.* Sharadin argues for the presence of 'linking principles' between current preferences and likely expected outcomes, claiming that in many cases you can know whether an experience is (likely to) have positive or negative valence for you even if you can't know exactly what that experience will be like (including *how* positive or *how* negative it will be).

In *1984,* Winston famously attempts to avoid being captured by the Thought Police. In this scenario, Winston is deciding whether to undergo a transformative experience. If he is captured by the Thought Police, it will be personally transformative: it will reshape his desires and his self-conception. He also has plenty of evidence that those who are captured and re-educated end up being very happy that they've undergone the process. And yet pre-Room 101 Winston strongly prefers not to be captured by the Thought Police, despite not knowing what it would be like to be re-educated, and despite the reported testimony of many that being re-educated is wonderful and fulfilling.

Pre-Room 101 Winston has many preferences, and among these are strong preferences not to live the kind of life that a person who is re-educated by the Thought Police will live. He wants to find out the truth. He wants to continue his relationship with Julia. He doesn't want to be slavishly devoted to Big Brother. He doesn't know what it's like to be re-educated by the Thought Police, but he knows that, whatever it's like, it's something he'd prefer to avoid.

Winston also has very good reason, of course, to suppose that all these preferences would change should he be captured by the Thought Police and re-educated. But, crucially, the fact that he knows his preferences would change doesn't affect how strongly he wishes to avoid being captured. If anything, it strengthens his desire to avoid capture. These preferences are part of who he is, and they matter to him. He quite simply doesn't want to be the kind of person who doesn't care about the things he currently cares about. Such a person—and such preferences—are *alien* to him. That is, they seem utterly foreign and bizarre (and unwanted) to his current self-conception and his current preferences. Nor is this sense of alienation lessened by the his knowledge that, were he to become such a person, he would likely not regret it. (Indeed, once he is captured and re-educated he ends content in the love of Big Brother and doesn't want to go back to his previous state. But that's something that makes his story tragic, not something that makes his previous desires to avoid capture irrational.)

Similarly, in *Star Trek: The Next Generation* Captain Picard wants to avoid being assimilated by the Borg.[3] Being assimilated by the Borg is a transformative experience. You discover what its like to be a member of a hivemind. Your desires and sense of self are completely altered by the will of the Borg collective. And so on. Moreover, there's good reason to

[3] I'm going to assume here, for simplicity, that there's still something that it's like to be you, as an individual consciousness, after you've been assimilated by the Borg. That's not clear from the fiction. I'm also considering Picard's choice before he is captured by the Borg. After being captured, he at least seems to have some sense of what it's like to be Borg.

suppose that this is a transformative experience that alters preferences. Borg don't regret becoming Borg. They think everyone should become Borg. They think being Borg is the best way to be.

And yet, despite not knowing what it's like to be Borg, it seems fair to say that Picard has reason to suppose that whatever it's like to be Borg is something he wants to avoid. He values autonomy and freewill. He values individuality and independence of thought. He values non-violence and the Prime Directive.[4] And he knows that becoming Borg would contravene all these values. Becoming Borg would also, of course, change his preferences such that he no longer value these things. But given what he values now, he doesn't want to become Borg, despite knowing that his preferences would change should he become Borg. Again, becoming Borg would change his preferences in a way that is *alien* (pun definitely intended) to him. Given who he is now—what we wants, what he desires, what he values—the preferences of post-assimilation Picard are utterly foreign to pre-assimilation Picard's sense of self.

Being captured by the Thought Police or assimilated by the Borg are unlike being attacked by a shark for the simple reason that the former involve predictable preference change in a way the latter does not. There are lots of people who value having been captured by the Thought Police or assimilated by the Borg, and assure you that you will too once it happens to you. The same isn't true for shark attacks.

But based on Winston's *current preferences*, getting captured by the Thought Police is like getting attacked by a shark. And based on Picard's *current preferences*, getting assimilated by the Borg is like getting attacked by a shark. Neither Winston nor Picard know what these things are like, but they have good reason to suppose that—whatever they're like—they're something they really don't want, given their values, hopes, dreams, and desires.

5. Rationally Preferring To Remain As You Are

Paul, of course, can say that Winston's choice to avoid the Thought Police and Picard's choice to run from the Borg *are* rational. But—assuming both choices are based on subjective beliefs and desires—they can only be rational insofar as they are choices to avoid 'revelation'. Winston's choice is rational to the extent that he is choosing not to undergo a new experience. And, likewise, Picard's choice is rational insofar as he is choosing to avoid finding out new phenomenological information (choosing to avoid finding out what it would be like to be Borg).

---

[4] It's not actually clear whether Picard does value the Prime Directive, rather than some specific applications of it, given how often he breaks it. Thanks to Prof. Heather Logue for discussion on this crucial point.

But this solution seems to (somewhat woefully) misdescribe the cases. Winston doesn't try to avoid being captured by the Thought Police because he doesn't want a new experience. He tries to avoid being captured by the Thought Police because *he doesn't want to be captured by the Thought Police*. And he doesn't want that experience because of what it entails—because he knows that being captured by the Thought Police will mean an end to his quest to find out what's really going on, an end to his resistance to Big Brother, and an end to his love for Julia. Winston doesn't need direct phenomenological awareness of what it's like to be re-educated by the Thought Police to know that—whatever it's like—it isn't what he wants.

Moreover, Paul's proposed solution seems to make transformative choice too coarse-grained. When we are faced with a transformative choice, the thought goes, we must choose to have a new experience or to forgo it, with the idea that the new experience itself is something of a phenomenological black box. And that's because its character—its *what it's likeness*—is hidden. But how, then, are we to explain choices between different sorts of transformations? Captain Picard does not want to be assimilated by the Borg: he does not want *that particular* new experience. But he is not at all averse to new and unknown experiences in general. Indeed, he seems to purposefully seek them out ('to boldly go where no one has gone before'). More often than not, Picard will choose revelation—he will choose to find out what a new experience is like, simply for the thrill of discovery. But he wants to avoid being assimilated by the Borg.

Paul grants that in some special cases—like getting attacked by a shark—your lack of *what it's like* knowledge doesn't impede your ability to make rational decisions based on projected outcomes. Yes, you're ignorant of (some of) the relevant subjective values. But you have enough other knowledge (your fear of sharks, your dislike of pain, your fondness for your leg, etc.) to make it the case that such ignorance doesn't interfere with your decision-making. What I'm claiming is that such cases aren't as rare as Paul seems to think.

In Paul's view, a predictable preference change seems to be enough to prevent standard rational decision-making:

> If you are to choose rationally, the preferences you have right now seem to have priority, such that to choose rationally you must choose in accordance with the preferences you have now. But your pre-experience preferences are dramatically incomplete, due to the epistemic inaccessibility of the values of the radically new outcomes. Under such circumstances, why should you be biased towards the preferences of your present self, the epistemically impoverished self? (p. 49)

But this seems too strong. Cases like Winston's and Picard's are those in which, *given their current preferences,* a particular transformative

experience is something they clearly don't want, even if they don't know what it would be like. True, if they underwent the transformative experience their preference would change. But, crucially, the way in which their preferences would change is itself a violation of their current preferences and values.

6. Character Planning Is Rational

What both Winston and Picard are engaged in is a type of *character-planning*. Despite not knowing what it would be like to be captured by the Thought Police or assimilated by the Borg, they know that these events would violate their preferences, their values, and even their sense of self. They also know, of course, that their preferences, their values, and even their sense of self would change if they underwent these experiences. But they would change in ways that are themselves violations of their preferences, their values, and their sense of self. That is, their preferences would change in a way that is alien to their current sense of self. Winston does not want to become the kind of person who is sycophantically devoted to Big Brother—developing those kinds of preferences is abhorrent to him as he is now. Picard does not want to become the kind of person who values the conquest and assimilation of other races—developing those kinds of preferences is abhorrent to him as he is now. Winston's choice to avoid the Thought Police and Picard's choice to avoid the Borg are both, I contend, completely rational when seen in this light. They are choosing to preserve their character, to continue to value what they value and pursue the projects they want to pursue.

But what of more mundane, ordinary cases like choosing *not* to have children? When my partner and I chose not to have children, it was for the simple reason that we didn't want them (and couldn't really imagine wanting them). Having children was a clear violation of our preferences and desires, even though we could predict that if we had a child our preferences would change. Paul argues that:

> the prospective parent who places a high value on remaining child-free faces an even worse dilemma, because, while friends and relatives tend to testify to their satisfaction after becoming parents, the empirical work suggests that well-being plummets. In this case, the evidence from testimony of friends and relatives suggests that the reluctant prospective parent should prioritize the preferences she'd have after becoming a parent, whereas the scientific evidence suggests she should prioritize the preferences she has before becoming a parent. The problem, then, is that there is no clearly correct decision-theoretic rule about which set of preferences to prefer at this level: those of the current, decision-making, child-free self or those of the future self who has become a parent. (p. 117)

But here I suggest that Paul—somewhat ironically, given how much emphasis she places of the first-person perspective in decision making—is demanding an overly objective stance from the would-be parent. On Paul's model, it seems as though we should be able to step back to a neutral, preference-free perspective and evaluate whose preferences should matter more: those of the current, child-free person or those of the future parent. But why should this kind of neutral evaluation be required for rational decision-making?

Winston doesn't need to weigh the potential preferences of post-Room 101 Winston when deciding to avoid the Thought Police. Picard doesn't need to weigh the preferences of post-assimilation Picard in deciding to avoid the Borg. In both cases, those preferences are alien to who they are now—that is, to the people currently making the choices. If such preference matter at all in Winston and Picard's choices, they matter only insofar as both Winston and Picard value never becoming the people who have such preferences.

Similarly, I don't think that I need to weigh the preferences of post-baby Elizabeth in deciding to take birth control. Winston and Picard do not want their preferences to change in ways which are completely alien to them, and which violate their current sense of self. Likewise for me, though admittedly to a milder degree. Having always actively desired not to have children, the preferences of post-baby me are completely alien—they violate both my current preferences and my sense of self. I can't imagine having such preferences, and having such preferences would be in tension with things about myself that I currently place great value on. It's rational for me to avoid such an experience, and such a change in preferences, even if I don't know what it's like to have the experience, or to have my preferences change in such a way. Whatever having kids is like, it's something that I can rationally predict that I don't want *given who I am now*.

When I choose not to have children, it isn't simply a choice to avoid 'revelation' or a choice made 'for the sake [or lack of] discovery itself'. It is, quite simply, a choice not to have children because I know I don't want them. And so it's a choice based on projected outcomes. I can project, given my current desires and presences, that having kids is a less good outcome than not having them, even though I'm ignorant of some of the relevant subjective values. In making such a choice—in assigning values to the relevant outcomes in this way—I'm engaged in a type of character planning. I'm choosing to value the person I am now, and I'm not placing weight on potential preferences that are alien—incomprehensible, foreign, in tension with my current preferences and values—to who I am now. I'm making a choice based on projected outcomes—a choice that values a non-kid-having future over a kid-having future—and evaluating those outcomes based on my actual, current preferences. That ought to be rational, I submit,

even if my preferences could change. In a choice like this, I'm saying whatever it's like to have kids (which I grant I don't know), I can reasonably assume it's something I don't want, given who I am now and given that who I am now is part of what I value.

And so, in a nutshell, what I'm arguing is this. For many of the happily child-free, having kids is kind of like getting assimilated by the Borg. We don't know what it's like. But we can still rationally choose to avoid it.

Philosophy and
Phenomenological Research

Philosophy and Phenomenological Research
Vol. XCI No. 3, November 2015
doi: 10.1111/phpr.12241

L. A. Paul's *Transformative Experience*

JOHN CAMPBELL
University of California, Berkeley

1. Authenticity as Requiring Imagination

Authentic decision-making requires imagining how your life will be if you decide one way or another. But what about cases in which your decision may affect your life in ways that you currently find unimaginable? In these cases, on the face of it, authentic decision-making is impossible. This is, I think, the main line of argument in L. A. Paul's brilliant *Transformative Experience*. It depends on the idea that 'authentic' decision-making must rely on imaginative understanding. But what conception of 'imaginative understanding' do we need to make this idea work?

The psychiatrist Karl Jaspers famously contrasted empathetic or imaginative understanding of patients with scientific explanation. On the one hand the psychiatrist can understand the patient's subjectivity, and grasp 'genetically by empathy' how one psychic event emerges from another. This may not always be possible; for example Jaspers thought that delusions were 'un-understandable' in this way. In the case of such patients, all we can achieve is 'scientific' explanation, in which we use our knowledge, from repeated experience, that 'a number of phenomena are regularly linked together', and on this basis, '*we explain causally*' (Jaspers 1913/1997).

Suppose we consider a psychiatrist confronted with a patient in a clinical setting. Suppose the psychiatrist does have a good empathetic or imaginative understanding of the patient. Jaspers doesn't put this point this way, but it seems evident that the psychiatrist here can achieve a certain 'authenticity' in their responses to the patient. The therapist's imaginative understanding of the patient can directly drive their emotional engagement with the patient. In contrast, consider a psychiatrist who has only a 'scientific' understanding of the

patient. The therapist has, we may suppose, ticked off a certain number of critical checkboxes relating to the patient's symptoms. Now she looks up a manual saying how to respond to a patient with those symptoms: what treatment is appropriate, what points to look out for in talking with the patient. Perhaps she also relies on knowledge of wide range of research articles, previous clinical encounters with other patients, and so on.

I said that there's a sense in which the emotional engagement driven by imaginative understanding is more 'authentic'. Does this notion of authenticity have anything to do with the notion of authenticity as 'being true to oneself'? I think it does: it's just that the connection between your imaginative understanding of the other person and your emotional engagement is made internally; it doesn't go by way of external connections such as books telling you how one should respond to such a patient.

This can really matter in practice. Consider a therapist confronted with a patient who has a huge problem with gambling. Let's suppose she has a lot of imaginative insight. She is appalled at what he's doing to his family and to himself. But her disgust is not the response mandated by the books. She tries to conform her response to him to what the books recommend. But the very fact that that's what she is doing means that there's a sense in which her response is not 'authentic'. This inauthenticity, if sensed by the patient, may in turn make a working relationship impossible.

Another, simpler way to see the connection between imaginative understanding and authenticity is to reflect on how you know that someone else's physical pain is a bad thing. It may be that you know perfectly well what pain is, and know that some other people have it, but don't particularly see why that should concern you, even in cases where the other person is right there in front of you. You might find a guru or spiritual instructor who tells you, and occasionally reminds you, that you should be concerned about other people who're feeling physical pain. But this is very different to the usual case, in which an imaginative understanding of someone else's physical pain immediately drives a compassionate response to them. A compassionate response in the ordinary case has an authenticity that the externally driven response does not.

2. What Concept of Imagination?

What concept of imagination do we need, to recognize the connection between imagination and authenticity? On the conception I recommend, imagination is (a) *de re*, or externalist, in the sense that it involves imagining how the external, mind-independent environment is, as well as the mental life that is located in that environment, and (b) affective, in that your exercise of imagination directly engages your emotions and actions, without any need for further reflection.

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. On this conception, imagination provides one only with knowledge of the internal mental life, which could be as it is however things are in the external environment, and whose exercise typically requires no immediate engagement of one's emotion and action. That's the conception suggested by reflection on Nagel's bat, for example, which is what made it vivid to most contemporary philosophers that imagination yields something not available by other means, or Jackson's Mary (Nagel 2002, Jackson 2002).

My main reservation about *Transformative Experience* is that although it's a great insight to connect, as Paul does, authenticity and imagination, that insight is undermined if you work with the internalist conception of imagination, as something whose function is to yield affectless knowledge of the internal mental life. On this picture, one's preferences with regard to the mental life are something external to the imaginative exercise itself. The function of the imaginative exercise is merely to provide the data on which one's preferences can operate.

Suppose you see an old friend who's been reduced to living in poverty. Your empathetic grasp of how things currently are for them, your concern and distress, can have all the authenticity you like. It seems absurd to represent your imaginative understanding here as confined in its scope to the qualia of your friend. Your imaginative understanding is central to your response: it's not merely a detached or 'scientific' response to your friend, viewed only as yet another terrible statistic. You're thinking about what it must be like to live '*in this*'. But that's not a concern with qualia, as usually conceived.

Of course Paul's principal concern is not with imaginative understanding of other people, it's with imaginative understanding of your own future life. But here the very same points apply. Suppose, for example, that you're contemplating a risky career move, which really may lead to you living in poverty. You want to know, what will that be like? This is a question that calls on your imaginative understanding, if you're to give an authentic response. But it's not a matter of trying to imagine what your qualia will be. It's a matter of trying to imagine what it will be like for you to live *in such circumstances*. And only a needlessly internalist conception of the mental will compel you to think that that's entirely a matter of speculating about what qualia will be produced in you by these external circumstances (or that, if you're not reflecting on qualia, you can only be engaging in an 'external', non-imaginative exercise).

The problem is exacerbated by the fact that one of the big ways in which philosophy of mind has found a place for imaginative understanding is in

providing knowledge of perceptual experience. That means that it's natural for Paul to try to get her point across by considering cases in which, for example, someone born color-blind is given the opportunity to see color, or someone born deaf is given the opportunity to hear. In these cases someone is being asked to make decisions regarding an outcome they can't understand imaginatively. So, given the internalism about perceptual experience of much contemporary philosophy, it's natural to suppose the role of imagination is to advise you as to the nature of your future qualia. Lacking that imaginative understanding, how can you make an authentic response?

Yet even in this case, I think we can see that the focus on an internalist, affectless conception of imaginative understanding doesn't get the picture correct. If you're being given the opportunity to hear, or to see the colors, this isn't properly conceived as a matter of getting a new set of inner qualia. Of course, the whole poignancy here is the impossibility of the imaginative exercise; but let's reflect for a moment on what it is you would like to do. What's important is that you're being asked to imagine, *of* the colors, what it will be like to see *them*. That's a *de re* imaginative exercise that you're being asked to perform. Unless you factor in that the possibility being offered to you is one in which you gain *knowledge* of some aspects of your environment, you've missed a principal factor that you ought to be taking into account in your assessment. And if you manage the imaginative exercise, but find yourself with no affective response one way or another to seeing the colors, that is in itself an important point about the possibility of 'authentic' decision-making. What would make authentic decision-making easy is being able to imagine what it would be like to see the colors; and once you have that, finding the prospect immediately irresistible, as part of the imaginative exercise. The immediate engagement of emotion by imagination is, as I've been emphasizing, a central component in authentic decision-making.

3. The Role of 'Knowing What It's Like' in *Transformative Experience*

The role that Paul envisages for imaginative understanding is to provide you with knowledge of what your future experiences will be like. When making an 'authentic' decision, you choose based on what the character of your future experiences will be. Only imaginative understanding can provide you with that knowledge of what your future experiences be like.

> When you consider what might happen in your future, your consideration involves an imaginative reflection on what it will be like, from your point of view, to experience the series of future events that are the most likely outcomes of whatever it is that you choose to do. You use this reflection on what you think these events will be like, that is, what you think your lived experience will be like, to authentically determine your preferences

> about your future, and thus to decide how to rationally act in the present. This . . . . fits our philosophical account of the responsible, rational agent. So there is a clear philosophical underpinning for the powerful sense of control and authenticity we get . . . .
>
> (106–107)

There seem to be cases of 'authentic' decision-making that aren't at all of the type Paul describes. Suppose you have a friend who has decided to accept a post as a high-school teacher in a bad part of town. You ask her, 'But have you thought about what your subjective experiences will be like when you are doing this work?'. And she explodes, 'My subjective experiences have nothing to do with it! Have you seen the state of these kids?'. The notion of 'authenticity', having something to do with the expression of one's deepest values, or true identity, is of course notoriously difficult to make fully explicit. But I can't see that there need be any failure of authenticity on the part of your friend in this case. The reasoning is still based on an imaginative exercise: how she could make life for those kids.

To put the point crudely, suppose you are choosing between A and B. And suppose someone comes along with a machine that, if set to A*, can generate experiences in you that you couldn't discriminate from experiences of A. The imaginative exercise required to find 'what A would be like', would presumably, on Paul's conception of it, be exactly the same as the imaginative exercise required to find 'what A* would be like'. So the imaginative exercise couldn't discriminate A from A*. So on this conception of it, 'authentic' decision-making couldn't justify choosing A over A*.

I think it's only very rarely that we reason in this way, and never in the context of anything that you might call a 'big life choice'. Consider an example Paul discusses a number of times, someone deciding whether to have a child. She writes:

> 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 subjective values of the experiences stemming from the choice about having a child play the central role in the decision to procreate.
>
> (75)

I am struck by the contrast between this way of framing the decision and the perspective of a friend of mine who said:

> 'I want to have children because my life has always been entirely centered on me, and I want to be forced to live in a different way.'

It doesn't seem to me that this can be represented as someone bowing to 'external or impersonal' factors, and failing to make an 'authentic' decision. But neither was my friend making a decision based on the expected qualitative characters of her future experiences. Rather, she was making a decision based on reflection about her own identity: what kind of person she was and what kind of person she wanted to be.

Consider someone wondering whether to have a second child. If 'authentic' imagination-based reasoning really centered on reflection about your own future experiences, then the impact of the second child on your first child would simply not be a factor; or rather, it would enter into your decision-making only because the first child might make a difference to your own future subjective experiences. But consider someone who says, 'Although things will be harder for me, I think it's important for Benjy to have a sibling'. In effect, you give a far higher weighting to Benjy's future subjective experiences than you do to your own. The turning point in your reflection might be a thought about how things will be for Benjy after you are dead; here reflection on your own future subjective life is not what is going on at all.

In a striking passage, Paul tries to bring out the importance of an imaginative understanding of your own future subjective life for authentic decision-making. She says:

> Imagine Sally, who has always believed that having a child would bring her happiness and fulfillment, deciding not to have a child simply because the empirical evidence tells her she will maximize her expected value by remaining childless. For her to choose in this way, ignoring her subjective preferences and relying solely on external reasons, seems bizarre.
>
> (87–88)

A slight modulation gives us the case of Billy:

> Imagine Billy, who has always believed that taking dangerous drugs would bring him happiness and fulfillment, deciding not to take them simply because the empirical evidence tells him that's how he will maximize his expected value. For him to choose this way, ignoring his subjective preferences and relying solely on external reasons, seems bizarre.

The general point here is that while it seems an extremely interesting idea to connect authentic decision-making with imaginative understanding, there is a certain 'internalism' about Paul's understanding of imaginative understanding that is skewing the picture. We should indeed, as Paul says, connect authenticity to imaginative understanding; that seems to me the great insight of the book. The trouble is the implicit restriction of imaginative understanding to 'knowing what it is like' to have various experiences,

where this is understood in 'internalist' terms, so that facts about the environment are thought of as 'external' or 'impersonal' factors, that can only undermine the 'authenticity' of one's decision-making. After all, slight modulations of Sally's case too might leave us feeling that she ought to be allowing her dreams to collapse, given the dangers she faces. We could agree that authentic decision-making has to be grounded in imaginative understanding, while insisting that imaginative understanding must always encompass facts about the environment. Suppose, for example, that Sally is working obsessively towards her dream of being a great landscape gardener. She has a vision of her beautiful gardens enriching every state in the country. She has an imagination-driven approach to decision-making on all the big things in her life. But that's not because what she cares about are her future subjective experiences. What she cares about are the gardens.

References

Jackson, Frank. 2002. 'Epiphenomenal Qualia.' In *Philosophy of Mind: Classical and Contemporary Readings*, edited by D. Chalmers. New York: Oxford University Press.

Jaspers, Karl. 1913/1997. *General Psychopathology*, translated by J. Hoenig and M. W. Hamilton. Baltimore, MD: Johns Hopkins University Press.

Nagel, Thomas. 2002. 'What is it Like to be a Bat?' In *Philosophy of Mind: Classical and Contemporary Readings*, edited by D. Chalmers. New York: Oxford University Press.

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

Philosophy and
Phenomenological Research

Philosophy and Phenomenological Research
Vol. XCI No. 3, November 2015
doi: 10.1111/phpr.12250

Transformative Experience: Replies to Pettigrew, Barnes and Campbell

L.A. PAUL
University of North Carolina, Chapel Hill

I am very grateful to my three symposiasts for their thoughtful, generous, and philosophically rich comments. The questions and challenges they raise bring out further issues and have helped me to develop the idea in several new directions.

Reply to Richard Pettigrew

Richard Pettigrew proposes a beautifully clear model for how to make transformative decisions under epistemic and personal transformation. The key to Pettigrew's model for epistemically transformative decisions is to replace what we can't grasp epistemically with uncertainty about possible utilities. He then proposes that we model epistemically and personally transformative decisions using sums of weighted local utility functions across time (where each local function can reflect local epistemic uncertainty at that time).

I agree with Pettigrew that decision theory can be reformulated in the way he proposes, but such models will still fail. They imply profound epistemic and metaphysical alienation for the decision maker, and, as he points out, will still leave us with difficult problems concerning how to weight local utility functions. What Pettigrew has shown us is how to make some versions of these questions precise, and, in cases of transformative choice, he has drawn out just how deep and fundamental the divide between deciding rationally and deciding authentically may be.

1. Replacing the unknowable with the uncertain

The first main feature of Pettigrew's model is to substitute uncertainty about utilities for the inaccessibility of subjective utilities in cases of epistemically transformative experiences. In effect, we are replacing a utility function

whose values for certain outcomes are undefined with a utility function that assigns sets of possible values for those outcomes.

The immediate problem with the model is that it is modeling the wrong decision problem. Mary's not knowing what it will be like for her to see red if she leaves her black and white room is not the same as Mary's being acquainted with a wide range of color experiences but not knowing which one of these experiences will be relevantly similar to her experience of seeing red when she leaves her room. Nor is it the same as Mary's being uncertain about which member of a (potentially enormous) range of possibilities, all of which are epistemically accessible to her, will obtain. Rather, what it's like to see red is simply epistemically inaccessible to Mary until she has the requisite experience.[1]

For this reason, I am not inclined to accept Pettigrew's redescription strategy for epistemically transformative choices. It isn't a redescription of the same decision problem: it replaces that problem with a different one.

Pettigrew recognizes this, but wants to collapse the difference for the purposes of rational decision-making.[2] He argues that "... you don't need to know the possible phenomenal characters of the experiences that you will have at each outcome in order to know what the possible utility hypotheses are ... You simply need to know the possible values of your utility function—and you do know that, since possible utility values are all real numbers."[3]

But something important has been lost. Consider InvertMary, who is functionally identical to Mary, but phenomenally inverted with respect to her color experience. If InvertMary sees a green apple, she'll have color experiences phenomenally identical to Mary's experiences of a red apple, and vice-versa. Let's also assume that, for Mary, the numerical value of the utility of what it's like for her to experience red is the same as the numerical value of the utility of what it's like for her to experience green.[4] Since Mary and InvertMary are functional duplicates, their numerical utility values for their experiences of seeing red and their experiences of seeing green are identical. So Mary and InvertMary can expect the same numerical utility if they decide to leave their respective black and white rooms. But it is obvious that what it is like for Mary to see a red fire engine for the first time is different from what it is like for InvertMary to see a red fire engine for the

---

1 We might hold that since she lacks the relevant phenomenal concept, she simply cannot represent her experience in the sense that's needed.

2 John Collins (2015) explores the possibility of rational neophobia (fear of the unknown) in these sorts of cases.

3 Pettigrew, 2015, p. 769.

4 If we make enough symmetry assumptions about Mary, we can do the case without InvertMary, but I think the InvertMary comparison is clearer.

first time, and so something important about the decision-making for each of them differs, even if the numerical values assigned to their subjective utilities are the same.[5]

Suppose, despite these worries, we adopt Pettigrew's redescription strategy. Then we can make transformative decisions rationally—but, as I shall argue, we *still* suffer an important kind of loss, and we cannot escape the problems raised by transformative choice. Agents who use Pettigrew's model for transformative choice face two distinctive types of alienation from their outcomes. I'll explore each type of alienation in the context of a thought experiment.

2. Alienation under rational choice

Consider a situation where you desire to have a baby. Your colleague has built a computer, call it "HAL", who can calculate your utilities for you. You consult HAL, and he tells you that you can expect a utility in the range between 2 and 3 if you have a child, and between 6 and 7 if you don't. You attach equal credences to each possible state given that the relevant act is performed. Given this, you will maximize your expected utility by choosing to remain childless, even if you are uncertain about just how much.

You can't understand HAL's assessments, because although you don't feel like you have a detailed grasp on what the future would be like (everyone tells you life changes dramatically), your own assessment of your utilities for having a child by imaginatively or introspectively prefiguring your future self as a parent assigns a very high utility to having a child, and a very low utility to not having one. In short, you desperately want to have a child.

Given that choosing to act in a way that does not maximize one's utility is not rational, then according to HAL's assessment of your utilities, you can't rationally choose to have a child, even though this conflicts with your assessment of yourself. In this situation, to choose rationally, you must revise your beliefs, allowing the computer's determination of what you are to believe about your utilities to replace your own introspective assessment of your heart's desires.

Nevertheless, you believe in HAL, and so you accept his assessment for you, even if it does not comport with what you believe about how you would respond. As a result, you are epistemically alienated from your rational choice by your imaginative incapacities.

---

[5] This example is *not* intended to suggest that subjective value is based on an internalist notion of experience. See my reply to Campbell below.

But what is HAL doing when he tells you what the range of your future utilities will be? HAL is, in effect, considering you in the actual world, @, at t1, and then assessing your utilities at t2 in different possible worlds W1 and W2. In W1 at t2, you have a baby, and in W2 at t2, you do not have a baby. HAL has to assess your utilities in different possible worlds because he is assessing what the actual world would be like under different possible changes of state. (Before you have a baby, as I discussed above, W1 at t2 is epistemically inaccessible to you, but HAL reports back about what he finds.)

Do you exist in W1 and W2? Yes—or at least your respective counterparts do. Let's call the person who exists in W1 at t2, "C1" and the person who exists in W2 at t2, "C2". There is a problem with C1, the person who is identical to you (or who represents you) in W1 at t2.

Here is the root of the problem. Normally, with a state change, the agent is kept fixed, in order to assess her utility in the new state. But the state change represented by W1 does not exist in isolation: because the state change involves an epistemically and personally transformative experience for you, changing the state of the actual world *also* changes your preferences and your psychological capacities. If C1 is you in W1 at t2, this represents a significant change in your first-personal perspective.[6]

The trouble is that at t1, in @, when you consider the choice to have a baby, from your first-personal perspective, C1's point of view is psychologically alien to you.[7] You cannot project your point of view into C1's point of view, or grasp her point of view as an extension of your own.

David Velleman's work on personal identity and persistence (1996) brings out the importance of having psychological access to one's future self: "The future 'me' whose existence matters [to me] is picked out precisely by his owning a point of view into which I am attempting to project my representations of the future, just as a past 'me' can be picked out by his having owned the point of view from which I have recovered representations of the past."[8] While C1 might be, strictly speaking, personally identical to you, from your actual perspective at t1, C1 is not an eligible future self, because C1 is not psychologically accessible to you in any first-personal sense.[9]

So, in this sense, the utilities that HAL is discovering in W1 are not the utilities of your future self. They are the utilities of C1 at t2, but from your

---

[6] It represents a change in the features of the agent whose utility is being assessed, not just the circumstances of the world in which the agent is embedded.

[7] Or, we might say, C1 isn't who you, from your @-at-t1 vantage point, would identify as your psychological counterpart.

[8] p. 76.

[9] On some metaphysical accounts of personal identity, the *same person* relation merely requires the right sorts of causal or other sorts of continuity. The point here is that *same person*, *future self*, and *same self* are different relations, and the relations that matter in these decision contexts are the *same self* and the *future self* relations.

first personal perspective at t1, C1 is an alien self, and in this sense, C1 is *not you*.[10] When you consider your decision at t1, you want to know how you'll respond to the experience at t2, that is, whether your preferences will be satisfied. Wanting to have *your* preferences satisfied carries with it an implicit, psychological, first-personal constraint: when you make an important personal decision to act in a certain way, you want to know the (range of) utilities that the person who you can first-personally identify as your future self will have. In other words, when you assess your possible acts, you want to have psychological access, in an anticipatory or imaginative way, to each of your possible future selves. For each possible act, you want to grasp the first personal perspective of the self who you'd be making yourself into, and who will live with the result of your choice.[11]

Because, from your first-personal perspective at t1, C1 is not you (or, if counterpart theory is preferred, C1 is the wrong counterpart), using Pettigrew's model entails a type of metaphysical alienation from your possible future selves. If you are facing a possible change, and you are psychologically alienated from the person who will result from this change, then the person who results is not your future self: the metaphysical relation of future-selfhood between who you are now and who you will be after the change does not obtain.

In fact, the person at t2 that you (in @, at t1) want to assess is a person in W3 at t2, a world with a state change in which you have a baby but your preferences and perspective remain the same as they were in @ at t1. That person, call her "C3", is (perhaps) psychologically accessible to you, but more importantly, she exists in a physically inaccessible world, and so she will not be the self that results from your having a baby.

So the sort of alienation you face doesn't arise from deferring to the dictates of morality or to your social group. We are not in the domain of traditional Existentialism. Rather, rational choice in transformative contexts entails epistemic alienation from the outcomes of one's choice and metaphysical alienation from one's future selves.[12] If we rely on decision models that replace our inability to know or to grasp utilities with uncer-

---

[10] Or, I'd be inclined to say, C1 is the *wrong* counterpart, given your first personal perspective. "It's the wrong trousers, Gromit, and they've gone wrong!"

[11] Counterpart theoretically: you want to know the (range of) utilities of a counterpart that is psychologically similar in the relevant first-personal sense to who you are now.

[12] There *is* a direct connection to morality, but it isn't through traditional existentialism. Rather, transformative choice suggests that what our best decision models propose conflicts with what is personally, socially, and morally acceptable. In other words, the models for such choice that fit with normative decision theory conflict with other traditional kinds of normativity: the normative ideals for personal, social and moral choices. At least, they conflict to the extent that our personal, social and moral choices rely on our ability to imaginatively access our own future perspectives and the perspectives of others. See, for example, Harsanyi (1977) and Holton and Langton (1998).

tainty over utility values that are given to us, the epistemic and personal changes that feature in transformative choice entail profound epistemic alienation from our possible future outcomes, and profound metaphysical alienation from our possible future selves.[13]

3. Can social science save us?

Pettigrew argues that we should replace utilities that are introspectively inaccessible with a range of possible utility values. He suggests that, in many cases, the values of our utilities can be determined using statistical evidence gathered by psychological and social science. If so, the fact that I must dispense with introspection is not a significant loss, for I can (and perhaps, in deference to science, I should!) replace my introspective utilities with (sets of) utilities determined by the statistical data. But this strategy will not evade alienation, for in cases of transformative change, statistical evidence cannot tell an individual what her own, individual-specific utilities are.

One important reason, which I will nevertheless set aside in what follows, is pragmatic. It is not currently possible for the psychological and social sciences to tell us, even allowing for some uncertainty, what our individual utilities are for big life decisions. We have nothing that's even close to good enough data. So, in the immediate future, there's no hope of the science even approximating the job that HAL did in the example above. Moreover, as technology and culture develops, the choices the world offers can be extremely complex and are constantly changing. Entirely new kinds of personal choices arise with major technological and scientific advances (for example, human egg donation). Thus, for real-world big decisions in the immediate future, adequate statistical data is unavailable, so we must immediately face the philosophical losses of transformative decision-making in the real world.

Let us set this issue aside for the purposes of discussion, and pretend that the statistical data is in fact adequate. If we assume that it is adequate, can I use such data to discover my own utilities? No, I cannot.

The first reason I cannot use such data to discover my own utilities is because what I know from the data is merely general. What the statistical data can tell me is what the *average* effect (or utility value) would be for any member of the population (which, by assumption, we take to be composed of individuals similar to me). The average effect, however, is perfectly consistent with wide and dramatic variation in the values assigned to

---

[13] If we do not employ the replacement strategy, we can frame the epistemic and metaphysical alienation differently. Since your preferences cannot even be formed until you've had the transformative experience (since you cannot conceptually represent the outcomes in the relevant way), your values for your outcomes are epistemically undefined and your counterpart relation is metaphysically indeterminate.

utilities (including the range of uncertainty) for any particular individual member who is included in this average. In fact, with real data, we see such variation all the time.

Intuitively, we may wish to use introspection to help us interpret average utility values and ranges of uncertainty with respect to an individual case. In particular, I might wish to consider average utility values for a member of my population with respect to how it comports with my introspective assessments of my own utilities, in order to refine my own personal utility values and my own range of uncertainties. But in the context of transformative decisions, no such introspective method is available.

It may be that discovering such average values are the best we can do, but we must be clear that *to discover the average utility values and its range of uncertainties for a member of my population is not to discover my individual utility values and my own, individual range of uncertainties*. The value of the average utility for a member of my population may be quite different from my own individual utility, given individual variation. Put another way, to choose the act that you expect to have the highest utility by the lights of the average member of your population is not the same thing as to choose the act that you expect to have the highest utility by your own lights.

The second reason I cannot use such data to discover my own utilities is because the data *conflates* two distinct types of utilities. Assume that the empirical data tells you the average utility values for members of your population after they undergo a transformative experience. Also assume that you assign these values to the relevant outcomes before you make your own transformative choice.

The trouble is that the data cannot distinguish between the future utility for the individual who is actually making the choice, and the future utility for a different individual who merely *replaces* the individual who is making the choice.[14] In the language of the previous section, the data cannot distinguish between the utility for a future self of yours and the utility for a replacement, alien self who merely results from your undergoing the experience. The data conflates two senses of "your" utilities, for the number it reports as the utility value in each case is the same. This is the deep problem with using statistical data to model your preferences about a transformative experience, and with a choice that is both personally and epistemically transformative, it is this distinction that is of the essence.

Normally, when faced with a choice that entails personal change, we introspect and reflect in order to be as sure as we can that we truly know our own preferences, but *also* so that we choose in concert with who we

[14] The replacement individual may well be personally identical to the original individual, but may not be the same self as, or a future self of, the original individual. See section 2 above.

really are, and especially, *with whom we want to become*. When undertaking a life-changing decision, we want to knowledgably control who we are making ourselves into, that is, to knowledgably choose our future self, and imaginative introspection is an important guide for doing so. But when the choice concerns an experience that is both epistemically and personally transformative, introspection cannot guide us in this way.

The problem is, neither can the data. For the social-scientific data that provides post-choice utilities can't distinguish between (i) a utility value that represents what your utility will be at t2 by your own lights, despite your inability to introspect to that result beforehand, and (ii), a utility value that results from replacing yourself with a different, psychologically alien self at t2, a value that represents what *her* utility is by *her* own lights.

4. Forming our future selves

How then, in the context of the alienation entailed by a choice that is both epistemically *and* personally transformative, are we to understand the difficult philosophical questions posed by Pettigrew's elegant model of personally transformative decisions?[15]

One problem involves how we are to weight changes in utility across time, since, as Pettigrew astutely points out, we cannot assume our higher-order utilities will remain constant. There is no epistemically neutral first-personal perspective that the agent can occupy in order to solve the decision problem in a principled way. Another problem involves how we are to address the fact that such changes are not first-personally forseeable, and so, if we assign future local utilities based on statistical data, we embed a kind of psychological indeterminacy and subsequent alienation into the model.

To sum up: First, in a transformative context, our future utilities, including our higher order utilities, may change, and we lack a normatively principled decision-theoretical rule for such changes. Second, we are epistemically and metaphysically alienated from these new future local utilities. We lack the ability to epistemically "see" these future utilities from the subjective perspective. Third, such alienation may force us accept a decision model in which the future utilities we assign do not discriminate between utilities for natural psychological extensions of our current selves and utilities for effective psychological replacements of our current selves.

This, then, is the problem. In contexts of transformative choice, how are we to make decisions within the constraints of deliberative decision theory? How are we to determine and follow the relevant diachronic rational norms?

---

[15] Also see Briggs (2015).

And how are we to do so while preserving a role for authentic, informed choice in high stakes cases of great personal importance?

In this way, exploring the questions of transformative decision-making brings new kinds of philosophical problems to our attention, ones that raise questions about authenticity, alienation, and the human condition framed in formal epistemological and phenomenological terms.

Reply to Elizabeth Barnes

In her fascinating piece, Elizabeth Barnes challenges me to explain why, for a person who has never wanted children and who strongly prefers to remain child-free, choosing *not* to have a child is not rational. In particular, she presses me to explain how, in this kind of case, choosing not to have a child is any different from other kinds of cases where it's rational to avoid having a transformative experience.

Now, Barnes understands the issue in a certain way. In particular, she understands it as involving the claim that one can rationally choose to avoid the experience based on the belief that whatever it's like, it's something you don't want *right now*. Even if having the experience would make you very glad that you'd had it, this is completely irrelevant to the rationality of the decision-making process. If, right now, you don't want to have a child, than any changes of mind you'd have afterwards don't matter.

To understand the force of her argument, we need to identify the relevant constraints and consider her examples. First, we are restricting our attention to decisions framed in terms of subjective values and preferences. The decisions of interest are decisions made by an individual and concern her immediate, first-personal future, and are framed in terms of the individual's preferences and values regarding the possible characters of her future lived experiences.

Barnes thinks that the choice to have a child is a member of a class of choices where one can rationally choose to avoid the experience based on the belief that whatever it's like, it's something you don't want right now. To identify this class, she considers two cases where, she argues, the agent can rationally choose to avoid having the transformative experience and its attendant personal changes. In each case, the rational basis for the agent's decision is his belief that whatever it's like to have that experience and to change in that way, and whatever he will think about the choice in the future, it's something he doesn't want right now.

The first example is from Orwell's *1984*: Winston wants to avoid being captured and "reprogrammed" by the Thought Police. Even if he knows that after the mental reprogramming he'll be very happy as a devotee of Big Brother, Winston can rationally choose to avoid reprogramming. The second example is from *Star Trek*: Captain Picard wants to avoid assimilation

into the Borg, a collective hive mind. Picard rejects assimilation even though he believes that after the assimilation he'll be very glad to be Borg.

The examples are chosen to support the claim that, even if I believe that the transformative experience of having a child is likely to result in an outcome where I will be very happy to be a parent and will take my lived experience to be very valuable, there is a sense in which I can disregard this fact. For, after all, who I'd be as a parent is not who I am now. So why can't I, now, ignore the preferences of that merely possible, vastly different self? (For an interesting, related argument, see Briggs (2015).)

Barnes's argument is that, for the child-free decision-maker in her example, the choice to have a child is just like the choice to avoid the Thought Police or to avoid becoming Borg. According to Barnes, in each case, the rational basis for the agent's decision is his belief that whatever it's like to have that experience and change in that way, and whatever he will think about the choice in the future, it's something he doesn't want *right now*, and that's all he needs to consider. According to Barnes, the same is true for the child-free decision-maker: having such preference changes are something she doesn't want *right now*, and a rational basis for a person's decision to choose not to have a child is the belief that, whatever it's like to have that experience and change in that way, it's something she doesn't want right now. So that's all she needs to consider. Case closed.

Case reopened. I agree that Winston can rationally choose to avoid reprogramming and that Picard can rationally choose to avoid becoming Borg, and that both agents are correct in disregarding the preferences and values of their future selves. But choosing to a child is not the same sort of choice. The real-life case of choosing to have a child is disanalogous.

Barnes's fictional cases *are* analogous to real-life cases of mind control, such as being hypnotized, being drugged, or becoming a member of a cult. In these sorts of cases, we are rational in disregarding the values and alien preferences of the selves who would result from the mind-altering manipulation being proposed. In such cases, the transformative experience involves a loss of mental autonomy. Mental autonomy is something that we rational agents currently have experience of, and something to which we assign a very high subjective value—so high that the value of having it can swamp the subjective value of almost any lived experience that lacks it. But there is a further difference. We also think mental autonomy has objective moral and social value, and often, this objective value trumps mere subjective value. Our assignment of a very high subjective value to mental autonomy reflects this objective underpinning, justifying the irrelevance of the subjective value of future lived experiences without mental autonomy, especially when one's current lived experience with mental autonomy is tolerably pleasant.

Becoming a parent is not like becoming a member of a cult, being drugged, getting hypnotized, or having one's mind assimilated into a collective mind. That is, becoming a parent does not involve a loss of mental autonomy.[16] While becoming a parent does involve significant changes in one's preferences, one's self-definition, and the character of one's lived experiences, external mind control or permanent mental impairment is not part of the outcome. The identity and preference change involved in becoming a parent is deep and far-reaching, but one's mental autonomy and mental capacities are not ordinarily lost or significantly impaired by the change.

The difference stems from whether making the choice should involve cognitively modeling yourself forward into the shoes of your possible future selves, the selves who would result from the decision. Some decisions should involve this sort of self-projection and some should not, for some sorts of decisions turn on the subjective value of what it's like to be that future self, and some do not. In particular, decisions involving the loss of mental autonomy do not. When an act results in a loss of mental autonomy that degrades the status of the future self's testimony and lived experience, it can be rational to disregard what it is like to be that future self when choosing how to act.

The problems raised by transformative choice, such a choosing to have a child, concern a different type of decision. Such decisions include cases where we must consider the possibility of becoming a self that is epistemically alien to us. But in addition, there cannot be an objective value that trumps considerations based on the subjective value of what it's like to be that alien future self. Transformative choices occur in cases where, in effect, our way of framing the decision presents us with an open field of possibilities. They are decision situations where, after the relevant objective moral, social and other nonsubjective values are taken into account (whether or not they are reflected in our subjective preferences), there are still multiple particular courses of action available to the rational agent. In such situations, we ordinarily want to make the choice in question by cognitively modeling ourselves forward into the perspectives of our possible future selves, so we can choose who we are making ourselves into in an informed, authentic manner. Choosing whether to have a child is just this sort of choice.

It is a distinctively modern situation to be in: one where we can choose different ways to realize our future selves without having our path laid down for us by the authorities. In this situation, we are permitted to make our own way through the field of possibilities, and we do so, in part, by assessing our subjective preferences for how we'd like our lives to go. Philosophers have devoted a lot of attention to decisions where objective values are the values we are concerned about. But many of our big life

[16] OK, apart from 3am feedings. There isn't a lot of mental autonomy there.

decisions involve "open field" cases where the choice to hand is focused on maximizing our expected subjective value, because purely objective values, empirical facts, and other constraints have already been taken into account.

Thus far, I have rejected Barnes's comparison, arguing that the cases of reprogramming by the Thought Police and assimilation by the Borg are disanalogous to the transformative choices of interest, such as the choice to have a child. Since becoming a parent is not analogous to being drugged or mentally controlled, there is no justification *of this sort* for the child-free person to dismiss what her future lived experience would be like and what her future preferences would be when she makes her decision.[17] However, as I argue in *Transformative Experience*, one *can* rationally refuse to discover what it's like to become a parent. How? By reframing the argument. If you are like Barnes's child-free decision-maker, you can rationally choose to forego the discovery of what it's like to become a parent by choosing to keep your current preferences rather than to discover new ones.

But there is a deeper way to interpret Barnes's argument. Part of what Barnes is suggesting is that Winston and Picard should disregard the values and desires of their possible, mind-altered future selves because those selves are damaged in some way. They are cognitively impaired agents, and thus their wishes, once they exist, should not be allowed to affect the rational decision-making process.

In other words, I interpret Barnes as raising a deeply interesting question: when, exactly, should we regard major cognitive changes in ourselves as destructive of our mental autonomy? When would making a radical epistemic and personal change in myself count—from my point of view before the change—as my making myself into a cognitively impaired agent? Obviously, if I were to choose to undergo a lobotomy I'd be making myself into a cognitively impaired agent. But when does radical change in myself, simply as mere radical change, effectively amount to a loss of control over how *I* think?

When I undergo a transformative change, I change my epistemic capacities and my core personal preferences, and as a result I change the character of the way I think and the way I first-personally experience the world. My response to a life-changing transformative experience will define and infuse the character of the ways I experience and value the world and myself, perhaps for the rest of my life. And essentially, in contexts of transformative change, I must decide whether to undergo such a change without being able to first personally forecast or model how the change will go, and thus without being able to grasp the nature of the cognitive change from my first

[17] Dear reader, please note: all this simply suggests that the choice to not have a child based on expected subjective value is no more rational than the choice to have one. We are all in the same boat.

personal perspective. This is precisely why the combination of radical epistemic change along with radical personal change is so threatening.

In such contexts, when does choosing to undergo such a change amount to giving up one's cognitive capacities in the pejorative sense? When is it the case that, before I make a decision to become a different sort of self, I can rightly regard my future self as cognitively impaired, relative to my current self? Should I, from the perspective of my current self with her current preferences, regard *any* dramatic change of my preferences, especially transformative changes to my core personal preferences, as a kind of cognitive impairment? Where is the line between revising one's preferences in response to experience such that one autonomously *learns from* the experience, versus having one's preferences *controlled by* the experience?

If just any transformative change counts as cognitive impairment, then Barnes's thesis endorses an unhappy conservatism: don't ever leave your small town, don't ever get a college education, and don't ever change your current political perspective, because, by your current lights, the self that results from such experience will be cognitively impaired. Barnes, of course, is not arguing for this sort of conservatism. But where do we draw the line in dismissing the epistemically inaccessible subjective perspectives of our possible future selves? What we've found is a connection to the point made by Richard Pettigrew in his comment (Pettigrew, 2015), who gives a formal presentation of a related question—how are we to weight our local utility functions over time when framing and contemplating the possibility of transformative change, especially when those future changes are first-personally inaccessible to us?

This, of course, is just the sort of question I intended to raise when writing *Transformative Experience*. I regret to say that I do not know the answer.

Reply to John Campbell

What must imagination be in order for it to play a suitable role in authentic decision-making? In his thoughtful and insightful comment, John Campbell argues that in some cases, for such a decision to be authentic, it must involve "...imagining how the external, mind-independent environment is, as well as the mental life that is located in that environment, and affective, in that your exercise of imagination directly engages your emotions and actions, without any need for further reflection." (Campbell, 2015, p. 788). I agree with Campbell, although my conception of imagination may not be quite the same as his.

1. Imaginative knowledge

To make an authentic decision in contexts of personal change, one that reflects an informed, first-personal grasp on who you are and what you care about, you often want to know how you'll respond to the effects of your

acts, including whom you'll become. 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.

When you make a decision in this way, you use your imagination to project yourself mentally forward into the first-personal perspectives of your possible future selves. On my view, for many big, life-changing decisions, you want to authentically assess your options by assessing the subjective value of your possible future lived experiences. Ideally, the assessment involves a determination of the subjective value of each possible outcome of your decision, that is, each possible lived experience, by imaginatively grasping what it would be like for you to live in that future. That is, you want to assess what it would be like for you to first-personally occupy the self who lives that experience in that outcome, and so you *imaginatively empathize* with your possible future selves.[18] In this way, imaginative empathy can play a central role in authentic future-self-creation, or authentic self-realization.

Intuitively, 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 Campbell puts it. 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.

So the character of one's inner life plays an important role in determining the subjective value of lived experience. But we need not understand this in a purely internalist sense. Often, what we care about is what the experience is subjectively like for a person, *given the circumstances she is in.*

Hence, subjective values need not be *merely* phenomenological or *merely* experiential. One way to put this is that, by assumption, an agent making choices about her futures assigns subjective values to outcomes concerning possible lived experiences, where the value of the lived experience can include what it would be like for her to "live" that outcome in the environment she is in.

Campbell is correct, then, that my approach to authentic decision-making must make room for an approach that extends past valuing experience understood as merely valuing one's purely internal, sensory phenomenal character, and thus should extend past an internalist conception of the imaginative task involved in grasping such value.

---

[18] It's worth noting the parallel here with Harsanyi (1977), who argues that "the basic intellectual operation . . . is imaginative empathy. We imagine ourselves to be in the shoes of another person, and ask ourselves the question, "If I were now really in *his* position, and had *his* taste, *his* education, *his* social background, *his* cultural values, and *his* psychological makeup, then what now would now be *my* preferences between various alternatives." (p. 638)

Subjective values are values of lived experience, and such experience often includes one's experience of the environment, and one's experience of her environment often includes her beliefs about her environment. So subjective values need not be internalist, at least not in the sense that Campbell urges us to reject. They do not need to be understood as merely concerned with the character of one's internal mental life, and I never intended them to be so understood. They are, instead, concerned with the character of one's lived experience, which can include her experience of her environment.[19] Since authentic decision-making can involve knowledgably imagining the subjective value of lived experiences, I would not want to be committed to an internalist conception of imagination that excludes this.[20]

Campbell, however, worries that I am implicitly endorsing a purely internalist conception of imaginative understanding, because my argument for authenticity and imagination exploits thought experiments in the philosophy of mind that assume internalist conceptions of qualia. Those thought experiments demonstrate the power and importance of experience in generating our imaginative capacities. However, they do so in a context where, at least arguably, an internalist conception of qualia is assumed.

I appreciate the chance to set the record straight. When using these thought experiments to frame my arguments, I am not doing so in an internalist context. That is, I do not assume that 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, nor do I adopt a restrictive, internalist or affectless approach to the imaginative act required to assess the subjective value of lived experience. In many cases, the subjective value of an experience, as well as the imaginative act needed to authentically grasp this value, will depend partly on the environment.

Fortunately, the thought examples will do their job as long as the information gained by the qualitative discovery (even a merely qualitative discovery), functions as a necessary element in the epistemic and personal transformation of the agent.[21] It is true that imagining the subjective value of your future lived experience may not merely involve imagining what

---

19 In the book, I attempt to capture this by noting that they are concerned with the character of one's *veridical* lived experience, which builds in an externalist safeguard.

20 Nor do I think imagination's function is to provide affectless knowledge of one's internal mental life.

21 Does my view require us to factor imagination into an "internal experience" part and an "external experience" part? As I've indicated, I don't think it does. I'm not a fan of internalism *or* of externalism about qualia: I find the distinction misguided. However, to defend this here would take me too far afield. I thank Campbell for pressing me on the point (in conversation). I think there are large and complicated issues about how to understand qualia: for example, see Mark Johnston's (2006) criticisms of what he describes as the "Wallpaper View".

your qualia will be like as you respond to the transformation. But you must still grasp what your qualia will be like in order for you to be able to cognitively evolve yourself forward under the transformative change, and in order to grasp what it will be like for you to live in your possible future circumstances. In cases of epistemic transformation, experience, even experience understood in purely internalist terms, is still needed to teach you what your future could be like, since qualitative experience of the relevant kind is needed to give you the imaginative capacity to first-personally represent or model your possible future selves in those possible future circumstances.[22]

That is, 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 and will have a significant effect on how your core personal preferences are formed and developed. So experience is needed in order for the agent to assess the subjective value of her possible future lived experiences, whether or not we endorse internalism.

Before Mary in her black and white room decides whether to exit and discover what it's like to see color *out there*, she cannot imagine what it would be like to see red, where this includes her inability to imagine and assess the nature of what her lived experience in a colorful world would be like. She cannot imagine what it is like to see red, nor can she imagine any of the effects of that discovery on the character of her lived experience, and so she cannot imagine what it is like to see red *out there*, nor can she imagine any of the effects of that discovery on the character of her lived experience *out there*.

If a congenitally blind adult were to decide to have a retinal operation that allowed him to see for the first time, he would discover what it was like to live in the world as a sighted adult. Before he has such an operation, he doesn't know what it would be like to be sighted, and thus he cannot imagine what it would be like for him to live in the world as a sighted adult, with all the gains and losses that entails.

And, of course, knowledge of the character of one's life experiences after the retinal operation, or of one's life experiences out in the world after growing up in a black and white room, along with one's preferences given those experiences, is precisely what is wanted for authentic decisions in these particular cases.

Campbell's point about authentic decision-making, then, is that in some cases, in order for your first-personal decision to be authentic, your imaginative assessment of your future must be knowledgeable. In such cases, the authenticity of the decision depends on a grasp on the nature of the world as well as a grasp on the nature of your self. As Campbell puts it:

---

[22] I make no assumptions about whether the discovery is, in fact, purely qualitative. My point is that the argument will still go through under these conservative assumptions.

"Unless you factor in that the possibility being offered to you is one in which you gain *knowledge* of some aspects of your environment, you've missed a principle factor that you ought to be taking account of in your assessment."

The point here is that while there is an essential role for the first-personal qualitative element to authentic decision-making, authenticity can require more. Authentic decision-making can require imaginative knowledge of what my future circumstances will be, where such imaginative knowledge carries with it a direct affective, emotional engagement that allows me to cognitively and emotionally empathize with my possible future selves.

2. The lived experience of others

There is another dimension to the subjective value of lived experience that Campbell emphasizes, and that I agree can play a role in our authentic decision-making. This is the subjective value of the lived experiences of other people affected by our decision. For some decisions, when determining the subjective value of an act, we desire to assess the subjective value of the lived experiences of others as well as our own.

This is especially important for decisions made for those dependent upon us, such as decisions made for our children, and decisions that are selfless to some extent, that is, decisions that are to some extent made for the sake of others. If I make a medical decision for my child, my decision is heavily influenced by my assessment of the quality of her future lived experience and the future lived experiences of the rest of the family.[23] If I decide to devote my life to teaching disadvantaged children, my decision is heavily influenced by my assessment of how the childrens' futures would be improved if I took the job.

Part of authentically engaging with the world around you and with your possible future selves can include imaginatively knowing how you understand yourself in relation to others. Such knowledge can be needed for you to first-personally understand how you'll respond to different possible scenarios involving other people, and in this way to grasp a dimension of who you are.

I might authentically decide to devote my life to others, and decide to privilege the positive first-order subjective values of others over my own first-order subjective values, perhaps because I think my selfless act may result in an outcome with a high objective value. I can make this choice authentically based on my imaginative knowledge of what I care about, given my understanding of myself in relation to others. If the choice is

[23] I discuss problems with informed consent and subjective decision-making for others in chapter three and in the Afterword of *Transformative Experience*.

made authentically, my affective engagement with bringing about the objectively valuable outcome can generate a higher-order subjective value that I'd describe as a kind of "meaningfulness".

3. Expert testimony

Campbell raises another problem for an overly internalist picture of authentic decision-making. Such a picture threatens to make it legitimate for Sally to privilege her personal desire to have a baby over the cautions of the experts, simply because the experts couldn't possibly understand what it would be like for her to have a child. If experience is necessary to understand major life changes, is there a tension between authentic decision-making and decision-making guided by scientific expertise?

There is, but authenticity had better not provide a cover for one's desire to follow the dictates of astrology or new age healing. Let me clarify what I intended to argue by examining the puzzle that Campbell poses for me.

Campbell points out that my example of Sally, who privileges her personal desire to have a baby over the cautions of the experts, might seem to be parallel to the case of Billy, who privileges his personal desire to take dangerous drugs over the cautions of the experts. In the example, Billy is clearly wrong to ignore the empirical evidence. But isn't Sally just as wrong?

Yes and no.

In the book, I use the example of Sally to emphasize the idea that unreflectively replacing one's introspective assessment of subjective value with an expert's assessments of subjective value is inauthentic. If a person unreflectively defers to experts when making a transformative choice, she is abdicating her responsibility for her actions in a way that makes her act inauthentic.[24]

However, Campbell is correct to point out that, in a context where the empirical evidence is perfectly clear-cut and clearly specifies the values (and credences) for Sally, it could be inauthentic for Sally to privilege her introspective assessments over what she knows from the scientists.[25] Authenticity in this sort of case requires her to recognize evidence provided by the experts as evidence she should use in her decision. This does not license unreflective replacement of one's introspective assessment with an expert's assessment, but it might license *reflective* replacement of one's introspective assessment with an expert's assessment.

---

[24] I'm particularly indebted to Tyler Doggett here.

[25] I say "could be", because there are in-principle problems with how Sally might be expected to interpret the evidence from the experts that I am eliding for the purposes of discussion. See footnote 26, below, and my reply to Pettigrew.

However, when we restrict the case to one involving transformative experiences where our empirical evidence for the outcomes is incomplete or insufficiently fine-grained, the situation changes, because we can find ourselves unsure about how the empirical results apply. In transformative contexts with incomplete evidence, the problem with replacing our introspective deliberations with an expert's assessment of subjective values has another dimension.[26] On a natural reading of the example, Sally's case exemplifies this additional dimension.

If the evidence against choosing to have a child in Sally's case were as clear-cut as the evidence against taking dangerous drugs in Billy's case (I will assume the evidence is clear-cut in Billy's case, for after all, the drugs are labeled as "dangerous"), then it would be bizarre for Sally to ignore that evidence. But if the situation is anything like it is in the real world, the empirical evidence in Sally's case would be nothing like as complete as we are assuming it to be in Billy's case.

Social science gives us excellent information for decisions involving populations, such as those concerning public policy or institutional guidelines. But when the empirical results can vary for different people, that is, when the evidence allows that more than one outcome is possible for the population of which you are a member, the data may be too general to guide you perfectly, as an individual, to the value for your very own outcome.[27] In fictional cases, we need not worry about the complications of empirical methodology. But in real-world cases of transformative experience, with real-world evidence, we must recognize real-world limitations.

The usual solution to this problem is for the individual to rely on introspection to attempt to close the gap between the empirical results for a population and the results she can expect as an individual. That is, she uses the empirical data in conjunction with introspection on the sort of person she is in order to assess how she is likely to respond to the experience. In such a context, for someone to make a big life decision without considering her

---

[26] In my book, I raise two kinds of problems involving incompleteness for our interpretation of evidence in cases of transformative experience. The first concerns the incompleteness of real-world evidence in terms of observational and external validity: the data might simply be too messy or incomplete in some practical sense, and the epistemic inaccessibility of transformative experience hampers our ability to use introspection to close the gap. For related issues, see Cartwright (2011). The second concerns a distinctive, in-principle problem for the individual who is asked to interpret the utility values given to her by an expert. Before she undergoes the change, we can assume she knows the value that the individual who results from the change will assign to the outcome. But she doesn't know if the self who results from the change is someone she'd first-personally recognize as *herself*, or whether she has been transformed into an epistemically alien self, a self with very different preferences from her current ones. This matters for how she is to interpret the utility value. See my reply to Pettigrew, above, for further discussion.

[27] See my reply to Pettigrew [ref]

introspective evidence *is* bizarre—and inauthentic. (In the book, I also raise the worry that in cases of transformative experience our introspective deliberations are ineffective, but this is a different point.)

So Campbell is correct to point out that truly authentic decision-making may require knowledge of the way the individual relates to the world, and thus may require knowledge of how she will respond to various events and interventions she may undergo. If the individual has perfect or near-perfect evidence of how she will respond, whether that evidence comes via testimony or via reflection, to decide authentically, she should reflectively recognize and employ that knowledge in her decision-making. However, when the evidence is incomplete, simply relying on the incomplete evidence *without* introspectively trying to close the gap is inauthentic, precisely because, in such a case, she lacks relevant knowledge of how she will respond.[28]

References

Barnes, E. (2015). "What You Can Expect When You Donr't Want to be Expecting", *Philosophy and Phenomenological Research*, 91, 775–786.

Briggs, R. (2015). "Transformative Experience and Interpersonal Comparisons", *Res Philosophica*, Vol. 92, No. 2, 189–216.

Campbell, J. (2015). "L. A. Paul's *Transformative Experience*", *Philosophy and Phenomenological Research*, 91, 787–793.

Cartwright, N. (2011). "The Art of Medicine: A Philosopher's View of the Long Road from RCTs to Effectiveness", *The Lancet* (377), April 23, 2011, 1400–1401.

Collins, J. (2015) "Neophobia", *Res Philosophica*, Vol. 92, No. 2, 283–300.

Harsanyi, J. (1977). "Morality and the Theory of Rational Behavior", *Social Research* (44:4), 623–656.

Holton, R. and Langton, R. (1998). "Empathy and Animal Ethics", In Dale Jamieson (ed.), *Singer and His Critics*. Oxford: Basil Blackwell, 209–232.

Johnston, M. (2006). "Better Than Mere Knowledge? The Function of Sensory Awareness", In John Hawthorne & Tamar Gendler (eds.), *Perceptual Experience*. New York: Oxford University Press, 260–290.

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

Pettigrew, R. (2015). "Transformative Experience and Decision Theory", *Philosophy and Phenomenological Research*, 91, 766–774.

Velleman, J. D. (1996). "Self to Self", *The Philosophical Review* 105, 39–76.

[28] I thank Herman Cappelen, Enoch Lambert, Casey O'Callaghan, and Alastair Wilson for very helpful comments. I am particularly grateful to Elizabeth Barnes, John Campbell, Tyler Doggett, Kieran Healy, and Richard Pettigrew for insightful discussions and generous written comments on multiple drafts.