

# The Problems of Transformative Experience

Yoaav Isaacs

## 1 Introduction

Decision theory is uncontroversially well suited for artificially simple decision problems. If you want to know how to bet when a randomly-selected marble is drawn from a randomly-selected urn, you need look no further than decision theory. The suitability of decision theory for more complex, realistic decision problems is more controversial.

Laurie Paul has recently argued that transformative experiences pose a problem for decision theory. According to Paul, agents facing transformative experiences do not possess the states required for decision theory to formulate its prescriptions. Agents facing transformative experiences are impoverished relative to their decision problems, and decision theory doesn't know what to do with impoverished agents.

Richard Pettigrew takes Paul's challenge seriously. He grants that decision theory (in its traditional state) cannot handle decision problems involving transformative experiences. To deal with the problems posed by transformative experiences, Pettigrew proposes two alterations to decision theory. The first alteration is meant to handle the problem posed by epistemically transformative experiences, and the second alteration is meant to handle the problem posed by personally transformative experiences.

I argue that Pettigrew's proposed alterations are untenable. Pettigrew's novel decision theory faces both formal and philosophical problems.<sup>1</sup> It is doubtful that Pettigrew can formulate the sort of decision theory he wants, and further doubtful that he should want such a decision theory in the first place. Moreover, the issues with Pettigrew's proposed alterations help reveal issues with Paul's initial challenge to decision theory. I suggest that transformative experiences should not be taken to pose a problem for decision theory, but should instead be taken to pose a topic for ethics.

---

<sup>1</sup>I will explicitly distinguish the formal problems from the philosophical ones. The more serious problems are the philosophical ones, as the formal problems can be palliated by formal means. Readers who are uninterested in formal issues should feel free to skip the formal sections of this paper.

## 2 Epistemically Transformative Experience

An epistemically transformative experience is one which provides information about the phenomenal character of an experience that can only be had by having that experience. Examples of such experiences abound. You can't know what it's like to see red without seeing red. You can't know what it's like to eat Vegemite without eating Vegemite. You can't know what it's like to have a child without having a child. And so on.<sup>2</sup>

According to Paul, an agent who must choose whether or not to have an epistemically transformative experience faces a difficult problem. "Epistemically transformative experience, viewed through the lens of epistemic decision theory, generates a problem. How? Because the transformative nature of the experience rules out the possibility that the agent can subjectively grasp her utilities, since she cannot 'see' these utilities or entertain the relevant propositions."<sup>3</sup> Since the agent hasn't had the epistemically transformative experience yet, the agent has no idea what he'd be getting himself into by choosing to have the experience. And since the agent has no idea how to evaluate the epistemically transformative experience, the agent has a difficult time making a rationally acceptable decision.

Pettigrew follows Paul in taking the problem posed by epistemically transformative experiences to concern an agent's utilities. Pettigrew notes that in order for him to make a decision in accordance with standard decision-theoretic norms he would need utilities defined over the possible outcomes of his decision. When deciding whether or not to undergo an epistemically transformative experience, he would need utilities defined over experiences he cannot fathom.<sup>4</sup> According to Pettigrew, "The problem is that... I do not have access to these utilities."<sup>5</sup> Pettigrew takes such access to be required for decision theory to function. "[P]resupposition: In order for an agent to follow this procedure, she must have access to her probabilities and to her utilities."<sup>6</sup>

Pettigrew's proposed solution is to redescribe decision problems involving epistemically transformative experiences with a finer grain. Instead of having an agent

---

<sup>2</sup>It remains to be seen how strictly the requirements for a transformative experience should be construed. It is, for example, a substantive question whether a Davidsonian swampman could know what it's like to see red. For more, see Davidson (1987).

<sup>3</sup>Paul (2015).

<sup>4</sup>To make a decision about whether or not to see red, he would need utilities defined over the various ways of seeing red and the various ways of not seeing red. To make a decision about whether or not to eat Vegemite, he would need utilities defined over the various ways of eating Vegemite and the various ways of not eating Vegemite. To make a decision about whether or not to have a child, he would need utilities defined over the various ways of having a child and the various ways of not having a child.

<sup>5</sup>Pettigrew (2015).

<sup>6</sup>Pettigrew (2016).

evaluate possible states of the world, Pettigrew suggests that an agent should consider conjunctions of possible states of the world and possible utilities for those states of the world.

[I]nstead 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  $u_1, \dots, u_n$  be the utility functions I might take 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:

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

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

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

Doing this solves the original problem of epistemically inaccessible utilities.<sup>7</sup>

In short, expected values for utilities do the work that would have been done directly by utilities, had those utilities not been problematically inaccessible.

It is not quite clear what Pettigrew has in mind when he talks about the epistemic inaccessibility of an agent's utilities, and (relatedly) it is not quite clear how he intends his proposed solution to work. I see two possible interpretations, one epistemological and one ontological. On the epistemological interpretation, agents *have* utilities concerning the experiences which they cannot yet fathom; the inaccessibility consists in the agent's being *uncertain* of what those utilities are. The expected value that Pettigrew defines would then be an expected value of the agent's current utility. On the ontological interpretation, agents *do not have* utilities concerning the experiences which they cannot yet fathom; the inaccessibility consists in the agent's *lack* of such utilities.<sup>8</sup> The expected value that Pettigrew defines would then be an expected value of the utility the agent would have after undergoing the epistemically transformative experience. But under either interpretation Pettigrew's proposal is untenable.

---

<sup>7</sup>Pettigrew (2015).

<sup>8</sup>Passages concerning a lack of access to certain utilities or an inability to see certain utilities suggest that the relevant utilities exist. But there is contrary textual evidence as well. For example, Paul writes, "[A]n agent without a value function for transformative outcomes is an agent without a (standard) model for a rational decision."

## 2.1 Epistemological Interpretation: Uncertain Utilities

If the problem posed by epistemically transformative experiences is that agents are uncertain about the relevant utilities, then Pettigrew's proposed solution faces three problems: the first problem concerns the uselessness of utilities given this proposal, the second concerns a role of access in decision theory, and the third concerns the difficulties inherent in averaging across different utility functions.

### 2.1.1 Philosophical Problem: Useless Utilities

Under this interpretation, agents have all the utilities needed for normal decision-theoretic analysis. Agents who have never seen red have utilities for seeing red. Agents who have never eaten Vegemite have utilities for eating Vegemite. Agents who have never had children have utilities for having children. And so on.<sup>9</sup>

For clarity, let's think about how the epistemological interpretation works in a particular case: Marie is deciding whether or not to have children. There are myriad possibilities in which she has children and there are myriad possibilities in which she does not have children. Marie has credences about which possibility will obtain if she chooses to have children and she has credences about which possibility will obtain if she chooses not to have children. Marie has utilities about each of the possibilities in which she chooses to have children and she has utilities about each of the possibilities in which chooses not to have children. Therefore the expected utility of having children and the expected utility of not having children are each well-defined, and thus there is a fact of the matter about which choice maximizes expected utility.<sup>10</sup> Because having children would be an epistemically transformative experience for Marie, she must be uncertain about her utilities in the possibilities in which she has children. It is also possible—whether due to other epistemically transformative experiences or for other reasons—that Marie is uncertain about her utilities in the possibilities in which she does not have children. So Marie will not be certain what her expected utilities are. But a few facts are notable. First, it's consistent with this scenario that Marie is certain about which choice maximizes expected utility. Uncertainty about fine-grained matters is compatible with certainty about coarse-grained matters. (For example, an agent might be uncertain about how much

---

<sup>9</sup>Note that this does not presuppose that agents assign utilities appropriate for the actual character of the epistemically transformative experience in question. It's not as though an agent has to assign a utility that corresponds to the actual, unknown taste of Vegemite. Similarly, it's not as though an agent has to assign a utility that corresponds to the actual, unknown effect of pushing a button. Instead, with both Vegemite and buttons, an agent can assign varying utilities to any possibilities that his credences differentiate. The unfathomability of an epistemically transformative experience might be taken to preclude such varying credences and utilities, but such a view takes the ontological interpretation rather than the epistemological interpretation, and will thus be treated in the next section.

<sup>10</sup>Barring a tie, in which case both choices maximize expected utility.

money Bill Gates has but still be certain that Bill Gates has more money than she does.) Second, even if Marie is uncertain about which choice maximizes expected utility, that doesn't change the fact that one of the choices does maximize expected utility. Marie might therefore be uncertain about which choice is mandated by traditional decision theory, but traditional decision theory would still mandate one of the choices.<sup>11</sup>

On the epistemological interpretation, Pettigrew allows that agents making decisions concerning epistemically transformative experiences have everything needed for a normal decision-theoretic approach<sup>12</sup>, but nonetheless wants to depart from the normal decision-theoretic approach. There's no need to alter the mechanisms of decision theory if they are in a position to function properly. As the platitude goes—if it ain't broke, don't fix it! On this interpretation of epistemically transformative experiences, Pettigrew's departure from decision-theoretic orthodoxy is unmotivated.

Moreover, Pettigrew's proposed heterodoxy introduces new problems. Note first that the utilities that agents have regarding transformative experiences do no work; they cannot influence any decision whatsoever.<sup>13</sup> Two agents with the same credences and with utilities that differ only regarding transformative experiences are guaranteed to be decision-theoretically equivalent. Regarding transformative experiences, Pettigrew proposes that expectations of utilities serve the role traditionally served by utilities. The expectation of a utility is a weighted average across all the possible values of the utility, those weights being determined entirely by the agent's credences. Thus which actions the agents should perform cannot differ, even though they have different utilities regarding transformative experiences! Pettigrew would make utilities regarding transformative experiences useless, and useless utilities are implausible.

It seems very strange for utilities to do crucial work in nearly every decision problem, and then drop into irrelevance when a transformative experience comes up.<sup>14</sup> It would make more sense for there to be a uniform treatment—for utilities

---

<sup>11</sup>It may be helpful to think about a very simplified case. Let us act as though there is only one possibility in which Marie has children and only one possibility in which she does not have children. The expected utility of a choice then trivially equals the utility of the relevant possibility. Suppose that Marie's utility for having children is 4 and her utility for not having children is 3. Suppose also that Marie is uncertain about her utilities; she has credence .5 that her utility for having children is 4 and her utility for not having children is 3, and she has credence .5 that her utility for having children is 3 and that her utility for not having children is 4. So Marie is uncertain about which choice maximizes expected utility. It is nonetheless a fact that the choice to have children maximizes expected utility.

<sup>12</sup>A normal decision-theoretic analysis requires only that an agent has credences and utilities. More on this shortly.

<sup>13</sup>Pettigrew's proposal may therefore be incoherent if it's necessary for utilities to serve a functional role.

<sup>14</sup>Note that the scope of the problem posed by epistemically transformative experiences will

to *always* be irrelevant, and for the work thought to be done with utilities to be done by credences about them. But it seems wrong for what an agent thinks to make what that agent wants be totally irrelevant. Decision theory rightly uses both credences and utilities; we can't make do with credences alone.

### 2.1.2 Philosophical Problem: Access

Pettigrew contends that agents need to be able to access their utilities. He writes,

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 relevant dependency hypotheses 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.<sup>15</sup>

It should be noted that Pettigrew's idea of access is not one that arises in the standard framework for decision theory. The standard framework for decision theory is only meant to specify which of an agent's available acts that agent should perform, and credences and utilities suffice for that. The central norm of decision theory—expected utility maximization—is formulated in terms of credences and utilities. Maybe some agent has access to his credences and utilities, and maybe that agent doesn't. But it doesn't make a difference as far as the norms of decision theory go. If the agent acts in a way which maximizes expected utility then he acts as decision theory says he should, and if the agent acts in a way which does not maximize expected utility then he acts as decision theory says he should not. The very notion of access is not part of decision theory; it's just expected utility maximization.

It is important that decision theory not require too much of agents. There may well be agents who have beliefs and desires, but who are not cognitively sophisticated enough to have either beliefs or desires about their own beliefs and desires. This poses no problem whatsoever for decision theory. The agent's credences and utilities suffice to determine what the agent ought to do. Decision theory can easily

---

depend on how decision problems are individuated. If the possibilities under consideration are individuated coarsely, then the problems posed by epistemically transformative experiences will be relatively rare. If the possibilities under consideration are individuated finely, then the problems posed by epistemically transformative experiences will be ubiquitous. Suppose you're facing a very pedestrian decision: whether or not to have a sip of water. If you have just one utility for having a sip and just one utility for not having a sip, then the problems posed by epistemically transformative experiences won't arise. On the other hand, if you have utilities for all the worlds in which you have a sip and for all the worlds in which you don't have a sip, then the problems posed by epistemically transformative experiences will arise, as you have epistemically transformative experiences in many of those worlds.

<sup>15</sup>Pettigrew (2015).

determine whether or not the agent takes the rational means for his ends. This agent won't understand decision theory, but that's no matter. Professional decision theorists have to understand decision theory in order to be professional decision theorists, but agents do not have to understand decision theory in order to be agents. Decision theory says what an agent ought to do. It doesn't say that agents know what they ought to do.

Pettigrew's concern about access cannot be brushed aside, however, even if access is not part of the standard framework for decision theory. Pettigrew thinks that more is needed than is provided by the standard framework, and he may well be correct. Pettigrew has clarified that he thinks that "Paul's challenge applies primarily to the deliberative understanding of decision theory."<sup>16</sup> A deliberative understanding of decision theory is one in which the process by which a decision is made is subject to rational evaluation, and not merely the decision itself. The standard framework for decision theory is not deliberative—it only evaluates the rationality of a decision, and does not evaluate the process by which the decision is made.<sup>17</sup> It should be a relief that a lack of access to utilities leaves the core of decision theory untrammelled. But it is worth looking more closely even at the status of distinctively deliberative ambitions for decision theory.

On the epistemological interpretation, the inaccessibility of utilities concerning transformative experiences consists in agents' being uncertain about those utilities of theirs. Such inaccessibility is consistent with agents having substantial confidence—just not quite certainty—about their utilities. (In fact, under this interpretation it seems to make more sense to say that agents do have access to their utilities, but that such access is imperfect.) An agent with substantial confidence about her utilities can still take her utilities into account when deliberating and make reliable judgments about which of her available actions maximizes expected utility. Her calculations will not be perfectly reliable, but epistemically transformative experiences are hardly the only factor leading to the imperfection of ordinary agents. Given the limitations of ordinary agents, merely moderate reliability is plausibly the most we can hope for in just about any context.<sup>18</sup> Under this in-

---

<sup>16</sup>Pettigrew, (Forthcoming).

<sup>17</sup>This may well seem strange. One might think this vision of decision theory insufficiently "internalist". Now admittedly, an agent's credences and utilities plausibly supervene on that agent's intrinsic properties, so there's still a measure of internalism there. But more than that just isn't part of the framework of decision theory. One can, of course, stipulate more about an agent's relationship to his credences and utilities. And much fruitful theorizing has been done on the basis of such stipulations. (For example, Robert Aumann's (1976) celebrated agreement theorem makes this stipulation along with several others.) But such stipulations are not essential to decision theory.

<sup>18</sup>For more about the epistemic significance of the limitations of ordinary agents see Williamson (2000). For a generalization of Williamson's reasoning to probabilistic contexts see Williamson (2008).

terpretation, epistemically transformative experiences thus do not pose any special problems, even for distinctively deliberative ambitions for decision theory.

Moreover, Pettigrew's proposal to replace the uncertain utilities about transformative experiences with the expectations of those utilities would do little to make deliberation easier for ordinary agents. Given that Pettigrew's model is one in which agents can be uncertain of their own utilities, there seems to be no reason to assume that agents cannot be uncertain of their own credences. Thus agents are liable to have just as much difficulty determining what Pettigrew would have them do as determining what orthodox decision theory would have them do.

### 2.1.3 Formal Problem: Averaging Across Different Utility Functions

Pettigrew's proposal involves taking an average across different utility functions. There are different possible utilities an agent may assign to some experience and different possible utility functions doing the assigning. But, standardly, there can be no such thing as an average across different utility functions. An average across different utility functions is not well-defined mathematically, as utilities are defined on interval scales and the operation of averaging is not well-defined across interval scales.<sup>19</sup> There is thus reason to be worried about the coherence of a proposal that requires averaging across different utility functions.

The situation is not hopeless, however. If the averages were taken across the utility functions of different people, then the situation would be more dire. But since all relevant utility functions are held by a single agent, some special maneuvering might help.<sup>20</sup> To my mind, the most plausible move would be to identify particular utilities in different utility functions with one another. Then an average taken across different utility functions could be meaningful. Suppose we can stipulate that value X in one utility function is equivalent to value Y in some other utility function. I'll call such a pair of values an *anchor*. In order for comparisons between two utility functions to be meaningful, there must be two distinct anchors between them. With two distinct anchors, both the relative positions and the relative scales of two utility functions could be determined, thus capturing their full structure.<sup>21</sup> So

---

<sup>19</sup>The philosophical and mathematical issues concerning averaging across utility functions are already widely understood, so I don't wish to belabor them. But neither do I wish to presuppose familiarity with those issues. For an explanation of them, see the appendix. For even more on this issue, see Briggs (2015).

<sup>20</sup>Speaking for myself, I do not think that special maneuvering helps. I think that averages across different utility functions are insuperably ill-defined. But I appreciate that others may think that the structure of utilities is more flexible than I do. Thus I feel I should explore the uses that such flexibility could be put to.

<sup>21</sup>Given only one anchor it would be impossible to capture the relative scales of the two utility functions. And the scale is always the root of the problem; merely additive differences always wash out on their own.



long as an agent had enough utilities which were not problematized by epistemically transformative experiences, those unproblematic utilities could act as anchors.

## 2.2 Ontological Interpretation: Utility Gaps

If the problem posed by epistemically transformative experiences is that agents lack the relevant utilities and only gain them (if ever) after having having an epistemically transformative experience, then Pettigrew’s proposed solution faces two problems: the first problem concerns the irrelevance of the future, and the second concerns the difficulties inherent in averaging across different utility functions.

### 2.2.1 Philosophical Problem: The Irrelevant Future

Decision theory mandates that an agent act so as to maximize expected utility. That expectation is calculated with the agent’s current credences and the agent’s current utilities. The agent’s credences may well change in the future<sup>22</sup> and the agent’s utilities may well change in the future. But those future credences and utilities are of no direct relevance to an agent’s decision. In just the same way that an agent can have credences and utilities about what the weather will be like tomorrow, an agent can have credences and utilities about what his credences and utilities will be like tomorrow. There’s nothing special about those future credences and utilities—they’re decision-theoretically inert. An agent can prefer that his future preferences be satisfied; he should then act so as to satisfy them. An agent can prefer that his future preferences not be satisfied; he should then act so as to not satisfy them. When an agent has occurrent preferences the question of what preferences the agent should aim to satisfy is straightforward—the agent should aim to satisfy those occurrent preferences. And when an agent does not have occurrent preferences then it is totally unclear what the aim of that agent’s actions should be (or even if there’s a fact about how agent should act in the first place). It is unreasonable to cast around for something to constrain the actions of an agent without preferences, and there’s no reason for that agent’s expectations about his future utilities to ground such constraints.

Pettigrew writes,

[P]resupposition: In the description of this procedure, we have not specified which of her utilities the agent should weight by her probabilities in order to obtain her expected utility for a given action—is it her utilities at the time she makes the decision, or at some other time, perhaps when the consequences of her decision will be making themselves felt? . . . Since this is not specified in standard expected utility theory, if that account

---

<sup>22</sup>And likely will as the agent gets more evidence.

is to be well-defined, there must be a presupposition that my utilities remain unchanged over time.<sup>23</sup>

But this is not so. Decision theory does not presuppose that an agent's utilities remain unchanged over time, nor does it need to. Decision theory no more makes the assumption that agents' utilities will remain fixed throughout the course of their lives than it makes the presupposition that agents will live forever. Agents do not even have to be *alive* when the effects of their decisions are felt; they certainly do not have to be *bouletically unchanged* when the effects of their decisions are felt. Decision theory does not depend on future utilities in any regard.

An agent's utilities express their preferences about states of affairs in which they may or may not exist, and in which (supposing that they exist) they may or may not maintain their preferences. An agent's utilities specify what an agent wants when it comes to these myriad possibilities. And without utilities, there is no decision-theoretic specification of what an agent wants.

A utility gap is a far-reaching thing. The only way for an agent not to have a utility for any experience he cannot fathom is for the agent to have no utility for any state of affairs in which he has an experience he cannot fathom. It's not as though the agent can have utilities for the things with which he's acquainted and leave out the utilities for the things with which he isn't acquainted. Utilities don't apply to things, but to states of affairs containing those things. An agent's future utilities after having had an epistemically transformative experience are then necessarily something about which the agent has no preferences. To say that the agent prefers that his future preferences be satisfied even though he has no current preferences is contradictory—asserting that the agent both does and does not have current preferences. The only way to make the case of a utility gap consistent is for the agent to have no utilities regarding his future utilities, to not care about those them in any way at all. And in that case the future utilities are devoid of immanent relevance.

### **2.2.2 Formal Problem: Averaging Across Different Utility Functions (Reprise)**

Under this interpretation, averaging across different utility functions poses the same problems as before. In this regard there is no difference between an agent averaging different utilities he thinks he may have and an agent averaging different utilities he thinks he may come to have. Just as before, the best hope for the meaningfulness of such averages is a plurality of anchor utilities.

---

<sup>23</sup>Pettigrew (2016).

## 2.3 Epistemically Transformative Experience and Decision Theory

The inaccessibility of utilities concerning epistemically transformative experiences can be interpreted either epistemologically or ontologically. Pettigrew's proposed alterations to decision theory are unwarranted on either the epistemological interpretation or the ontological interpretation. Thus any decision-theoretic problems posed by epistemically transformative experiences remain unmitigated.

But this analysis suggests a deeper question: Do epistemically transformative experiences actually pose any problems for decision theory? The answer depends a bit on whether we take the epistemological interpretation or the ontological interpretation, whether agents are uncertain of what their utilities regarding epistemically transformative experiences are or whether they instead lack utilities regarding epistemically transformative experiences altogether.

If agents are uncertain of what their utilities regarding epistemically transformative experiences are, then it seems that epistemically transformative experiences would pose no problem for decision theory. Decision theory can function as normal given agents who are uncertain about their utilities. Moreover, ordinary agents are liable to be at least somewhat uncertain of at least some of their utilities quite apart from issues involving epistemically transformative experiences. It thus seems as though epistemically transformative experiences have no distinctive status in decision theory.

If agents lack utilities regarding epistemically transformative experiences, then it seems that epistemically transformative experiences would pose a problem for decision theory. The problem that they would pose, however, seems implausibly large. A utility gap regarding epistemically transformative experiences is a utility gap regarding any option in which an agent has an epistemically transformative experience. If options are individuated finely, then these utility gaps will be hugely problematic. Nearly every possible world we consider involves us undergoing an epistemically transformative experience at some point.<sup>24</sup> Epistemically transformative experiences would not only mean that there's no rational way to choose whether or not to adopt a child; they would equally mean that there's no rational way to choose whether to eat one's favorite lunch or instead eat one's least favorite lunch. One might hope to avoid this problem by individuating options coarsely, so that the options an agent faces in mundane situations have nothing to do with epistemically transformative experiences. But such coarse-graining is only legitimate when it doesn't make a decision-theoretic difference. An agent with utility gaps really does not have all the preferences that an agent without utility gaps has, and it would be inappropriate for decision theory to treat an agent with utility gaps

---

<sup>24</sup>Moreover, if one cares at all about the experiences of others, then the unfathomability of someone else's experiences would pose an identical problem.

as though she didn't have utility gaps. Decision theory must fall silent regarding any decision involving any possibility with a utility gap. That would be a major problem. If epistemically transformative experiences require utility gaps, then it is implausible that there are epistemically transformative experiences.

The relationship between epistemically transformative experiences and decision theory depends on how epistemically transformative experiences are understood. But however they are understood, it seems doubtful that epistemically transformative experiences pose a problem for decision theory.

### 3 Personally Transformative Experience

A personally transformative experience is one which alters an agent's values—which things are valued and the extent to which they are valued. In a decision-theoretic framework, transformative experiences are ones which alter an agent's utilities. Paul takes it that such changes pose a problem for decision theory. “[D]ue to the incommensurability of the preference change, [the agent] cannot adopt a principled decision rule to prefer one set of preferences over the other. Because she cannot step into a neutral first-personal perspective in order to evaluate and compare each possible successor first-personal perspective, she cannot formulate a higher-order rule that will adjudicate the decision for her.” Pettigrew agrees that decision theory faces a problem in any situation in which an agent's utilities are not guaranteed to remain fixed. He writes,

Suppose I face a choice between a range of acts; and suppose that one of those acts has a possible outcome that involves a [personally transformative experience]. 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 us say, very crudely and ignoring the fact that becoming a parent may be an [epistemically transformative experience] as well as a [personally transformative experience], 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.<sup>25</sup>

Pettigrew's proposed solution is to require that agents have a new sort of utility function. A traditional utility function represents how an agent value complete scenarios (in the limit, how the agent values entire worlds). This new sort of utility function, which Pettigrew dubs a local utility function, instead represents how an agent values possible present time slices. Pettigrew's proposal is to aggregate local utility functions together to make what he calls a global utility function, and then to use that global utility function as traditional decision theory uses a traditional utility function.

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,  $t_1, \dots, t_n$ . And let  $lu_{t_1}, \dots, lu_{t_n}$  be my local utility functions at those moments. Thus, the outcome in question is:

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

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 utility functions  $lu_{t_1}, \dots, lu_{t_n}$ . 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

---

<sup>25</sup>Pettigrew (2015).

$\beta_{t_1}, \dots, \beta_{t_n} \geq 0$ . 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  $j = 1, \dots, n$ , the weight  $\beta_{t_j}$  specifies the extent to which I take into account the values I had or have or will have at time  $t_j$ . Thus:

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

This solves the formal problem.<sup>26</sup>

Unfortunately, Pettigrew’s proposal faces three substantial problems: the first concerns the irrelevance of the future, the second concerns local utilities themselves, and the third concerns the difficulties inherent in averaging across different utility functions.

### 3.0.1 Philosophical Problem: The Irrelevant Future (Reprise)

It is simply false that there is any presupposition in decision theory that an agent’s utilities will not change over time. It is easy to say what an agent should do in the face of potentially changing utilities—maximize expected utility relative to his current credences and current utilities.

The decision-theoretic primacy of an agent’s current credences and current utilities is not some arcane technicality of the formalism—it makes perfect sense. Suppose that you currently value the happiness and wellbeing of your family. Nonetheless, you know that tomorrow you will undergo a horrifying (albeit temporary) change, and will come to instead value the misery and wretchedness of all those you currently hold dear. Note that the only thing that will change is what you value; your psychological reactions will remain the same. Even tomorrow you wouldn’t take an iota of pleasure from the suffering of your loved ones. You’d just choose that they suffer, and be utterly horrified that you had so chosen.

What should you do? It would be wrong to suggest any sort of compromise between your current utilities and your future utilities. You needn’t feel compelled to help your future self get what your future self will want. In fact, since what your future self will want is so diametrically opposed to what you want, you should feel compelled to *prevent* your future self from getting what your future self will want. You should take steps to make sure that you won’t be in a position to do any harm when tomorrow takes its strange toll on your utility function. You might consider

---

<sup>26</sup>Pettigrew (2015).

locking yourself in a cell. If you can get yourself into a medically induced coma, so much the better.<sup>27</sup>

It might seem strange that decision theory attaches no direct significance to one's future desires. One might think that adequately anticipating one's future desires is a hallmark of rational action. One might worry that traditional decision theory forces an agent into an unnatural disregard for her future self. I believe, however, that traditional decision theory can validate our natural judgments about what sorts of action are rational. But the relevant cases must be framed carefully, or else there is liable to be confusion. Consider the following case—

**Future Plans:**

Charlotte is wondering whether remain at home and watch television or to attend a university and study politics and literature. As things stand, Charlotte likes watching television and doesn't care about politics and literature. But Charlotte knows that if she attends a university she will come to care about studying politics and literature much more than she cares about watching television.

What should Charlotte do—should she watch television or study politics and literature? As things stand, I don't think this question can be answered yet. The case is underdescribed, and can be precisified in two different ways: Either (1) going to university would bring about a change in Charlotte's tastes, or (2) going to university would bring about a change in Charlotte's utilities. Let's look at these precisifications more closely.

(1): For convenience and clarity, let's assume that Charlotte is a hedonist. Charlotte's preferences between options correspond to her quantity of pleasure in those options. As things currently stand, Charlotte gets more pleasure out of watching television than out of studying politics and literature. But Charlotte knows that if she goes to university her tastes will change, and she will get much more pleasure out of studying politics and literature than she gets out of watching television. In this case, Charlotte assigns higher expected utility to attending university, as she knows she will get more pleasure attending university than she would staying at home. According to decision theory the rational thing for Charlotte to do is to attend university, and that verdict is eminently sensible.<sup>28</sup>

---

<sup>27</sup>Note that my point is not that Pettigrew's modified decision theory cannot deliver the natural verdict about the case—the weighting function he employs can discount sufficiently abhorrent utilities. My point is instead that Pettigrew's modifications are not necessary to deliver that verdict, that standard decision theory comes to it as well. There is no need to explicitly discount abhorrent future-utilities; current utilities can fully express the extent to which an agent cares or does not care about abhorrent future-utilities. For more on this sort of decision problem, see Elster (1979).

<sup>28</sup>At bottom, this case isn't any more fraught than that of working up an appetite so that one will enjoy dinner more.

(2): Again, let's assume that Charlotte is a hedonist. As things currently stand, Charlotte's preferences between options correspond to her quantity of pleasure in those options. But Charlotte knows that if she attends university and studies politics and literature she will cease being a hedonist, and will come to value other things than pleasure. As a hedonist, Charlotte only values her hedonism instrumentally—she'd happily become a stoic if by doing so she could have more pleasure, and not otherwise. Let us assume that the change in her utilities would not bring about a change in her tastes (or any other relevant change) that would make her life of study more pleasurable than her life of television-watching. In this case, Charlotte assigns higher expected utility to staying at home, as she knows she will get more pleasure staying at home than she would attending university. According to decision theory the rational thing for Charlotte to do is to stay home, and that verdict is eminently sensible.

An agent's current utilities are perfectly capable of valuing her future utilities in any way she likes. There is therefore no call to give a special status to the utilities that an agent may come to have in the future. Concomitantly, there is no problem posed by utilities changing over time.

### 3.0.2 Philosophical Problem: Trouble with Local Utilities

Pettigrew's local utility functions do great violence to the notion of the utility function. They are not a little different than regular utility functions. They are drastically different.

Given Pettigrew's proposal, there is literally no way to care about anything that happens after you die. Agents have no local utility functions covering times after their deaths, so those times are irrelevant to the global utility function which is stitched out of local utility functions. It is wrong to have a decision theory in which it is not so much as *possible* to work for the wellbeing of the family one will leave behind, for the good of future generations, or for the grim purpose of posthumous vengeance against one's enemies. A decision theory in which the future cannot matter is a decision theory that isn't worth having.

Pettigrew's local utility functions only value particular time slices. There is therefore no way for local utility functions to value or disvalue the various ways that time slices fit together in a world. Thus Pettigrew would force agents to evaluate each moment in isolation. There would be no way, for example, for one to care about the persistence of one's spouse. Being a newlywed is pretty spectacular; it would therefore be hard to avoid valuing most highly a weird, patchwork world in which one gets married (apparently for the first time) over and over again. The weird, patchwork world doesn't make sense, but there is no way to disvalue that lack of sense.

Local utility functions are also not particularly well-defined. Since local utility



functions are operant only over possible present moments, it's tremendously hard for them to make the sorts of comparisons they need to. Suppose you're deciding whether or not to have a child. At any time in which you know you have a child, your local utility for not having a child will be undefined. At any time in which you know you don't have a child, your local utility for having a child will be undefined. Agents can only have local utilities for epistemic possibilities, and that's a big problem. Even if you don't know whether or not you will have a child in the future, your normal-looking futures are ones in which you know whether or not you wound up having a child.

Local utility functions are deeply problematic.<sup>29</sup> Decision theory should not incorporate them.

### 3.0.3 Formal Problem: Averaging Across Different Utility Functions (Finale)

The problems of averaging across different utility functions present themselves again, but more starkly. Personally transformative experiences straightforwardly change an agent's utility function, making comparisons across utility functions more difficult. As before, a plurality of anchor utilities is the most plausible way to make the comparisons mathematically and philosophically meaningful. The problem is that the existence of such anchor utilities is more doubtful.

The only hope for getting anchor utilities is a specification of which utilities were changed by the personally transformative experience. Unchanged utilities might then act as anchors, as before. If some personally transformative experience leaves no utility unchanged, then there is simply no hope for a comparison. Note that the specification of which utilities changed and which remained the same would have a practical effect. Changing these utilities and leaving those the same would be very different from changing those utilities and leaving these the same. Suppose an agent has utilities defined over four options: A, B, C, and D. Suppose that the agent's utilities for these options can be given the following numerical representation: 2, 4, 6, and 8. (Remember, the agent's utilities are unique only up to positive affine transformation; other numerical representations for the agent's utilities are also available.) Consider a change of the agent's utilities such that A and B become worse and C and D remain the same. The agent's utilities are now 0, 2, 6, and 8. Now consider a change of the agent's utilities such that C and D become better and A and B remain the same. The agent's utilities are now 2, 4, 8, 10. These two utility functions, 0, 2, 6, 8 and 2, 4, 8, 10, would have different effects on an average. But note that, standardly, 0, 2, 6, 8 and 2, 4, 8, 10 are two numerical representations of *the same utility function*. Utilities are unique only up

---

<sup>29</sup>The weighting function over the local utilities is also something of a blank check, and it is unclear how deal with different possible life-spans.

to positive affine transformation, and the latter utilities are each just 2 larger than the former utilities. This means that Pettigrew needs to invoke more finely-grained utility functions than are employed in traditional decision theory. Averaging across different utility functions always requires some machinery not contained in the standard framework of decision theory. This sort of averaging across different utility functions requires more additional machinery than usual.

### **3.1 Personally Transformative Experience and Decision Theory**

The possibility of personally transformative experiences led Pettigrew to propose alterations to decision theory. Those proposed alterations are untenable. Thus any decision-theoretic problems posed by personally transformative experiences remain unmitigated.

But this analysis suggests a deeper question: Do personally transformative experiences actually pose any problems for decision theory? It seems that they do not. Decision theory is well-equipped to deal with agents whose utilities change over time. A personally transformative experience could well also involve an epistemically transformative experience. But then since it seems doubtful that epistemically transformative experiences pose a problem for decision theory, it seems doubtful that personally transformative experiences pose a problem for decision theory.

## **4 Conclusion: Prospects for Transformative Experience**

Transformative experiences do not seem to pose a problem for decision theory. Decision theory is not problematized by uncertainty about utilities nor by changes in utilities. Decision theory would be problematized by gaps in utilities, but such gaps would be so broadly ruinous that it is implausible that transformative experiences produce them.

It should not be surprising that transformative experiences do not pose a problem for decision theory. The framework of decision theory is merely formal, and thus very flexible. Thus transformative experiences can still have significant relevance to practical reasoning without impugning the viability of standard decision theory. And there are substantive issues posed by transformative experiences. What sorts of expectations regarding transformative experiences are reasonable, and what expectations regarding transformative experiences are unreasonable? To what extent ought agents to care about future changes in their preferences? Decision theory can operate so long as numbers for credences and utilities get plugged in, but what

numbers should get plugged in? These questions have been asked before.<sup>30</sup> But their answers have not yet been settled, and it is plausible that their answers cannot be settled with adequate consideration of transformative experience. I therefore suggest that transformative experiences should not be taken to pose a problem for decision theory, but should instead be taken to pose a topic for ethics.

## Appendix

It might seem easy to average the utilities of two different people. Utilities are numbers, and averaging numbers is easy. But averaging utilities is not so easy a thing. It's well-known among economists<sup>31</sup> that averaging the utilities of two different people—indeed, making any interpersonal comparisons of utility whatsoever—is not meaningful given the standard structure of decision theory. It's also well-known that with some additional technical apparatus more can be meaningfully done. Such issues are not as well-known among philosophers, however. I therefore include this explanation of the philosophical issues at stake in the mathematical structure of utilities.

There are limits to what sorts of mathematical operations can be meaningfully performed. Suppose someone wondered, “What’s the height of a tree that’s as tall as the temperature of a hot summer day?” One could reply “A hot summer day is around 95 degrees Fahrenheit, so the tree would be around 95 inches tall. It wouldn’t be a particularly tall tree, then, but more likely a sapling.” But this reply is obvious nonsense. The whole idea of directly equating some height with some temperature is absurd. Note that the numbers don’t help even though the same numbers appear when measuring heights and temperatures. There’s no equivalence to be had between 95 degrees Fahrenheit and 95 inches. Degrees Fahrenheit and inches are just very different things; the common ‘95’ doesn’t help. Of course, the selection of those units was arbitrary. One could have equally well taken degrees Celsius and centimeters and gotten a different bogus answer. But the basic problem is not the plurality of units. Then the problem would be that there were too many viable answers to the question. There are, however, no viable answers to the question. The question itself is deeply misguided.

Numbers (specifically, the natural numbers) can be put to a great many uses. Different uses capitalize on different aspects of the numbers’ numerical structure. We can use the numbers 1, 2, and 3 to describe how many coins Alice, Bob, and Carol have in their pockets. In this case, it makes sense to use the additive structure of those numbers. Just as  $1 + 2 = 3$ , it makes perfect sense to say that Alice and Bob have as many coins in their pockets as Carol does in her pockets. We can also

---

<sup>30</sup>See Bricker (1980), Parfit (1984), Gibbard (1992), and Bykvist (2010) for more.

<sup>31</sup>See Harsanyi (1955) and Arrow (1963) for treatments of this issue by two eminent economists.

use the numbers 1, 2, and 3 to describe the order in which Alice, Bob, and Carol completed a race. But in this case it makes no sense to use the additive structure of those numbers. There's no sense in which Alice's race result and Bob's race result, taken together, are equivalent to Carol's race result. But the ordinal structure of the numbers still applies. Just as  $1 < 2 < 3$ , it makes perfect sense to say that Alice finished before Bob, who in turn finished before Carol.

The mathematical structure of utility functions is quite modest, too modest to allow for interpersonal comparison. Utilities have the structure only of an interval scale. Utilities have an order, and comparisons of differences between utilities can be made, but that's it. Utilities can express that an agent prefers A to B, and utilities can express that an agent prefers A to B by twice as much as he prefers C to D. But utilities standardly express nothing more. It is, for example, not meaningful to say that A has twice as much utility as B. The reason is that utilities are standardly meant to capture preferences and nothing more. The preferences defined over simple outcomes are purely ordinal. Preferences over mixtures<sup>32</sup> of simple outcomes get only a little more structure. An agent prefers A to B by as much as he prefers B to C just in case the agent is indifferent between the certainty of B and a 50 / 50 mixture of A and C. And so on. The numerical representation of an interval scale is—as Pettigrew correctly notes—unique only up to positive affine transformation (that is, adding or subtracting any number to all utilities and multiplying all utilities by any positive number leaves those utilities unchanged). All the properties that matter in a utility function will be preserved by any positive affine transformation. Suppose an agent has three options, A, B, and C. Giving them utilities of 1, 2, 3 is the same as giving them utilities of 5, 6, 7 (just add 5), is the same as giving them utilities of 10, 20, 30 (multiply by 10), is the same as giving them utilities of 15, 25, 35 (multiply by 10 and then add 5), is the same as giving them utilities of 50, 60, 70 (add 5 and then multiply by 10). All the properties a utility function has are preserved in those assignments. Those utility functions all just represent preferring C to B to A, and preferring C to B and B to A by the same amount. Those utility functions are the same utility function, just as  $(2 + 2)$  and  $(1 + 3)$  are the same number. The representations are different, but it's the same thing being represented.

On a mathematical level it makes sense that you can't average across different utility functions. Suppose you're trying to take a straight average of two utility functions, giving them each equal weight. What numerical representations of the two functions should you take? If you average 1, 2, 3 and 3, 2, 1, you get 2, 2, 2. But if you average 100, 200, 300 and 3, 2, 1, you get 51.5, 101, 150.5. The latter representation of the first utility function gives it massively more of an effect than the former representation does. But any representation-dependent operation

---

<sup>32</sup>Such as lotteries.

is nonsense.

On a philosophical level it makes sense that you can't average across different utility functions. Utilities don't measure absolute amounts of desirability. It's not as though you can average liking X a lot with liking X a little and thereby get liking X moderately. Two agents, one of whom likes every possibility and the other of whom dislikes every possibility, can easily have the same utility function for those possibilities. Utilities show only relative desirabilities, and standings in two different hierarchies of relative desirability can't be averaged together.

Note that—even in the worst-case scenario—formal problems with averaging across different utility functions do not mean that information about other people's utilities has no bearing on what you should think about your own utilities. Suppose you're uncertain about your own preferences, but have reason to believe that your preferences are similar to a friend's preferences. Then learning that your friend prefers A to B can easily give you evidence that you prefer A to B. There are plenty of ways that the utilities of others can bear what one should think about one's own utilities. The formal problems have a limited scope. The worry isn't that information about other people's utilities will be categorically unhelpful; the worry is that certain sorts of comparisons—including the sorts of comparisons that Pettigrew's proposal requires—will not be possible.

One can, however, define non-standard utilities which express more than relative desirability according to some agent. The structural poverty of an interval scale can be enriched. And if one is to alter decision theory in the ways Pettigrew prescribes, then such enrichment is necessary.

## References

- Arrow, Kenneth. *Social Choice and Individual Values*. New York, NY: John Wiley & Sons, Inc, 1963.
- Aumann, Robert, 1976. 'Agreeing to Disagree' *The Annals of Statistics*, 4 (6), pp. 1236-1239.
- Bricker, Phillip, 1980. 'Prudence'. *Journal of Philosophy* 77 (7):381-401.
- Briggs, Rachael, 2015. 'Transformative Experience and Interpersonal Utility Comparisons'. *Res Philosophica* 92 (2):189-216.
- Bykvist, Krister, 2010. 'Can unstable preferences provide a stable standard of well-being?'. *Economics and Philosophy* 26 (1):1-26.
- Davidson, Donald, 1987. 'Knowing one's own mind'. *Proceedings and Addresses of the American Philosophical Association* 60 (3):441-458.
- Elster, Jon. *Ulysses and the Sirens*. Cambridge: Cambridge University Press, 1979.
- Gibbard, Allan. 'Interpersonal Comparisons: Preference, Good, and the Intrinsic Reward of a Life'. In Jon Elster and Anand Hylland (eds) *Foundations of Social Choice Theory*, Cambridge University Press, 1992.
- Harsanyi, John C, 1955. 'Cardinal Welfare, Individualistic Ethics, and Interpersonal Comparisons of Utility.' *Journal of Political Economy* 63 (4): 309-321.
- Paul, Laurie. *Transformative Experience*. Oxford: Oxford University Press, 2014.
- Paul, Laurie, 2015. 'Transformative Choice: Discussion and Replies'. *Res Philosophica* 92 (2):473-545.
- Paul, Laurie, 2015. 'What You Can't Expect When You're Expecting' *Res Philosophica*, 92 (2), pp. 1-23.
- Parfit, Derek. *Reasons and Persons*. Oxford University Press, 1984.
- Pettigrew, Richard, 2015. 'Transformative Experience and Decision Theory' *Philosophy and Phenomenal Research*, 91 (3), pp. 766-774.
- Pettigrew, Richard, 2016. Review of *Transformative Experience*, by L. A. Paul. *Mind* 125 (499):927-935.

Pettigrew, Richard. 'Transformative experience and the knowledge norms for action: Moss on Paul's challenge to decision theory'. In Lambert, E. and J. Schwenkler (eds.) *Becoming Someone New: Essays on Transformative Experience, Choice, and Change*, Oxford University Press, Forthcoming.

Williamson, Timothy. *Knowledge and its Limits*. Oxford University Press, 2000.

Williamson, Timothy, 2008. Why epistemology cannot be operationalized. In *Epistemology: New Essays*, ed. Quentin Smith, pp. 277-300.