1

Who Will I Become?

L. A. Paul

1 Introduction

Life brings opportunities. Opportunities bring change. Sometimes an opportunity is unexpected. You receive a job offer out of the blue. You fall in love. You get pregnant. Other times, it comes after months or years of hoping and planning. You are admitted to the college of your dreams. You decide to get married. You emigrate to a new country.

Such an opportunity can be the chance of a lifetime. It's exciting. You have the chance to discover a new way of living. You have a chance to make something new for yourself, to fashion a new you.

It's also frightening. Change brings risk. Having a baby will change what you care about and how you'll live for the rest of your life. Going to that fancy college will take you into an unfamiliar, challenging new world where you'll have to find new friends and meet high expectations. Moving to a different country means leaving your home and everything familiar behind. Embracing your new love means betraying your partner and destroying your family. Whether you are ready for it or not, having the opportunity to start a new life opens up a whole new world of possibilities, possibilities where you succeed, but also possibilities where you fail.

There are other kinds of life changes we can face. Not all life changes hold promise for the future. Life brings love and friendship and opportunity, but it also brings loss and misfortune. You get divorced. Your sister is diagnosed with late-stage pancreatic cancer. Your son is killed in a car accident.

All of these experiences, good and bad, chosen and unchosen, can be transformative. Transformative experiences are momentous, life-changing experiences that shape who we are and what we care about. By transforming us, they structure the nature and meaning of our lives and the lives of others. They change us, and in the process they reveal ourselves to ourselves, as we recreate ourselves in response to the experience. They make us who we are.

Part of the power of a transformative experience is that having it involves discovery. Until you have it, you don't know what it will be like. As you have the transformative experience, something new is revealed to you—what it's like to be in that situation or what it's like to have that experience—and as you discover what it's like, you discover how you react to it. You discover how you'll respond, and in particular, who you become, as the result of the experience.

L. A. Paul, *Who Will I Become?* In: *Becoming Someone New: Essays on Transformative Experience, Choice, and Change.* Edited by: Enoch Lambert and John Schwenkler, Oxford University Press (2020).
DOI: 10.1093/oso/9780198823735.003.0002

The way we form and change ourselves through life-changing experiences underlies the way that such transformations define our lives. They shape what we believe and care about, and in this way they make us who we are.

2 Transformative Experience

Transformative experiences, as I define them, affect you in two deeply related ways.

First, they are epistemically transformative: they transform what you know or understand. They do this because they are experiences that are new to you—that is, they are experiences of a new kind, or experiences of a sort that you've never had before, and you have to have this kind of experience yourself in order to know what it's like. By having it, the experience teaches you something you could not have learned without having that kind of experience. When the experience teaches you what that kind of experience is like, and gives you new abilities to imagine, recognize, and imaginatively model possible states involving that kind of experience. Second, such experiences are personally transformative: they transform your preferences. They do this by changing or replacing a core preference, through changing something deep and fundamental about your values. Thus defined, *transformative experiences* are experiences that change you in both of these ways: they are both epistemically and personally transformative.

Leaving home for college can be a transformative experience. Imagine the moment of departure. Your bags are packed. You've said goodbye to your friends. Your family is waiting at the door. It's time to leave. It's time to start this new part of your life, and you couldn't be more excited. The promise of the open future, of having a world of ideas spread out at your feet, the freedom of having control over your own schedule and your own choices, the thrill of meeting new people and exploring new possibilities: you will stretch your mind in unexpected directions as you enter a new and exciting stage of your life. And a part of you knows that, once you go, you can never come back. Even if you come back, the place will be different. The people will be different. Most importantly, you'll be different. You can return to the place and the people you once knew, but it won't be the same. In this sense, leaving now is leaving forever, because you will never be the same.

Moving to a new place with new challenges, and a new kind of life, confronts you with all the possibility, excitement, and risk that a transformative change can offer. As you prepare to take this momentous, life-changing step, you know that a new life is before you, a life that's very different from the life you've lived up to now. What will it be like? Who will you meet? What will you do? How will you change? Who will you become? This sense of the open future captures the way the experience will be epistemically transformative: the experience will change you psychologically, but until you are actually there, having the experiences of college life, you can't really know what it will be like. You might know or be told that you are in for these kinds of changes ahead of time, but actually living it teaches you something you couldn't know ahead of time, and in the process, it changes who you are.

Many of life's big personal decisions concern experiences that are transformative in this way. They involve the real possibility of undergoing a dramatically new experience that will change your life in important ways. If you've never done

something like it before, the experience will be new and different and mentally expansive, and is likely to change what you care about and how you define yourself. If you are making a decision like this, you face a choice. Should you choose to do it, and discover this new way of living your life? Or should you pass up the chance?

3 Decision-Making

On a natural way of thinking about a life-changing decision, making the choice in the right way is a way of taking charge of your own life. Choices involve responsibility, and to choose responsibly, you need to assess how your choice will affect the world and others in your life. Of course, this is your life we are talking about, and so you also need to assess how your choice will affect you more personally. This is because your choices structure your own life experiences and what happens to you in the world.

An ordinary story of how we are supposed to choose responsibly involves making a rational assessment of the nature of each option. You assess the different possible ways you could act and the different possible results of your act. You map out the different ways the future could develop if you act one way rather than another. You think about what the world could be like, and what you could be like, for each way you could choose. You estimate the value of each path you could take, and the likelihoods of the expected outcomes. Of course, you also take into account expert advice and any moral or social facts that bear on the question of what to do. To choose *rationally*, you evaluate the options by weighing the evidence and considering the expected value of each act from your own perspective, and then act in a way that maximizes expected value.

How do we go about assessing the values of the options we are supposed to compare when making these sorts of choices? To determine their values and preferences in such cases, people use different types of reasoning, depending on context and previous experience. Often, they rely on one of two basic types of reasoning that are much discussed in the psychological literature: "model-free" or "model-based" reasoning.

Model-free reasoning involves judging options based on retrospective, memory-based assessments (Crockett 2013). When a person reasons this way, they evaluate future options and possible actions based on cached values they've assigned in the past to similar situations. Such judgments are computationally less demanding. We may be especially likely to reason this way when the possibilities are too difficult or complex to evaluate otherwise.

In contrast to relying on cached values, we can approach decision-making in a different, more expansive way. We can model the hypothetical evolution of our lived experience in response to each possible consequence of each act, in order to assess, judge, and evaluate possible choices. Such "mental modeling" can assist us in developing our point of view in order to determine our preferences concerning the different acts. This is a type of model-based reasoning.

Model-based reasoning is especially useful in high-stakes, deliberative contexts. It is a type of reasoning that

generates a forward-looking decision tree representing the contingencies between actions and outcomes, and the values of those outcomes. It evaluates actions by searching through the tree and determining which action sequences are likely to produce the best outcomes. Model-based tree search is computationally expensive, however, and can become intractable when decision trees are elaborately branched. (Crockett 2013)

Model-based reasoning is what's most naturally used when you want to deliberate about and carefully assess novel possibilities. You can approach the decision by imagining how you'd respond to different events, reverse-engineering your preferences based on your imagined response. You might start by simulating yourself acting in different ways to bring about different hypothetical options, and then, as you imaginatively represent the outcomes of your actions, assign them value. By imaginatively simulating and assessing your possibilities, you can compare them and determine which act will maximize your expected value.

Your reverse-engineering task could be used to discover your preferences about a possible outcome. However, importantly, it might also be used to *create* your preferences about a possible outcome. That is, you might need to imaginatively simulate yourself embedded in various events in order to form value judgments about them in the first place, not merely to figure out what you actually prefer given your antecedent values for engaging in those events. Either task involves a form of model-based reasoning.

Using imaginative representation like this can be very useful and important for rational deliberation. We do it all the time when we make ordinary decisions. For example, when you are considering whether you would rather visit a museum or take a stroll in the park, you might start by imagining yourself in the museum, contemplating a series of paintings, in order to assess the desirability of that outcome. You might then imagine yourself walking in the park and admiring the spring flowers, and use your assessment of the appeal of this option to determine your preferences regarding the choice between a visit to the museum and a walk in the park. If you are deciding whether to go for a swim or to go for a run, you might reflect upon whether it would be unbearably hot to run in the afternoon, while being refreshingly cool to swim. Or it might be numbingly cold to swim in the morning, while being invigorating to run in the cool before the dawn. You plan your daily exercise accordingly.

The same sort of imaginative assessment can be important when we deliberate about major, life-changing decisions for ourselves (or for others). Perhaps you are choosing between two very different types of colleges. If so, you might imaginatively model studying and learning on one kind of college campus, and then imagine doing this on the other kind of campus, and then compare these reflections as you choose where to matriculate. Or perhaps you are pregnant, and you must choose between having the baby or going to college to get an education. To decide, you might imagine life as a parent, with all the joy and sacrifice this entails, and compare it to life as a college-educated person with a wide range of career options. Perhaps, like the French post-impressionist artist Paul Gauguin, you must choose between a life of drudgery and sacrifice where you work to support your wife and family, versus abandoning them for a creative yet self-indulgent life doing what you love. For all of these choices, you might imaginatively model each kind of life you might lead and assess its value,

including its moral value, before deciding what to do. For these kinds of life-defining situations, your imaginative assessment of the value of your future life, as well as the future lives of others who will be affected, plays a central role in the way you determine your preferences about which act to perform.

There is a natural connection here with authenticity: roughly, we can think of authenticity in terms of a relation that holds between our current self and our future selves. If I authentically form my future self, my current self intentionally forms my future self as an extension of who I am now, in a way that is consistent with the values of my "true self." Empathetic imagination of one's future self is an important tool that we can use to authentically form ourselves into whom we want to become.

The existence of this sort of reasoning in situations when people face novel, potentially life-changing decisions is documented in McCoy et al. (2019). In this survey, participants were asked about their preferences for a series of high-stakes, fictional options to have transformative experiences (e.g. you are given a one-time only chance to become a vampire, or, a one-time only chance to travel to alien planets). After reporting their preferences, 75 per cent of people sampled from the standard US population and 53% of the philosophers who took the survey decisions indicated that they had learned something about themselves, suggesting that they had discovered something about their preferences by thinking through the novel scenarios.

Imaginatively exploring how we respond in novel scenarios that could be transformative, then, seems to be an important way of discovering how we'd value them, and thus of discovering various truths about ourselves. Such simulation is an important tool to use to discover our preferences when faced with a choice between hypothetical options. Once we assess the values, we can know our preferences and decide how to act.

But there is a problem with relying on your imaginative capacity to envision possibilities for your future self in contexts of transformative choice: that is, when you are choosing whether to undergo an experience that is epistemically and personally transformative. The problem arises when you want to determine your preferences with respect to these novel, hypothetical options.

If you face a decision involving a choice to undergo an experience that will transform you both epistemically and personally, a "transformative decision," you may not be able to imaginatively assess the nature of your future life. This is because you don't have the right sort of epistemic access to your future self. The problem comes from the fact that you can't accurately imagine or simulate what the transformative experience is like. When an experience is a radically new kind of experience for you, a kind you've never had before, you don't know what it will be like before you try it. But you also don't know what you will be missing if you don't try it. You have to actually experience it to know what it will be like for you. As a result, you can't accurately imagine or simulate what it would be like for you to undergo the transformative experience involved. You are in an epistemically impoverished state, facing a distinctive kind of unknown, because you don't know what the experience will be like.

It's a very special kind of situation to be in. In this sort of situation, you have to make a life-changing choice. But because it involves a new experience that is unlike

any other experience you've had before, you know very little about your possible future. And so, if you want to make the decision by thinking about what your future would be like if you undergo the experience, you have a problem.

Metaphorically, it's as if you face a blank concrete wall, where you can't see what lies beyond. Perhaps you know that whatever happens in the future, past the wall, will involve you somehow. You know you'll be there, in that future moment, living that future experience. But you don't know what it will be like to be that self. As I will describe it, you face an "epistemic wall."

It's the unknowability that creates the problem, because you can't "see" the outcomes. The basic idea is that if you can't properly represent the points of view of the future selves that are the possible outcomes of your choices, you can't accurately imagine these future lived experiences, and so you can't model them in order to assess how you'd value them as the self who is living that experience. In technical terms, your subjective value function goes undefined for these outputs.

To get a sense of how facing transformation involves facing the unknown, imagine the epistemic situation of a congenitally blind man who is about to gain ordinary vision. Like all of us, his lived experience is formed by his way of experiencing the world through his senses. As a blind person, his dominant sense modality is audition, and thus his way of living in the world is highly defined by his sense of hearing and touch. He has never seen a sunset or watched a movie. This will change when he becomes sighted. Until then, before he gains ordinary vision, there is something he can't know: what it will be like for him to live in the world as a sighted person.

Importantly, descriptions and testimony from others aren't enough to teach him what this is like. Think of admiring the color of the sky just after the sun sets. That color has a particular character, and you couldn't accurately describe what it looks like to him if he'd never had this kind of experience. You could use metaphors, images, and poetry to try to capture its quality by suggesting evocative comparisons, but unless he's already had the right sorts of color experiences, he won't be able to grasp what it is like. For you to be able to accurately describe to him what it's like to experience a sensory quality like light pink, he has to have had the right sort of kinds of experiences beforehand. (And even then, you'd have to describe by using comparisons—for example, you'd tell him it looks like a shade of pink he's seen before, or maybe like a lighter shade of a color he's already seen.) This is because descriptive language lacks a certain type of expressive power. As a result, some things can only be communicated through experience.

It isn't just sensory experiences that are like this. Many of life's momentous experiences have a special, distinctive character about them, the nature and quality of the *experience as lived*, that's simply impossible to know about without actually having the experience. It's easiest to see in examples involving the discovery of new sensory qualities, but it isn't confined to them. Other kinds of new experiences can also be like this: for example, living in a world with dramatically new kinds of technology, or experiencing earth-shaking weather events due to climate change, can introduce us to new kinds of lived experiences that we can't know about beforehand.

Even ordinary kinds of experiences, if they are new to you, can be imaginatively inaccessible before you have them. Sometimes this is because they involve kinds or

combinations of sensory qualities that are new to you. If the new sensory contribution can't be isolated, or somehow pulled out and separated from the rest of the experience, then to grasp the nature of the lived experience you must actually undergo it, because the sensory contribution is an essential element of the overall lived experience.

For example, think of the distinctive feeling of being in love. Somehow, being in love is made up of a blend of emotion, belief, and desire, and this gives rise to a distinctive kind of experience with a distinctive kind of feeling. Being in love is partly composed of sensation—that is, it consists partly in an experience that has a particular kind of feel or quality, a feel that is inextricably bound up with the experience of being in love. The distinctive nature of the experience of being in love arises, at least in part, from the contribution made by the feeling involved. You couldn't subtract this experiential element out of being in love and still be in love, yet (despite what some popular songs might claim) being in love isn't just a feeling. It's an experience that involves feelings, beliefs, desires, and other rich mental states that constitute the relation you stand in to your beloved. But the phenomenology is still necessary, even if it isn't sufficient. If you've never been in love, you don't know what it is like, and my descriptions here won't be able to teach you. You can know all the things about love that philosophy and science can tell you, and still, when you fall in love for the first time, you'll learn something new. The nature of this complex experience can't be captured with descriptions any more than descriptions can capture what it's like to see light pink. So if you don't know what it's like to be in love, there's an essential element of the nature and value of being in love that you can't appreciate.

It isn't just the character of experiences like love, fear, awe, and joy that defy description. A person leaving for college can be in the same epistemic boat. They can get descriptions and stories from others about what it will be like for them to start this grand new phase of life, but before they actually leave home and start their new life, there is often a basic and extremely important sense in which they can't know what they are in for.

Moreover, what you discover when you have new kinds of life-transforming experiences isn't just the nature of the new experience. You also discover how you change in response to it, that is, you discover who you become as the result of that new experience. The epistemic transformation changes the way you think and what you care about, and this translates into a new way of understanding yourself and the world around you.

Distinguish between a person persisting over time and the series of selves that realize the person (perhaps realized in turn by a series of more fine-grained temporal parts). On this model, we can think of the new kind of experience as changing a person's life through creating, through the fire of epistemic change, a new self, a new realizer of the persisting individual.

4 Facing the Epistemic Wall

The fact that mere descriptions of experiences can lack expressive power means that when you face a new kind of life-changing experience, you can't rely on descriptions

and testimony from others to learn everything you need to know about who you will become as the result of that experience.

You can learn a lot of things beforehand: for example, you can learn a lot of things about falling in love or leaving home to go to college. But actually knowing what the nature of this new kind of experience will be like for you, and by extension what your new lived experience will be like, remains elusive. Because you can't learn from others about what the new kind of experience is like, you can't learn enough to know what it will be like to be the new self that this experience will make you into.

It's the combination of epistemic with personal change that makes this elusiveness worrying. The elusiveness of minor, non-life-changing experiences, like trying a new kind of cereal, isn't something that we worry about. Such experiences are not personally transformative: they don't change who you are. If a new experience isn't a big deal for you, it's easy to skip it or just try it for the sake of discovering what it's like. Trying a new kind of food or reading a new kind of book is like this. If you don't like it, you just move on. If you pass on trying it, it wasn't that important anyway. The epistemic change doesn't scale up into self-change.

A life-changing experience is a much bigger deal. When the new kind of experience is both epistemically and personally transformative, having such an experience is a game-changer. Think about it this way: when the blind man's experience changes in dramatic ways, who he is as a person will also change. But because the experience of becoming sighted will be personally transformative as well as epistemically transformative, his future self is epistemically foreign to him. If he can't know what his future lived experience will be like, there is a deep sense in which he is alienated from his future, sighted self. As I will put it, his epistemic wall generates a *self-alienation problem*.

The problem of being alienated from one's future and possible selves arises for all transformative choices, both chosen and unchosen. Given the desirability of reasoning rationally when making high-stakes, life-changing decisions, this creates special difficulties for practical deliberation using model-based reasoning. The problem, very simply, is this. You can't know, for some hypothetical future, what it would be like to be the self you'd become in that future. So you can't accurately imagine or simulate this future self. This means you cannot construct an accurate internal model of this lived experience in order to assign it value and determine your preferences. Thus, you've lost one of the main cognitive tools you have for mapping your way through your possible futures and constructing a deliberative response to the choices you face.

Imagine you are facing a choice of whether to undergo a transformative experience. (Alternatively, imagine facing a situation where you are forced to choose between different possible transformative experiences.) In this situation, you don't know what it will be like to have the transformative experience you are making a decision about. This means that you can't accurately imagine or first personally represent what the nature of the lived experience will be like in a way that allows you to imagine who you'll become as the result of the transformation.

If you can't imaginatively represent this possible future lived experience, you can't assess its subjective value—that is, you can't assess the experiential value of the nature and character of this future lived possibility, and thus you cannot determine your preferences. You lack the ability to imaginatively simulate the transformative

experience and the future self it could create. As a result, you cannot represent yourself in the way you need to in order to form value judgments about that self or decide which self you prefer to be.

We can draw out the nature of the self-alienation problem in the context of a thought experiment showing how model-based reasoning breaks down when we lack epistemic access to the subjective values for the possible outcomes.

5 Choosing to Have a Child

Imagine yourself in the following situation: you and your partner are trying to decide whether it's time to start a family. In particular, you are trying to decide whether you'd like to have a baby. Your financial situation and physical health make the decision to become a parent largely up to what you choose—you have the necessary resources, so it's about what you want your future lives to be like. This is a paradigmatic "big decision": the stakes are high, and the choice is irreversible in the sense that, once you've had the child, you can't undo its existence. Even if you give your child up for adoption, you've still become a biological parent.

There are many ways to approach a big decision like this, and, if you have any uncertainty about what you'd prefer, you'll want to think carefully about what you value in order to make the best choice for yourself (and your partner).

Model-based reasoning, where you think about each way you could act, build out the likely consequences of each possible action, and then evaluate and compare these consequences, is the natural way to approach this high-stakes deliberative task. To deliberate, you assess your possibilities to compare them and create (or discover) your preferences. If you can accurately assess the expected value of each act you might perform and compare these values, then when you choose the act that maximizes your expected value, you are choosing rationally.

How are you to assess the expected values of different ways to act? First, you have to be able to assign values to the possible consequences of your actions. Crockett and Paul (forthcoming) show that many people, when they are uncertain about a transformative decision, want to find out what their possible futures would be like. In a study asking people how they would evaluate the possibility of becoming a parent, regardless of whether participants leaned strongly toward wanting children or not, being uncertain about that preference significantly increased the likelihood of wanting to take a (magical) transporter that would allow them to visit a hypothetical future for 24 hours to discover what it would be like to have their child. For those participants who did want children but were uncertain about it, a whopping 96 % of them indicated that they would pay to take the transporter (Figure 1.1).

When a magical transporter to the future is not available, the imaginative simulation of possible futures is a natural substitute.[1] When considering whether or not to become a parent, assessing your possibilities by imaginatively simulating what it would be like for you to have a baby seems like the most natural way to approach the task of assigning value in order to compare alternatives.

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

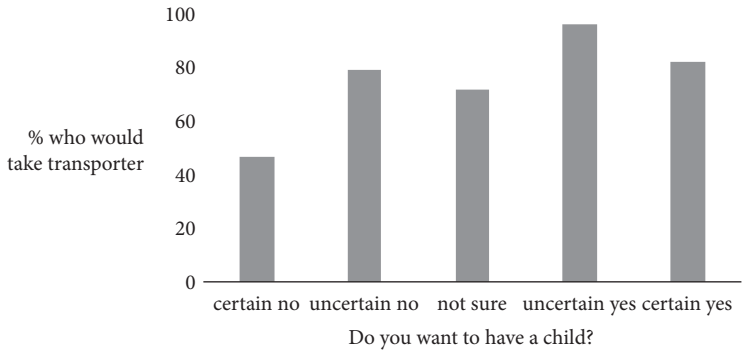

Figure 1.1 Crockett and Paul (forthcoming) found that uncertainty about the preference for having children was significantly correlated with stated preference to take a magical transporter allowing one to visit the future and discover what this is would be like.

The trouble is, for many people, becoming a parent is transformative (Paul 2015b). Having your first child can transform you both epistemically and personally, and for this reason it is an experience that you need to actually undergo in order to know what it is like. When you actually hold your newborn in your arms, you can have an experience like no other, and this can change you in ways that change some of your deepest preferences.

There are, in fact, at least two new kinds of experiences involved: physically gestating and giving birth to your baby, and forming the loving parent–child attachment bond with your child. The second kind of new experience is one that all parents can have (and for women who have the first kind of experience, it can feed into the second one). These new experiences create a new kind of love in a person, a kind of love that's different from romantic love, and different from the kind of love that you can feel for parents and other family members. It brings intense joy, as well as a new capacity for suffering and vulnerability.

Forming a loving parent–child attachment relation is the source of the foundational shift that parents experience with respect to what matters most to them. Many parents shed their old selves and create new ones, forged by the deep and powerful love they feel for their baby.

The shift involves a core value change. Before they become parents, many people prioritize things like their career, financial success, or social achievements. After they become parents, they care more about their child than anything else. Selfishness turns into selflessness. Put more precisely, most of us have a core preference to pursue our own interests, and this preference is replaced with a preference to pursue our child's interests above our own. This is expressed most keenly by a change in our natural instinct for self-preservation. In a life-threatening situation, while we would want to help others, many of us would save ourselves first, or at least *think* about saving ourselves first. All of that can change when you have a child. The child (at a level that feels almost instinctual) comes first, even at your own expense. Speaking personally, until becoming a parent, I had never truly understood what it was like to love

someone selflessly enough to be willing, without a moment's hesitation, to sacrifice my life for them. This is just one of the deep and important ways that becoming a parent changed me, and it was a type of change in how I understood myself that I was unable to anticipate until after I became a mother.

The situation is perfectly analogous to that of the blind man before he becomes sighted. The epistemic wall blocks his imaginative capacity to project himself forward into his future self. Given who he is now, he cannot imaginatively see who he will become. In the same way, prospective parents know, before the baby arrives, that a lot is going to change. And yet, they can't know how things will change, in the important experiential sense of truly understanding what their new lives will be like. This means that they cannot prospectively assess their future lived experience as parents. Their epistemic wall blocks their ability to use their imaginative capacities to project themselves forward into their future lives.

If, before you've had your child, you lack the ability to accurately imagine what it will be like, you cannot accurately imagine and assess the nature of this possible lived experience. If you can't do this, you can't use your imaginative abilities to assess the subjective value of this outcome: that is, before you actually have your baby, you cannot assess the value of what it will be like. You can't (accurately) mentally evolve the world forward in order to imaginatively assess what it would be like to have your baby. Therefore, you cannot compare this subjective value to the value of what it would be like to live a child free life. Your subjective value function, which takes as inputs various lived experiences and gives as outputs their respective values, goes undefined at a crucial point.

Epistemic walls create two serious problems.

The first problem, the problem of *undefined subjective value*, stems from the epistemic transformation involved. It arises for any theory of decision that requires you to assess and compare values in order to maximize your expected utility. If you can't assign the relevant subjective values, you can't compare them to decide which life choice is the best one for you.

The second problem, the problem of *self-alienation*, stems from the way that the epistemic transformation is the source of the personal transformation. It arises for any view that assumes that rational, reasonable life planning is defined by prospective, informed assessment of one's future possibilities "from the inside," or from the first-person perspective. For many of life's biggest choices, an epistemic wall blocks our psychological access to who we are making ourselves into. A choice to transform becomes, in effect, a leap into the unknown. You make a choice to replace your current self—that is, who you are now—with a new, alien, unknown self.

5.1 Undefined Subjective Value

You might think that the problem of undefined subjective value is easy to solve. Here is one way to solve it for our case of choosing to have a child: forget about deciding based on what you and your partner's future life will be like. Instead, decide to have a child based on something else, such as whether it will make your mother happy, or because you want to pass on your DNA to future generations. This isn't especially satisfactory, but it is an option.

Another way to solve it is to get friends and relatives with parenting experience to tell you what your subjective values will be (Pettigrew 2015). Now, it's not clear that this is really the right way to get the information that you need. How can someone else know what it will be like for you to have your child? Your mother's experience as a parent is likely to be quite different from what yours would be like. There are further issues here. Your child's nature will have a dramatic effect on your experience as a parent. Before your child is born, can your friends and family members really know enough about what your future child will be like to give you accurate advice about what it will be like for you to have that child? Do we ever really know what our child will be like before we actually produce them? (This may be especially true if your child has a serious mental or physical disability, as the nature of your life as a parent will be significantly affected.)

You could decide instead to rely on what we can know from science and medicine.[2] To determine your expected utility from having a baby, you could draw on the statistical data about happiness and life satisfaction for parents. Such data can't tell you directly what your subjective value for having a child would be. What the data can tell you is what the *average* effect (or utility value) would be for any member of a population that is composed of individuals like you.

This average effect, however, is perfectly consistent with wide variation in the values assigned to utilities (including the range of uncertainty) for any particular individual member who is included in this average. With real data, we see such variation all the time. At best, you will be able to infer that your subjective value will be within this range.

It's important to see how limited this information is. In particular, you might hope to interpret the average utility values for a member of your population through the filter of your own introspective assessments in order to get a more precise fix on your utilities. This is what people ordinarily do when they want to inform their decisions using scientific evidence. The trouble is that this is precisely what you don't have the ability to do. You can't accurately introspect, because the experience is epistemically transformative. You must simply accept the average utilities, and therefore accept that you are the average person. What this means is that you must replace your own utility assessment with the assessment that applies to the average person. Importantly, you are not informing or updating your own prior assessment, because (given that you cannot assess the subjective value) you do not have an informed prior opinion.

Put another way, to choose the act that you expect to have the highest subjective 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. This may be the best we can do in real-world transformative contexts. The value that

[2] This sounds appealing, but it's important to note that it's hypothetical at this stage. It's not currently possible for the psychological and social sciences to tell us what your subjective utilities would be for this choice. We have nothing that's even close to good enough data. In fact, for most real-world big decisions at this level, like choosing to have a child, or to emigrate, or to enlist in the Marines, adequately fine-grained statistical data about subjective utilities is unavailable. For a related debate, see my (forthcoming 2020) debate with Paul Bloom in *Rivista internazionale di filosofia e psicologia.*

people attach to the possibility of having a transporter, were such available, to actually visit their future selves suggests that, implicitly, as decision makers, we realize that this is non-ideal.

So the problem of undefined subjective value can be "solved" by eliminating the role of introspection in how you assign your subjective values. You can do this by eliminating subjective value from the decision model, or by assigning yourself a subjective value that is determined solely by the average subjective values of people who are like you in certain ways. "Solving" the problem this way, however, leads us straight into the problem of self-alienation.[3]

5.2 Self-Alienation

Recall that getting testimony about a transformative experience does not teach you what it's like to have the experience: language lacks the expressive power needed to communicate what actually having the experience teaches you. Even if you use the testimony of others to predict the valence and range of your expected subjective value (or if, instead, you dispense with subjective value altogether), when you undergo the transformative experience, you will undergo an epistemic shift that will bring you a dramatically new kind of life. Before you have experienced that epistemic shift, you lack psychological access to your future self. Thus, you lack psychological access to who you are making yourself into until you actually undergo the transformation. As I put it above, you face an epistemic wall.

David Velleman's work on personal identity[4] and persistence 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. (Velleman 2005: 76)

The way self-alienation arises in transformative contexts can be brought out by exploring two case studies of the transformative choice to have a child.

In the first case, imagine that you are suddenly confronted with the decision of whether to become a parent. Perhaps your partner unexpectedly gets pregnant. Or (if you are female), perhaps you get pregnant by mistake. As you deliberate about what to do, you ask around for advice.

Pressure to have a child often comes from well-meaning friends and relatives. They say, rightly, that most parents will say they are very happy they decided to do it (Harman 2009). The odds are, then, that you will be happy that you did it. People who know you may also tell you about their own experiences, thinking that your experience will be like theirs. Your friends tell you that they think you should do it because it's the best thing they've ever done. Your mother tells you that she is sure you will be happy if you have a baby.

[3] I don't think these purported "solutions" are at all satisfactory. For discussion see Paul (2015a, 2015b, 2014).

[4] For further discussion see Parfit (1984) and Callard (2018).

However, you are not convinced. When you demur or raise worries about the way it could negatively affect your current life, friends and family admit that there are costs, and yet they tell you that once you've become a parent, you'll be willing to make those tradeoffs. You are unimpressed. You tell them that you've seen the haggard parents on the local playground, hair askew, smelling of baby vomit and urine, and that you don't find the thought of being a parent at all appealing. In response, they laugh and explain that you won't mind any of that very much once you actually have your baby.

This is a case where you are to solve the unknown subjective value problem by using testimony from others to determine your subjective values. You are to substitute the judgments of friends and family about your expected subjective value for your own judgment. When their judgments are in alignment with your own, the replacement is easy to make. But when the values they assign are in conflict with your own assessment, the inadequacy of this solution becomes apparent, because it exposes the self-alienation the solution creates.

The self-alienation is obscured when your solution to the value problem lines up with your pre-choice, or "ex ante" beliefs and desires. When there is no value mismatch, and the "ex post" self you become is happy enough, it might not matter much that, after the fact, you couldn't first-personally grasp who you would become. After all, both the ex ante and the ex post selves agree on the choice, and now that it's done, there isn't anyone around to be unhappy about the loss of that old self.[5] But when the self before the choice, the ex ante self, prospectively disagrees with the self that is to be created by the choice (the ex post self) but nevertheless is expected to choose to become that self, the alienation of the ex ante self from the ex post self becomes apparent (Paul and Quiggin 2018).

The issue is not that the judgments of your friends and family members are incorrect. Assume their judgments about your future subjective values are in fact correct. (If their judgments are false, you haven't solved the unknown subjective value problem.) The trouble is that, at the time of choosing, their judgments about the best choice to make conflict with your judgment. The choice that is deemed "rational" by their testimony cuts deeply against what you want and believe now. This means that, to choose rationally according to your friends and family, you should choose to become a parent, even though you don't want to. The lesson is that, to be rational, it doesn't matter what *you* think. What matters is what the people who know you (especially the people who are already parents) think.

This seems bad. But why? After all, choosing rationally is the choice that maximizes your expected value. Their testimony guides you to the rational choice. Why isn't this obviously and uncontroversially the best thing to do?

Part of what's bad is that you are trading in your autonomy for the sake of your rationality. Your solution to the unknown subjective value problem—one that relies on others to tell you the subjective value of your future life—eliminates an important part of your role in your value assessment. For one of the most important and

[5] And as luck would have it, the human psyche is enormously successful at adapting its preferences to be happy with whatever outcome it finds itself in.

personal decisions of your life, you are forced to rely solely on the judgments of others.

The real problem, however, isn't *merely* that you have to rely on others to make one of your most important personal life choices. That's just what leads to the real problem. The real problem is that, to be rational, you must make this choice by rejecting what you care about. In order to choose rationally, you must to choose to become a self that is alien to who you are now. Solving the value problem through relying on testimony from others creates the self-alienation problem.

The self-alienation you face, at its root, stems from the deep epistemic change you will undergo. When there is a value mismatch between your ex ante and your ex post selves, the solution to the unknown subjective value problem alienates you from your choice and, by extension, from the self you are choosing to become. The self you will become is first-personally foreign to who you are now. And it is because this self is so alien that you face the unknown subjective value problem in the first place: you can't imagine yourself into that self's perspective in order to assign value to that new lived experience. The deep epistemic change of the transformation is the common source of the unknown subjective value problem and the self-alienation problem, and this is why merely solving the value problem won't eliminate the alienation problem.

We can explore a second case in order to draw out the formal structure of the self-alienation problem. In the second case, we flip the results. Instead of being skeptical, you find yourself incredibly keen to become a parent. You've read novels and seen films about how joyful and satisfying it can be to have a family. Your sister just had a baby and she can't stop talking about how happy she is. You've always wanted to become a parent, and feel that having a baby would make your life fulfilling and meaningful.

However, your friends and family counsel you to avoid becoming a parent. Perhaps they think that you and your partner are not ready, or that you would be unhappy as a parent. When you consult the scientific evidence, it also goes against you: it tells you that the average subjective value for people like you is negative. The evidence suggests that the quality of your future life will decline if you become a parent. Again, to choose rationally, you must substitute the judgments of others in place of your own. Let's assume that you believe in the science, and in the wisdom of your friends and family members. So you accept their assessment, 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. This is simply the bargain you must make in order to solve the unknown subjective value problem.

Let's add detail to see how the reasoning might go. You have precise credences for each of the relevant hypotheses and their associated outcomes: having a baby versus not having a baby. You consult the best scientific sources available, and the research clearly tells you that you can expect a low utility if you have a child, and a high utility if you don't. (Friends and family agree with this result.) In effect, the science tells you that you will maximize your expected utility by choosing to remain childless, even if you are uncertain about just how much.

You can't understand this intuitively, because although you don't feel like you have a detailed grasp on what the future would be like (everyone tells you to expect a

dramatic life change), 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, and it "feels right" to you to have one.

Given that it's not rational to choose to act in a way that does not maximize your utility, then according to the expert's assessment of your utilities, you can't rationally choose to have your child, even though this conflicts with your personal assessment. In this situation, to be rational, you must allow the expert determination of what you are to believe about your utilities to replace your introspective assessment of your heart's desires.

How does the evidence predict your expected utility? Let's walk through the way the counterfactuals work. To assess your utilities in different possible circumstances, we start by 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. We assess your utilities in different possible worlds because we are assessing what the actual world would be like under different possible changes of state, viz. having a baby or not having a baby. (Recall that, before you have a baby, because of the transformative nature of the experience, world W1 at time t2, where you have a baby, is epistemically inaccessible to you. However, we assume the science tells you what utility to expect in those circumstances.)

Do you exist in W1 and W2? Yes—or at least your respective counterparts do. Call the person who exists in W1 at t2, "C1" and the person who exists in W2 at t2, "C2." The self-alienation problem rises with C1, the person who is identical to you (or the counterpart who represents you) in W1 at t2.

Here is the root of the problem. Normally, when we counterfactually assess the value of a state change for an agent A, the salient dispositions and preferences of A are kept fixed in order to assess A's proposed utility in the new state. In other words, we preserve *act–state independence*.[6] After all, at t1, we are considering what A wants, and trying to assess whether a potential change in circumstances (a change in the state of the world) suits A's preferences. If we are interested in what the value of the change would be for agent A, then we want to compare A (ex ante) in their current circumstances to A in their new (ex post) circumstances. If A's preferences also change when there is a change in circumstances, then at t2 we aren't getting information about whether the proposed change suits A's current (ex ante) preferences.

In other words: when you prospectively assess what the value of a change in circumstances would be for you, you want to compare yourself in your current circumstances to your counterfactual self in your new (changed) circumstances. But if *who you are also changes* in the new (counterfactual) circumstances, finding out your future utilities (from science, or via testimony) in those future circumstances doesn't tell you whether the change fits who you are right now ex ante (that is, at the time of choosing). The problem is that act–state independence has been violated.[7] Example: I might be happier if I had a frontal lobotomy tomorrow. After

[6] This is often characterized as an "independence" axiom or theorem.

[7] For a developed discussion of this problem, see Paul and Healy (2016).

the lobotomy my ex post self might even testify to my new preference to have been lobotomized, finding it quite pleasant indeed. But right now, I (ex ante) definitely don't want to get a lobotomy! I feel quite sure I, ex ante, should disregard the preferences of that hypothetical lobotomized ex post self. That prospective future self's testimony is not relevant to me, precisely because act–state independence has been violated in the creation of that self.

Choosing to have a child puts you in an interestingly similar position.[8]

For the change modeled by W1, becoming a parent, act–state independence *fails*. This is because, by hypothesis, the state change represented by W1 at t2 (you with your baby) does not exist in isolation. The change you undergo by having a child is transformative. That is, changing the state of the world to make you into a parent would *also* change your preferences and your psychological capacities. If you are (or are represented by) C1 in W1 at t2, in W1 you are a person with a radically changed first-person perspective.[9]

We can put it this way: at t1, in @, when you consider the choice to have a baby, from your first-person perspective, C1's point of view is psychologically alien to you.[10] 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.[11]

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.[12]

So the utilities that the science and the anecdotal testimony predict you'll have in W1 are not the utilities of anyone you can recognize as your future self. They are indeed the utilities of C1 at t2, but from your first person perspective at t1, C1 is *not you*.[13] 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, 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 choices, you want to have psychological access, in an anticipatory or imaginative way, to each of your possible future selves.

[8] See Barnes (2015).

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

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

[11] This brings out an interesting mismatch between how we think about personal identity from an impersonal, bird's-eye point of view, and how we think of it from within. Bernard Williams (1970) captures this mismatch in his discussion of conflicting intuitions in his "The Self and the Future."

[12] On some metaphysical accounts of personal identity, the *metaphysically same person* relation merely requires the right sorts of causal or other sorts of continuity. The point here is that *metaphysically same person* and *same self* are different relations, and the one that matters in these decision contexts is the same self-relation. We can think of this as a problem of personal ambiguity, as opposed to a problem of personal identity.

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

For each possible choice, you want to grasp the first-person perspective of the self who you think you could become, and who will live with the result of your choice.[14]

Because, from your first-personal perspective at t1, C1 is not you (or, if counterpart theory is preferred, we can say that from this perspective C1 is the wrong counterpart), relying solely on testimony to tell you about your future utilities creates alienation from your possible future selves. You are psychologically alienated from who results from this change.

(Perhaps in the strict metaphysical sense you are the same person after having your child, just like you are the same person now as you were when you were three years old. But in another, very important sense, you aren't the same person—that is, you are not the same self, and this may be what people mean when they say "I'm not the same person I used to be." The metaphysics here involves distinguishing "same self" from "same person." Think of a person as composed of a series of selves over time, and of those selves in turn as composed, most fundamentally, of a series of temporal stages, or temporal parts. We can then mark different requirements for being the same person in some literal or strict sense at different times and for being a different self over time (and further, we can distinguish between different temporal parts of the same self, or of the same person, at different times). The important distinction, then is between being very different selves over time and being literally a different person at different times. Strictly speaking, I'm still the same person: after all, I have the same birth certificate and the same parents. I'm just not the same self. I've changed in core ways, and colloquially, that's usually what we mean by saying "I'm a different person." In this way, a person can be literally the same person over time while also being composed of a chain of changing selves, which is in turn composed of a chain of temporal parts. Each link of the chain is a different part, and the links summed together compose the person over time.[15])

The point of all this isn't that you shouldn't have a baby. Maybe you should. The point is that choosing to have a child, like choosing to become sighted after a life of blindness, involves facing an epistemic wall, and the implications of this are personally significant. Having a baby can be a transformative experience, and so choosing to become a parent can mean you are choosing to become a different kind of person. You are choosing to become a self who is unknown to you now.

And this is an extraordinarily salient implication of transformative experience and transformative choice. You might think that, when you undergo a transformative change like becoming a parent, you are just realizing a future version of who you are now. But you aren't: you are *replacing* who you are now with some radically different, alien self. That self is you, in some sense. But not in any first-personally accessible sense.

Drawing out the nature of the choice in this way can help us to understand some of the implications of the first case we considered, where you don't want to become a parent, but everyone advises you to have a baby. In that example, testimony from those who have children seems to suggest that, even if you can't really understand

[14] 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.

[15] For relevant discussion see Parfit (1984), Paul (2017), and Pettigrew (2020).

what it's like to be a parent, you should do it anyway, because you'll maximize your expected utility. In that case, you admit that parents are in general happy with their choice, and you even admit that, if you became a parent, you'd probably be happy with your choice, even if, right now, you don't want to be a parent.

In such a case, what should you infer? That is, what if your friends are right that, like them, you'd be very happy and satisfied as a parent? What if your mother is right that, after having your baby, you won't care about all the things you care about now? Does this mean that, deep down, you'd really prefer to be a parent, even if you can't know it now? Does this mean that your mother really does know you better than you know yourself?

The suggestion seems to be that your friends and family members understand something you don't. Once you become a parent you'll understand that they were right all along. In more technical terms, the implication seems to be that, once you have your child, your underlying preference to become a parent will be revealed. You can't understand this ahead of time, but that's because the experience is epistemically transformative.

I reject this. To start: why is this sort of well-meaning advice so irritating? It's even more upsetting when science is brought to bear. When the scientific evidence supports a choice that you intuitively reject, the implicit suggestion seems to be that resisting it involves some sort of magical thinking. The suggestion is that you are confused: you think you are unique or that the science doesn't apply to you, but if you had a proper understanding of the way empirical evidence worked, you'd know better. Aspersions are thus cast on your ability to understand the nature of statistical inference or to make rational decisions.

What is it that feels so objectionable here? Is it merely the smug paternalism of such suggestions? Are you simply wrong, and unable to face this fact? No.

Again, assume your friends and family members are, in fact, correct in the conclusion that, if you become a parent, you'll be happy with your choice. And assume there is scientific evidence to support this conclusion. But this fact, considered alone, does not mean that you should choose to become a parent. The philosopher Elizabeth Harman (2009) has argued persuasively that "I'll be glad I did it" reasoning can fail in some cases. This seems to be such a case. (Harman might not agree—see her 2015.)

We can see the mistake when we focus on the transformation involved. Recall that, under the state change we are considering, act–state independence is violated. Having a child is not just epistemically transformative: it is also personally transformative. Becoming a parent transforms what matters to you, and can make you happy and satisfied to be a parent. This transformation can happen even if, before you have a child, you really, truly, don't want to become a parent. So, again, does this imply that when you have a child, your underlying preference to become a parent is suddenly revealed?

No. There is another, very plausible explanation that fits your situation: something about becoming a parent eliminates your old preferences and implants new preferences in you. When you are transformed, your preferences are transformed. What you care about is transformed by the process itself. Even if you don't want to become a parent, the process of forming an identity-defining attachment to your child can

create or implant new preferences, replacing your old pre-kid cares with new kid-focused ones. What a person cares about can change, hugely, when they have a child, and this happens in virtue of the psychological and biological changes that make them a parent.

If so, then your concerns about the choice are perfectly legitimate. You are not being perverse. You are not confused. You are not ignorant of your own preferences. Your worry is not about whether you'll be happy with who you've become *after* you've been transformed.

Your worry is that, right now, what you care about—now—isn't consistent with being transformed. Becoming a parent would change you in ways that, right now, you reject. If you do not want to have a child, then, in your current childless state, you don't care about the things you'd care about as a parent, and, even more importantly, you don't want to care about them. You want to preserve who you are *now*, and what you care about *now*. In these circumstances, it's perfectly reasonable to resist the pressure you are getting from the experts. That's because there is no implication that somehow, becoming a parent would be better for the self you are now. Rather, becoming a parent would *replace* the self you are now with a different self, an alien self: a self that, right now, you don't want to become.

So the well-meaning advice from friends and family is too simplistic. Their advice is flawed, because it does not account for the true structure of the problem. For the same reason, the scientific evidence fails to apply in the clean way that the results might suggest. The clean application assumes that act–state independence is preserved. But when act–state independence is violated, it is unclear how one should interpret the statistical results, and thus it is unclear what one can infer about what is rational to choose in these circumstances.[16]

What we have here is a first-personal version of a Kuhnian revolution. In transformation, we replace our old point of view, our self-understanding of who we are, with a new, incommensurable point of view, a new self-understanding of who we are. Instead of a conceptual revolution writ large, like that brought on by the discovery that the earth revolves around the sun (which replaced the old idea that the sun and other planets revolve around the earth), we experience a conceptual revolution writ small.

The point can be made another way. When you face a transformative experience, even if friends, family, or science can tell you about the utilities involved, you still face an existential problem, one that they are unqualified to address: Will *you* be happier after the transformative change? Or will you just become someone else?

References

Barnes, E. 2015. "What You Can Expect When You Don't Want to Be Expecting." *Philosophy and Phenomenological Research* 91(3): 775–86.

Callard, A. 2018. *Aspiration: The Agency of Becoming*. New York: Oxford University Press.

Crockett, M. J. 2013. "Models of Morality." *Trends in Cognitive Sciences* 17(8): 363–6.

Crockett, M. J., and L. A. Paul. forthcoming.

[16] Paul and Healy (2016).

Harman, E. 2009. '"I'll Be Glad I Did It": reasoning and the significance of future desires'. In J. Hawthorne (ed.), *Ethics*, 177–99. Hoboken, NJ: Wiley–Blackwell.
Harman, E. 2015. "Transformative Experiences and Reliance on Moral Testimony." Res Philosophica 92(2): 323–339.
Lewis, D. 1990. "What Experience Teaches." In W. G. Lycan (ed.), *Mind and Cognition*, 29–57. Hoboken, NJ: Wiley-Blackwell.
McCoy, J., T. Ullmann, & L. A. Paul. 2019. "Modal Prospection." In A. Goldman & B. McLaughlin (eds), Metaphysics and Cognitive Science. Oxford University Press, 235–267.
Parfit, D. 1984. *Reasons and Persons*. Oxford: Oxford University Press.
Paul, L. A. 2014. *Transformative Experience*. Oxford: Oxford University Press.
Paul, L. A. 2015a. "Transformative Choices: Discussion and Replies." *Res Philosophica* 92(2): 473–545.
Paul, L. A. 2015b. "What You Can't Expect When You're Expecting." *Res Philosophica* 92(2): 149–70.
Paul, L. A. 2017. "The Subjectively Enduring Self." In I. Phillips (ed.), *The Routledge Handbook of the Philosophy of Temporal Experience*, ch. 20. Abingdon: Routledge.
Paul, L. A., and K. Healy. 2016. "Transformative Treatments." *Noûs* 52(2): 320–35.
Paul, L. A., and J. Quiggin. 2018. "Real World Problems." *Episteme* 15(3): 363–82.
Pettigrew, R. 2015. "Transformative Experience and Decision Theory." *Philosophy and Phenomenological Research* 91(3): 766–74.
Velleman, J. D. 2005. *Self to Self: Selected Essays*. Cambridge: Cambridge University Press.
Williams, B. 1970. "The Self and the Future." *Philosophical Review* 79(2): 161–80.