



# Rational preference in transformative experiences

Saira Khan<sup>1</sup>

Received: 2 September 2020 / Accepted: 16 February 2021 / Published online: 9 March 2021  
© The Author(s) 2021

## Abstract

L. A. Paul’s *Transformative Experience* makes the claim that many important life decisions are epistemically and personally transformative in a way that does not allow us to assign subjective values to their outcomes. As a result, we cannot use normative decision theory to make such decisions rationally, or when we modify it to do so, decision theory leads us to choose in a way that is in tension with our authenticity. This paper examines Paul’s version of decision theory, and whether this version in fact admits of the challenge she wants to raise. I focus on her psychological realist view of utilities and beliefs and her notion of rational, authentic preferences as informed by imaginative acquaintance. I argue that Paul fails to engage critically with traditional accounts of decision theory and, on closer inspection, it is not clear that her version of decision theory entails a tension between rational and authentic choice. More importantly, I argue that if her contribution is instead to bring to light the importance of authenticity alongside rational decision-making, the definition she provides of authenticity in fact undermines her argument.

**Keywords** Decision · Transformative · Authentic · Utility · Belief · Preference

## 1 Introduction

Consider Sally, who has long believed that having a child would bring her a sense of fulfilment and happiness. From the perspective of normative decision theory, Sally’s preference to have a child would only be rational if it maximised expected utility. In *Transformative Experience*, L. A. Paul argues that in order to determine whether or not having a child would maximise utility, Sally would need to know “what it is like” to have a child (Paul, 2014a, pp. 19, 63). She would need to know the characteristics of her future child and the particular attachment relation she would develop with the child. She would also need to know how her preferences change after undergoing

---

✉ Saira Khan  
sairak1@uci.edu

<sup>1</sup> Department of Logic and Philosophy of Science, University of California Irvine, Irvine, CA, USA

the experience and the things she now comes to like or dislike. Since she does not have access to this information, Sally cannot make the decision in a fully informed, rational way.

Suppose she supplements this lack of knowledge with evidence from surveys about motherhood and the testimony of friends and family regarding their experiences. Perhaps now we might say she can make her decision rationally. Suppose further that this evidence suggests that most people experience lower subjective well-being after having a child. Paul thinks Sally cannot simply inform or adjust her subjective values based on such evidence because she does not know the relevant psychological similarities between herself and the standard respondents in such surveys (Paul, 2014a, pp. 89, 133). Even the potential scale or shape of the value space for having a child is something we cannot calibrate based on other experiences whose value we know. Therefore, Paul argues, normative decision theory would demand Sally “replace” her initial preference to have a child with that of not (Paul, 2014a, p. 89). For her, this is a major flaw in the theory, as rationality comes into conflict with what we would ordinarily consider Sally’s authentic preference. Here, authenticity is to be understood as self-governance with an emphasis on the expression of one’s core principles and commitments (Paul, 2015c, p. 493).<sup>1</sup>

Paul argues that such conflict arises in many important life decisions which have the character of “transformative experiences”. Transformative experiences are constituted by two features. First, they are such that the outcome of the decision is epistemically inaccessible to the agent prior to making the decision—where outcomes are to be understood in terms of “what it is like” to experience the particular outcome in question. For example, it is not possible to know what seeing the colour red is like before one sees it. This poses a problem for rational decision-making as Paul holds that, in order to assign a value to an experience, it is necessary to be able to imaginatively access what it is like (Paul, 2014a, p. 13). Second, the nature of the experience is such that one’s preferences change upon making the choice. For example, perhaps prior to having a child, Ann valued sleeping in and working late, but upon having a child, she now comes to value waking up early and spending time with her child much more. This poses a problem for rational decision-making as it is unclear whether to maximise the utility of one’s current or future self, particularly if the values of one’s future self are also epistemically inaccessible.<sup>2</sup>

Paul’s challenge is that the decision theorists’ ways of dealing with transformative experiences render authentic first-personal preferences secondary to third-personal evidence, and this is not how we would like to make important life decisions. I argue that this challenge is misplaced and decision theory does not demand that

---

<sup>1</sup> See also Paul’s Teaching Guide for Transformative Experience (Paul, Undated). Paul notes such replacement of one’s preference is not a problem in all cases: throwing oneself under a bus is unanimously considered a harmful thing to do, so third-personal evidence ought to justifiably inform our preferences about such an action. She concedes that if the evidence was clear-cut, Sally’s introspection should allow her to reflexively change her preferences in a way that preserves her authenticity, but this is not true of the transformative experiences in question (Paul 2015a, p. 811).

<sup>2</sup> If an agent does not have epistemic access to her future utilities, it is not possible to appeal to use of weighted or higher order utility functions to make such decisions (Jeffrey 1974; Pettigrew 2015; 2016b).

Sally give up her preference to have a child. In Sect. 2 I argue that Paul’s challenge in fact only makes sense under her own, unorthodox conception of decision theory, which relies upon psychologically real utilities and credences, places emphasis on the phenomenal value of experience, and considers imaginative forecasting a means to making informed choices. Since none of these features are prevalent in standard accounts of decision theory, the standard framework is unaffected by her challenge. One may respond that Paul is engaged in a different normative project than that of the standard theory, but if this is the case, she has failed to motivate her normative project before bringing to bear the tension she wishes to address within it. In Sect. 3 I argue that even considering the framework under which her challenge makes sense, her conclusion does not necessarily follow. She must do much more work to convince us that any version of decision theory that relies on psychological realism about utilities and credences entails that Sally change her preference.

Leaving aside these issues, a charitable reading of Paul may understand her as highlighting the importance of authenticity for transformative decision-making. However, given that she has neither provided a clear definition of authenticity in *Transformative Experience*, nor is the definition she later provides consistent with her account, I believe this more charitable reading is also problematic. Section 4 address this point. The definition of authenticity Paul provides requires that authentic, like rational, preferences are informed by way of imaginative acquaintance with the outcomes of one’s decision. I show that this requirement entails that Sally’s preference was never authentic to begin with, and in fact, could never be. This is something which has not been addressed in existing discussion of her work, yet it is imperative that we are clear about what authenticity amounts to, and that it is congruent with Paul’s argument, in order for us to be persuaded that there is a tension that must be addressed between the dictates of rationality and those of authentic choice. I conclude by considering whether her account could be made coherent using alternative notions of authenticity.<sup>3</sup>

## 2 Standard decision theory

Throughout the book, Paul presents her challenges as raised against “standard” decision theory (Paul, 2014a, pp. 21, 91).<sup>4</sup> In this section, I argue that Paul’s version of decision theory is rather non-standard, and it is only on her version that her criticism makes sense. First, on Paul’s account of decision theory, beliefs and utilities are treated as psychologically real entities that an agent can look to when making decisions—it is from these that one derives or discovers preferences. I refer to such a position as mentalist. More than this, though, Paul’s account holds that outcome descriptions are isolated from objective considerations, and are described uniquely

<sup>3</sup> I do not address Paul’s proposed solution to the problem of transformative experience since I am concerned with showing that the problem does not arise in the first instance. For criticism of her proposed solution, see Bykvist and Stefánsson (2017).

<sup>4</sup> See also Paul’s (2014b) interview with *Edge*.

in terms of what it is like to experience the state in question (Paul, 2014a, pp. 12, 26, 124). Additionally, she requires that value assignments must be rationally justified by way of evidence, and that evidence is constituted by being able to imaginatively access outcomes, which requires the relevant prior experience (Paul, 2014a, pp. 13, 23, 37). In what follows, I draw out the facets of this account that are non-standard and the centrality of these claims to her argument. First, I turn to her presentation of the standard theory.

Though Paul does not spell out what she refers to as standard decision theory in *Transformative Experience* itself, in other places, she presents the following formulation.<sup>5</sup> She holds that the standard theory tells us to list the possible acts we can perform; the outcomes of each act; the utilities assigned to each outcome; and the probability of the outcome obtaining conditional upon performing the act. She argues we should perform the action that maximises the probability-weighted sum of these utilities.<sup>6</sup> Is this the “standard” theory? Indeed it is “standard” in a very general sense, since it does not specify *what* the demand to perform that action amounts to. In fact, the presentation is consistent with a description of Leonard Savage’s (1954) framework, which is widely accepted, but the Savage framework is notably different from Paul’s account.<sup>7</sup> The key difference being that Savage’s theory is anti-realist, or what I will refer to as behaviourist, about beliefs and utilities. That is, beliefs and utilities are not psychologically real entities which one can refer to when making decisions. This means the theory cannot demand we directly attend to the utilities we assign and maximise their probability-weighted sum, rather rationality just is the performance of the act that maximises the probability-weighted sum of utilities.

In particular, on a behaviourist framework, justified beliefs and justified utilities do not combine to give us rational preferences, but rather, *from* rational preferences, beliefs and utilities can be described which mathematically capture an agent’s preference ordering. This is known as a representation theorem and will be discussed in detail in Sect. 3. What is important to note for our current discussion is that this version of decision theory does not tell us which preferences to have; it only requires that the set of preferences we have is *consistent* (discussed in Sect. 3). So if this is our standard account of decision theory, Sally need not give up her preference to have a child as long as her other preferences are consistent with this. Indeed, for the theory to demand Sally give up her preferences, the framework must be mentalist, since it is only on this understanding that utilities determine preferences, and it is precisely when Sally supplants her utilities with those informed by third-personal

<sup>5</sup> See Paul’s Teaching Guide for Transformative Experience (Paul, Undated).

<sup>6</sup> The presentation of standard decision theory in Paul’s (Undated) Teaching Guide may also be read as Jeffrey’s (1965) framework. However, since the example she uses here satisfies state-act independence, for simplicity, I restrict our discussion to the Savage framework. Note that the points I raise concerning the uniqueness of the additional requirements she imposes stand regardless of whether we consider the standard theory Jeffrey’s or Savage’s.

<sup>7</sup> See Buchak (2016) for an analogous presentation of Savage’s decision theory.

evidence that rationality supposedly demands she must overturn her preference to have a child.<sup>8</sup>

We must then ask, is the behaviourist account in fact standard? The origins of decision theory lie in Daniel Bernoulli's suggestion that there is a measure of subjective happiness that differs from the monetary value of a gamble.<sup>9</sup> This conception of subjective happiness is mentalist. However, whilst Bernoulli introduced the concept of expected utility maximisation, he did not offer a reason as to why agents ought to care solely about expectation of utility rather than other features of the gamble, such as its expected value, range, or variance. That is, he did not *justify* the principle of expected utility maximisation. It was the modern decision theory of the twentieth century developed by Ramsey (1926), von Neumann and Morgenstern (1953), and Savage (1954), which sought to deduce the principle of expected utility maximisation from a set of axioms, which would provide the normative basis for the principle lacking in Bernoulli's account, and this normative basis rested on behaviourist grounds (Okasha, 2016).

The usefulness of the behaviourist mathematical framework is widely acknowledged. However, to apply it one must determine how preferences are to be understood and there are differing views on the matter, particularly on the relationship between preference and choice behaviour.<sup>10</sup> There are also some prominent mentalist positions in the field, which adopt various positions on the relationship between mentalism and the representation theorem. For example, Nozick (1969) and Lewis (1981) present decision theory with reference only to utilities and beliefs, and do not mention their derivation from a representation theorem, whilst Joyce (1999, pp. 19–22) addresses the importance of rational representation on a mentalist picture.<sup>11</sup> As such, there may potentially be many frameworks considered standard in the literature. However, the important feature to note is that Paul's challenge that Sally needs to give up her preference only makes sense on a framework quite different to any of the existing positions. Paul's discussion is limited to what she refers to as "first-personal" choices. These are ones in which the subjective value of the experience is of utmost importance to the agent, and objective values of the outcomes may be ignored (Paul, 2014a, p. 124).

---

<sup>8</sup> That Paul must hold a mentalist interpretation of decision theory is also supported by Collins (2015) and Pettigrew (2019). She herself denies she is committed to non-constructive realism or what I will later refer to as strong mentalism (Paul 2015c), though I believe her claims are best understood on such a reading.

<sup>9</sup> For example, a \$5 gain with certainty is often preferred to a \$10 gain or loss on the flip of a coin, though both have the same expected value. Bernoulli's suggestion stemmed from the observation that people do not act to maximise expected monetary value in the St. Petersburg paradox.

<sup>10</sup> Some behaviourists (and many economists) understand preferences to be revealed by choice behaviour directly. Others believe that there is a larger metaphysical gap between the two—preferences being a state of mind or comparative attitude and choice behaviour being an action that results from the appropriate combination of such preferences with the external environment. Even among those behaviourists who consider there to be a strong relationship between the two concepts, there are yet multiple ways to construct preferences from observed choice behaviour. For more on the relationship between preference and choice, see Hansson and Grüne-Yanoff (2018) and Hausman (2012).

<sup>11</sup> See Okasha (2016) for discussion.

Yet it is precisely this focus on *experience* that entails Sally would be misguided in assigning justified values to the outcomes of her decision. In fact, Sally has access to non-experiential and objective information about the value of having a child, for example, how much it would cost. She may also have access to some experiential information of her own—she may have interacted with other children or child-minded her sister’s son for long periods of time. However, for Paul, what Sally is really missing is the *particular* experience of having her own child and standing in a parent–child attachment relation to that particular child. In Paul’s words, this particular experience is “an outcome whose value might swamp the other outcomes” (Paul, 2015b, p. 165).<sup>12</sup> It is only if we require this that Paul can maintain Sally’s preference to have a child is irrational. Yet few decision theorists save Lewis (1999) focus on the phenomenological value of experience. Indeed, Paul argues that this is something she brings to the theory which had previously been overlooked. That is, it is not a requirement of the standard account that one must assign values based on the experience of an outcome in order to rationally prefer it. If we do not require this, Sally need not change her preference based on the fact that she cannot assign a justified value to the experience of having her own child.

Posing a further problem is Paul’s requirement that value assignments are only rationally justified when informed by way of *imaginative acquaintance* with the outcomes of our decision. Why might we require such imaginative acquaintance? To illustrate a well-known example, if Mary grew up entirely in a black and white room, one might argue she cannot assign a value to what it is like to experience colour (Jackson, 1986). She can only assign a value to seeing the colour red when she can in fact imaginatively access the experience of seeing red, which she cannot do prior to the experience. Yet it is only if we require that Sally has *already had* the relevant prior experience that we run into a problem since Sally has not had the prior experience of having her own child. If we do not require this in order to imaginatively access outcomes, as above, Sally need not change her preference based on the fact that she cannot assign a justified value. If, for example, a more standard account required only that Sally take into account all available information in her value assignment, Sally’s initial preference is rational and need not be overturned.<sup>13</sup> This is not to say the requirement that value assignments be justified constitutes a flaw in Paul’s theory; indeed, many other decision theorists and philosophers would agree with such constraints. It is only to say it is not a requirement of more standard accounts that the assignment be justified by way of imaginative acquaintance, and it is precisely this unorthodox requirement which renders Sally’s preference irrational.

So whilst Paul argues that the standard decision theorist’s version of rationality has Sally replace her authentic preference with an inauthentic one, the form of decision theory that Paul targets is one that is focused on psychologically real, introspectable utilities and beliefs; that sees imaginative forecasting as the means

<sup>12</sup> See also Paul’s (2014b) interview with *Edge*. For criticism of this requirement, see Friedman (2015) and Chang (2015).

<sup>13</sup> For more on the value of imaginative forecasting in decision making, see Nanay (2016), and for criticism, see Van Leeuwen (2016).

by which our utilities and beliefs are rationally assigned; and places considerable weight on the phenomenal value of experience. Even if mentalism is prominent in decision theory, these further features are not widely accepted. Thus, more standard frameworks hold strong against her criticisms. If we describe the outcome in terms of its objective features and do not require that Sally knows what it is like to value it, then Sally may be rationally justified in her preference. In fact, if we take standard decision theory to be any behavioural variant, the theory in fact never tells Sally which preferences to have; it requires only that her set of preferences be consistent in order for her to be described as an expected utility maximiser.

Of course, it might be said Paul is simply engaged in a different normative project than that of traditional decision theorists. In particular, if her criticisms do not affect other decision theories that are more “standard” and are not focused on subjective phenomenal experience, she could argue that this is so much the worse for these theories, since they should be considering questions such as the metaphysics of selves and the value of phenomenal experience. A more charitable view of her aim, then, would be to bring to light the following: though it is possible to construct a decision theory on which Sally’s preference need not change, this version of decision theory is missing important aspects of real life decision-making. It is when we introduce these further aspects, which Paul thinks are essential to an account of rationality, that further complications arise, and it is these complications which she seeks to address in her work.

However, if Paul is not engaging with standard decision theory, she has failed to motivate her project since she has not provided us with reason to care about her normative standard. We have a notion of rationality given to us from standard decision theory. Paul bases her concept of rationality on the importance of phenomenal experience and imaginative acquaintance in order to draw out complications with respect to transformative decisions, yet we have received no justification of why this is an appropriate account of rationality. She only suggests this is the “ordinary” or “natural” or “culturally sanctioned” way we would like to make decisions (Paul, 2014a, pp. 20, 74, 83, 85). Why is this focus on experience and mental simulation considered ordinary? Indeed, Bykvist and Stefánsson (2017) note that transformative decision-making is only problematic for a particular type of person—the person who fetishizes the phenomenal experience of “what it is like” to have a particular experience over all other values. They go on to argue this is not the “normal approach” many people take to making life decisions, and that agents often focus on factors which are non-phenomenal.<sup>14</sup> So even if we grant that Paul is engaged in a different normative project than that of standard decision theorists, there may be no such agents for which her argument is valid. Without offering justification of her view of rationality as distinct from the standard theory, it seems *Transformative*

---

<sup>14</sup> Bykvist and Stefánsson’s (2017) point is congruous with, but not the same as, mine: that Paul’s criticism only makes sense on a particular, unorthodox view. However, whilst they focus on the peculiarity of the agent, I focus on the unorthodox nature of Paul’s conception of rationality, and consider other features in addition to the focus on phenomenal value that further articulate the differences between her view and the standard account.

*Experience* constructs a straw man version of decision theory, to then criticise and offer a solution.

### 3 Strong and weak mentalism

We have seen that Paul's challenge only has force on her own, non-standard conception of decision theory. In this section, I precisify her claims within the more general mentalist framework, and show that her conclusion still does not follow: neither the strong nor weak versions of mentalism considered below entail that Sally must change her preference from an authentic one to an inauthentic one in the absence of further assumptions. I begin by elucidating the role of a representation theorem in behaviourist and mentalist decision theories.

In decision theory, a representation theorem establishes a biconditional relation between an agent's preferences satisfying certain rationality axioms and her maximising expected utility.<sup>15</sup> In particular, von Neumann and Morgenstern prove that if an agent has a probability distribution over lotteries of outcomes over which we can describe a preference relation that respects completeness, transitivity, continuity and independence, then one lottery is preferred to another just in case it has greater expected utility.<sup>16</sup> So if an agent's preferences obey these rationality axioms, she maximises expected utility. Note that for the behaviourist, it simply would not make sense to conceive of the biconditional acting in the opposite direction—where the maximisation of expected utility entails the rationality of preferences.<sup>17</sup> Since utilities and credences are not psychologically real, there is nothing for an agent to attend to in order to maximise expected utility directly: the maximisation of expected utility is a mathematical representation. However, for the mentalist, the injunction to maximise expected utility *can* tell us something significant about what preferences to have. Exactly what it demands, and whether this entails that Sally replace her authentic preference, will be made clear by further dividing mentalism as follows.<sup>18</sup>

---

<sup>15</sup> In what follows, I consider the biconditional relation of von Neumann and Morgenstern's representation theorem, rather than the conditional relation of Savage's representation theorem. I believe it is this that meets the mentalist on her own grounds. A conditional relation is inconclusive on the matter of whether the maximisation of utility entails that an agent has a rational preference ordering. In Savage's theory, satisfaction of *some* but not all axioms is implied by the maximisation of expected utility, that is, those rationality axioms that are necessary for the existence of the representation (or equivalently, those axioms closed under submodels) but not those concerning the structural environment of the decision. See Suppes (1974).

<sup>16</sup> Completeness of preferences requires for any two lotteries, either one is preferred to the other, or the agent is indifferent between them. Transitivity requires that if one lottery is preferred to another and that is preferred to a third, the first is also preferred to the third. Continuity requires that there is some middling lottery where the agent is indifferent between this and a mix between a better and a worse lottery. Independence requires that the preference ordering is preserved when we introduce the possibility of another outcome.

<sup>17</sup> That is, it would not make sense in an ontological or pragmatic sense. Of course, the logical entailment runs in both directions.

<sup>18</sup> The terminology which follows is my own, but see Buchak (2016) and Zynda (2000) for analogous distinctions.



**Table 1** The intelligibility of entailment under different interpretations of decision theory

	Rationality of preferences to maximisation of expected utility	Maximisation of expected utility to rationality of preferences
Behaviourist	Yes	No
Weak mentalist	Yes	No
Strong mentalist	No	Yes

Strong mentalism takes utilities and beliefs to be primitive: our preferences are constructed from them. The injunction to maximise expected utility then says that the rational preference to have is the preference for the act that yields the outcome with the highest expected utility. But it would not make sense to say that the rationality of an agent's preferences entail that she is maximising expected utility, because it is the fact of her maximising expected utility given her psychologically real utility function that tells us that her preferences are rational. In contrast, weak mentalism takes beliefs and utilities to be psychologically real, but constructed from preferences which are the primitive entities. Given this primitiveness, the agent cannot *bypass* her preferences to directly attend to utilities and beliefs—we therefore cannot read the principle of maximising expected utility as having any bearing on the rationality of her preferences.<sup>19</sup> This is summarised in the table below (Table 1).

Paul is most plausibly read as a strong mentalist. Sally maximises the particular utility function that Paul claims is the correct one for her to have given her available evidence, and this determines the rationality of her preference. However, on such a reading, it appears that Paul loses the normative strength on which decision theory is traditionally grounded: the more primitive rationality constraints on preferences. The merit of the behaviourist approach is that the normative justification is much easier to derive; the principle of maximising expected utility is not a rationality demand on the utilities and beliefs that one attends to directly, but is rather a consequence of more primitive constraints on one's preferences.<sup>20</sup> The reason this matters is because the maximisation of expected utility is an automatic consequence of having well-ordered preferences; it requires no justification as a principle of rationality apart from this. In other words, if Paul is a strong mentalist, she must explain why the principle to maximise expected utility is justified.

Specifically, if one believes that the agent has psychologically real utility and credence functions and follows the normative injunction that she should maximise her utility with respect to these, her preference relation satisfies the rationality axioms according to the von Neumann and Morgenstern representation theorem.<sup>21</sup> But crucially, even if she does not maximise her utility function, her preferences *may still* satisfy the axioms (Okasha, 2016). This is because if her preferences accord with the

<sup>19</sup> This position is similar to behaviourism but posits that what results from our preference ordering is not simply a mathematical representation but a psychologically real entity.

<sup>20</sup> Of course, one may argue over the justifiability of individual axioms as intuitive constraints, but this does not undermine that this is where the normative justification of the principle stems from.

<sup>21</sup> Though the mentalist might not appeal to the representation theorem in explaining why the resulting preferences are rational.

axioms, the representation theorem guarantees there is *some* function she is maximising, whether or not it is her psychologically real expected utility function. Therefore, the strong mentalist cannot appeal to the axioms to normatively justify the principle of maximising expected utility, since one can obey the axioms both while being an expected utility maximiser and while not. If strong mentalism lacks normative backing from a representation theorem, what is it that tells us that the rational thing to do is the act that maximises expected utility? If Paul cannot substantiate this demand, decision theory has no claim on Sally's actions. If decision theory has no claim on Sally's actions, *contra* Paul, Sally need not give up her preference to have a child.

Another way one may seek to justify the principle of maximising expected utility without appeal to representation theorems is to consider long-run arguments. These claim, drawing on the strong law of large numbers, that in a long run of gambles, the average gain per trial approximates the gamble's expected value. However, such arguments face many common conceptual problems. Most notably, it is not clear that many ordinary decisions are such that we could repeat them over infinitely many trials or that the relevant factors would remain unchanged, i.e. that trials are probabilistically independent and the subjective values of each trial are independent. Indeed, winning on one gamble may plausibly affect our utility assessment of winning on another. In particular, these conditions definitely do not hold for the kinds of decisions Paul is concerned with, so this cannot be the justification of the normative injunction that she is seeking. Whilst some decision theories could in principle provide satisfactory reason for the strong mentalist to claim that Sally ought to change her preference—for example, the mentalist frameworks of Pettigrew (2016a) and Easwaran (2014)—they are not widely considered standard, and are thus not the frameworks against which Paul claims to be raising her challenge.

Perhaps Paul is making a weak mentalist claim—that to maximise expected utility simply means there is a utility function whose expectation one is maximising. If this is the normative injunction of decision theory, must Sally replace her preference? No.<sup>22</sup> It is entirely consistent to hold that, though Sally is not maximising the utility function she considers to be psychologically real, insofar as her preferences obey the rationality axioms, she is representable as an expected utility maximiser.<sup>23</sup> In order for Paul to uphold that Sally needs to change her preference, there would

---

<sup>22</sup> One might argue this follows directly from the fact that the conditional relation is unintelligible in this direction: decision theory would not demand that Sally change her preferences based on her beliefs and utilities as, for the weak mentalist, it is her preferences which determine her beliefs and utilities.

<sup>23</sup> In the case of Sally, the reason she does not maximise her psychological real utilities is because they are incomplete. However, there is a sense in which her preferences are complete since Sally either chooses to have a child or not (though one may in principle deny the tight connection between choice behaviour and preference). If her preferences are also consistent, then by the representation theorem, there is some other utility function according to which Sally is an expected utility maximiser. Note this will not be the same as her psychologically real utility function since it is complete. One might think there is something strange about the weak mentalist position. In particular, it appears strange that an agent could be rational whilst completely failing to attend to her psychologically real utilities. Indeed, this oddity in weak mentalism has also been pointed out by Buchak (2016), who similarly notes it is a flaw of weak mentalism that “maximizing the expectation of some utility function doesn't guarantee that you've maximized the expectation of your own utility function” (footnote 12).

need to be some other principle that necessitated a connection between the utility function that falls out of the rationality constraints on an agent's preferences, and the utility function to which the agent actively attends. For example, Zynda's (2000) *Reality Condition* or Christensen's (2001) *Representation Accuracy*.<sup>24</sup> That is, if an agent's preferences can be represented as resulting from a unique (up to positive linear transformation) utility function and probability function by expected utility maximisation, then these functions *in fact describe* the agent's degrees of belief and utilities. In the absence of such a connection, Sally is acting rationally in preferring to have a child as long as Sally is maximising some utility function, regardless of whether this is the one she treats as psychologically real. Note that this preference to have a child is not only rational but also authentic, since Paul has stipulated from the beginning that Sally's authentic desire is to have a child. Thus, the tension between rationality and authenticity does not arise on the weak mentalist reading.<sup>25</sup>

So neither the weak mentalist nor strong mentalist injunctions to maximise expected utility entail that Sally must replace her preference to have a child with that of not. In the case of weak mentalism, this is because if multiple utility and credence functions can be constructed from an agent's preference ordering, it is not clear that there is a connection between her maximising expected utility and her maximising her psychologically real expected utility. In the case of strong mentalism, the demand lacks normative justification in the absence of appeal to a representation theorem, and the versions of the theory which could provide such justification are not the versions Paul claims to be targeting, since they are not "standard" decision theory.

A caveat to the above: we have noted that part of Paul's project is to draw out the problems in standard accounts of decision theory. To then say, in response to her work, that the representation theorem is constitutive of rationality would be a position in retreat. The standard theory is in fact inadequate for Paul's project, which seeks to address decision theory for the imperfect agent in non-ideal circumstances. This is indeed not the version of decision theory that receives its force from the representation theorem. So though the normative justification for the behaviourist framework is easier to derive, for Paul, this rests on an erroneous conception of rational decision making—one that ignores subjective phenomenal value. As a result, we should not see this as a merit of the theory. In the previous section, I raised concerns that her particular conception of rationality is unmotivated. In what

<sup>24</sup> See Meacham and Weisberg (2011) for discussion.

<sup>25</sup> One may object: just what are these "real" credences and utilities, if they are to be different from the mathematical entities? To provide an account of the psychological underpinnings, if any, of utility is beyond the scope of this paper. However, it suffices to say that one may plausibly hold that there is a metaphysical difference between this entity—insofar as it can even be posited as something we might introspectively access—and the idea that we can be mathematically represented by a utility function, of which there may be many that appropriately represent us. One only needs such an intuition of this difference to see why the above necessitating connection is important. Indeed, Paul clearly possesses this intuition as she is a strong mentalist, yet she does not acknowledge the need for this connection.

follows, let us further meet Paul on her own grounds, granting her version of decision theory the normative force it requires in the absence of a representation theorem, and ask whether her argument is internally consistent.

#### 4 Rationality and authenticity

So far, we have seen that Paul's challenge is not in fact raised against standard decision theory, and that even within the mentalist framework in which her challenge is intelligible, her conclusion does not follow absent other assumptions. In this section, I grant for the sake of argument that strong mentalism possesses the normative force she requires: the injunction to maximise expected utility means Sally must change her preference. I show how, if this is the case, Paul's argument is self-defeating since her definition of rationality and her definition of authenticity rest on the same condition of imaginative acquaintance. Thus, if Sally's preference to have a child is not rational, neither is it authentic, and the tension between rationality and authenticity that Paul wishes to draw out does not arise in the first instance.

First, it is important to note that Paul believes values and credences must be assigned by way of evidence if they are to be rationally justified (Paul, 2014a, p. 23). Such evidence is constituted by imaginative acquaintance. That is, Paul believes that in order to make justified value and credence assignments, we must be able to mentally simulate the outcomes of our decisions and this mental simulation is only possible if we have undergone relevantly similar experiences: "we gain the ability to assess the subjective value of the experience by gaining the ability to grasp it using our first-personal, imaginative perspective... that is, the experience is needed to give us access to the subjective value" (Paul, 2014a, p. 13). For example, I may possess imaginative acquaintance with what it is like to eat a cookie, allowing me to rationally decide whether eating the next cookie would maximise my expected utility, but I do not have such acquaintance with raising a child.<sup>26</sup> This is what Paul refers to as the "normative gold standard" of rationality, or what I will refer to as Paul-rationality.

**Definition** Paul-rational preferences are those which follow from credences and subjective values informed by way of imaginative forecasting.

This is an unattainable standard in cases of transformative experience. Indeed, Paul requires exactly that information which is *by definition unavailable* to us prior to making the decision: the experience of what it is like to have made that decision. Paul recognises this and thus considers whether Sally would be rational to supplement her lack of information with third-personal evidence. The problem, she claims,

---

<sup>26</sup> One might argue that the next cookie could be transformative, and Paul concedes that there is not a clear line. She appeals to a notion of experiential natural kinds to justify the difference between experiencing the next cookie and the experience of having a child (see Paul's (2014b) interview with *Edge*).

is that this notion of rationality demands that Sally replace an authentic preference with an inauthentic one (founded upon third-personal evidence). However, the argument only follows by obscuring two different notions of rationality, one of which is actually coextensive in important ways with her notion of authenticity, and the other being the Paul-imperfect rationality of the “standard” decision theorist.<sup>27</sup>

To make this clear, let us return to Paul’s definition of authenticity mentioned earlier. Crucially for Paul, not only are authentic choices those which are true to our deepest passions, but authenticity also involves “self-governance informed by knowledge under subjective modes of presentation, that is knowledge via experiential or imaginative acquaintance with the objects of deliberation.”<sup>28</sup> This is something which is not made clear in her original work and it is this ambiguity concerning the definition of authenticity that first allows her argument to seem appealing. However, in subsequent writing, authenticity for Paul requires imaginative acquaintance and, as noted above, imaginative acquaintance requires mental simulation, which requires having undergone a prior experience of a relevantly similar sort.

**Definition** Paul-authentic preferences are those which are informed by way of imaginative forecasting.

So can Sally’s initial preference to have a child be authentic? No, since it is not formed by way of imaginative acquaintance with the outcome of having a child. Since the precise definition of authenticity is not offered in *Transformative Experience*, the reader is tempted to ascribe their own common-sense notion of authenticity to Sally’s initial preference, obscuring the fact that, for Paul, authenticity requires imaginative acquaintance. So Paul-rational and Paul-authentic preferences rest on the same condition, a condition which by definition cannot be met by someone in the context of a transformative decision where outcomes are epistemically and personally inaccessible to the agent.

Sally’s initial preference to have a child is presented as *authentic* but *irrational*—in the sense of Paul-rationality. However, it is in fact both Paul-inauthentic and Paul-irrational as it was never formed by way of imaginative acquaintance. Sally’s later preference to remain childless after acquiring evidence is presented as *inauthentic* but *rational*—however, here, we have slid into talking about a different kind of rationality: a weaker notion of rationality adopted by the “standard” decision theorist that takes into account only available information; we have abandoned our notion of the “normative gold standard” and conceded that third-personal evidence suffices. This obscures the fact that Sally’s later preference is again both Paul-inauthentic and Paul-irrational as it is not formed by imaginative acquaintance. There

<sup>27</sup> Authenticity may involve further features, for example, recognition of the fact that one must live with the effects of one’s choices (Paul 2014a, p. 120). Thus, in principle, the two may come apart. However, this does not undermine my argument that both authenticity and rationality have in common one necessary condition: imaginative acquaintance, and this is enough to show that Paul’s argument is self-defeating.

<sup>28</sup> See Paul’s Teaching Guide for *Transformative Experience* (Undated). See also Paul (2015c).

is no change. In fact, Paul is committed to saying that in cases of transformative choice, any preference is Paul-irrational and any preference is Paul-inauthentic. This means that her argument does not get off the ground, as the force of her challenge rests on the presupposition that Sally's preference has changed from an authentic one to an inauthentic one by the dictates of decision theory.

One may object that there are many other notions of authenticity present in extant literature, which we may draw upon, and with which Paul's argument would not be self-defeating. On a more permissive reading than the one I have offered here, authenticity might amount to the ability to figuratively understand or occupy the shoes of a possible future self as an extension of one's current self. If Sally is not able to grasp the nature of a particular future self (the one where she remains childless), she is alienated from that future self. Yet in many traditional accounts of authenticity—those of Kierkegaard (1849), Heidegger (1927) and Sartre (1946)—the self is defined through concrete life choices an agent makes, which constitute the agent's identity over time.<sup>29</sup> Authenticity is not derivable from a set of pre-given opinions or feelings, subject to introspection, as Paul requires it to be. Indeed, Sartre writes, “feeling is formed by the deeds that one does; therefore I cannot consult it as a guide to action. And that is to say that I can neither seek within myself for an authentic impulse to action, nor can I expect, from some ethic, formulae that will enable me to act” (Sartre, 1946, pp. 36–7). Authentic choice is not a matter of introspection, but a process of self-creation.

Frankfurt (1988a, b, c) takes a volitional view of authenticity whereby to be true to oneself means to be the person one wants to be: structuring one's identity through one's commitments. Persons decide which of their (conflicting) desires they want to be “effective” or action-guiding (Frankfurt, 1988a). Yet Frankfurt also has a decision-centered notion of determining which desires are truly internal to the agent—that is, what reveals which of the two conflicting desires are internal, or higher in the agent's ordering, is in fact what the agent decides to act on (Frankfurt, 1988b).<sup>30</sup> He writes “the decision determines what the person really wants by making the desire on which he decides fully his own. To this extent the person, in making a decision by which he identifies with a desire, *constitutes himself*.” (Frankfurt, 1988c, p. 170). One way we could still see this as consistent with Paul's work is by understanding the third-personal evidence which determines Sally's rational decision as in fact an externally imposed desire. Or, in Frankfurt's (1988b) terminology, Sally's desire to remain childless is external to her in the sense that it is not part of the same “ordering”—it is a desire which she seeks to *reject*, not something which she wants *less* than having a child. This is one possible way in which, as Paul would have it, Sally's initial desire to have a child is authentic but her later desire to remain childless is inauthentic.

However, the existentialist account of self-creation may better accord with our ordinary notion of authenticity in cases of transformative life decisions. It is reasonable to think an agent does not possess an authentic preference when faced with such

<sup>29</sup> See Varga and Guignon (2020) for discussion.

<sup>30</sup> Note that deciding is not the same as choosing but rather forms an intention to act (Frankfurt 1988b).

a major choice shrouded in epistemic inaccessibility. Perhaps prior to making the decision the agent is unaware of what action would best respect her deepest desires, but after having the experience, the person comes to acquire authentic preferences which align with this experience. Of course, this does not provide a satisfactory formulation for the agent deciding whether or not to undergo the experience. However, the important thing to note is that, if even on more standard existentialist accounts of authenticity, any introspected preference in cases of transformative choice is not constitutive of authenticity, then Paul's argument does not get off the ground. If Sally's preference is not authentic until a choice has been made, then there is again no tension between the dictates of rational choice and of authenticity. If Paul seeks to convince us of this tension, she must provide a defensible alternative definition of authenticity, and one which is not based on imaginative acquaintance, since this renders her argument self-defeating.

## 5 Conclusion

The challenge to decision theory that Paul raises is thus: in cases of transformative experience, where our utilities are inaccessible, the normative injunction of standard decision theory would have us replace our preference with the preference informed by the utilities of other agents or third-personal evidence. This results in an inauthentic choice. First, as I have argued in Sect. 2, Paul's target is not in fact standard decision theory, but her own specific conception of decision theory which strips outcomes of their objective considerations, focusing instead on the phenomenal value of experience, and requires that psychologically real and introspectable utilities and credences are rationally assigned only by way of imaginative acquaintance with outcomes. Decision theories which do not include these features are unaffected by her criticisms. One might argue that Paul sees herself as engaged in a different normative project than that of the traditional behaviourists, but if this is so, she has not motivated why we ought to care about this particular notion of rationality, instead relying on an intuition that this is how we would ordinarily like to make decisions.

In Sect. 3, I argued that on a strong mentalist reading of the principle of expected utility maximisation, Paul loses the normative force on which decision theory is grounded through representation theorems. We no longer have clear grounds for accepting the principle of maximising expected utility as the basis for rational choice and therefore for accepting that Sally ought to give up her authentic preference. On a weak mentalist reading, decision theory does not demand that Sally replace her preference to have a child with the inauthentic choice not to, as when she chooses to have a child, there is *some* function whose expectation she is maximising as long as her preference ordering is consistent. Therefore, in neither case are the normative demands of rationality in conflict with authentic choice. Paul may argue in response that, though the normative justification for the behaviourist framework is easier to derive, it rests on an impoverished view of rationality to the extent that it ignores subjective phenomenal value.

Granting this, in Sect. 4, I consider whether Paul's argument is internally consistent. If Paul's work does not bear critically upon standard decision theory and does

not offer a defence of her own notion of rationality, as argued in Sect. 2, we might take her project instead to be one of bringing to bear the importance of authenticity in decision-making. Yet in order for her to convince us that that this is something worth addressing, she would need to provide a clear and explicit account of authenticity, which she does not in *Transformative Experience*. I argued that when we unpack her notion of authenticity provided in later work, we find that Sally's initial preference to have a child is as inauthentic as her later preference to remain childless, as neither can have been reached via imaginative acquaintance. Thus, the tension which Paul wishes to draw out between the dictates of authenticity and the dictates of rationality in decision theory does not arise in the first instance. Perhaps her argument is tenable on a more permissive notion of authenticity, such as that of Frankfurt's, but much more work would need to be done to define and defend a view of authenticity that does not see an agent's identity as constituted by the agent's choices over time, as is the case in more traditional existentialist accounts.

Whilst I do not think that Paul's argument is sound, this is not to say transformative experiences do not raise questions for decision theory. Paul is right to emphasise the uncertainty and inaccessibility we face when deciding to undergo lived experiences that may fundamentally change us in ways unknown. Her discussion brings to bear the difficulty that is faced when we apply seemingly elegant and intuitive normative standards to muddled, real life decisions. It brings to bear that such decisions affect different interpretations of decision theory in different ways. It also brings to bear that standard decision theory pays little attention to issues of authenticity, which might be an avenue for future work in the field.

**Acknowledgements** My sincerest thanks to Simon Huttegger, Brian Skyrms and Cailin O'Connor for their comments on earlier drafts of this paper, as well as the support and wise words of Calum McNamara, Daniel Herrmann, Gerard Rothfus, Nathan Gabriel, Adam Birnbaum and Josiah Lopez-Wild during its subsequent development. I am indebted to Laurie Paul for discussing her work with me and improving my understanding of it.

**Authors' contributions** Not applicable.

**Funding** None.

**Availability of data and material** Not applicable.

**Code availability** Not applicable.

**Declarations**

**Conflict of interest** The authors declare that they have no conflict of interest.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission



directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Buchak, L. (2016). Decision theory. In A. Hájek & C. Hitchcock (Eds.), *The oxford handbook of probability and philosophy* (1st ed., pp. 789–811). Oxford: Oxford University Press.
- Bykvist, K., & Stefánsson, O. (2017). Epistemic transformation and rational choice. *Economics and Philosophy*, 33(1), 125–138. <https://doi.org/10.1017/s0266267116000274>
- Chang, R. (2015). Transformative choices. *Res Philosophica*, 92(2), 236–282. <https://doi.org/10.11612/resphil.2015.92.2.14>
- Christensen, D. (2001). Preference-based arguments for probabilism. *Philosophy of Science*, 68(3), 356–376. <https://doi.org/10.1086/392889>
- Collins, J. (2015). Neophobia. *Res Philosophica*, 92(2), 283–300.
- Easwaran, K. (2014). Decision theory without representation theorems. *Philosophers' Imprint*, 14(27), 2–30.
- Frankfurt, H. G. (1988a). Freedom of the will and the concept of a person. In H. G. Frankfurt (Ed.), *The importance of what we care about: Philosophical essays*. Cambridge University Press.
- Frankfurt, H. G. (1988b). Identification and externality. In H. G. Frankfurt (Ed.), *The importance of what we care about: Philosophical essays*. Cambridge University Press.
- Frankfurt, H. G. (1988c). Identification and wholeheartedness. In H. G. Frankfurt (Ed.), *The importance of what we care about: Philosophical essays*. Cambridge University Press.
- Friedman, J. (2015). Epistemically transformative experience. Presented at the Pacific APA in Vancouver, April 2015 as part of an Author Meets Critics session on L. A. Paul's book, *Transformative Experience*.
- Hansson, S. O., & Grüne-Yanoff, T. (Summer 2018). *Preferences*. The Stanford Encyclopedia of Philosophy. Retrieved February 13, 2021, from <https://plato.stanford.edu/archives/sum2018/entries/preferences/>.
- Hausman, D. M. (2012). *Preference, value, choice, and welfare*. Cambridge University Press.
- Heidegger, M. (1927/1962). *Being and Time* (J. Macquarrie & E. Robinson, Trans.). Harper & Row. Harper Perennial Modern Classics: Reprint edition (July 22, 2008).
- Jackson, F. (1986). What mary didn't know. *The Journal of Philosophy*, 83(5), 291–295. <https://doi.org/10.2307/2026143>
- Jeffrey, R. C. (1965/1983). *The logic of decision* (2nd ed.). University of Chicago Press.
- Jeffrey, R. C. (1974). Preference among preferences. *Journal of Philosophy*, 71(13), 377–391. <https://doi.org/10.2307/2025160>
- Joyce, J. (1999). *The foundations of causal decision theory*. Cambridge University Press.
- Kierkegaard, S. (1849/1980). *The sickness unto death: The christian psychological exposition for upbuilding and awakening*. Princeton University Press.
- Lewis, D. (1981). Causal decision theory. *Australasian Journal of Philosophy*, 59(1), 5–30. <https://doi.org/10.1080/00048408112340011>
- Lewis, D. (1999). What experience teaches. In D. Lewis (Ed.), *Papers in metaphysics and epistemology* (1st ed., pp. 262–290). Cambridge University Press.
- Meacham, C. J. G., & Weisberg, J. (2011). Representation theorems and the foundations of decision theory. *Australasian Journal of Philosophy*, 89(4), 641–663. <https://doi.org/10.1080/00048402.2010.510529>
- Nanay, B. (2016). The role of imagination in decision making. *Mind and Language*, 31(1), 127–143. <https://doi.org/10.1111/mila.12097>
- Nozick, R. (1969). Newcomb's problem and two principles of choice. In N. Rescher (Ed.), *Essays in Honor Carl G. Hempel* (1st ed., pp. 114–146). Reidel.
- Okasha, S. (2016). On the interpretation of decision theory. *Economics and Philosophy*, 32(3), 409–433. <https://doi.org/10.1017/S0266267115000346>
- Paul, L. A. (2014a). *Transformative experience*. Oxford University Press.
- Paul, L. A. (2014b). *L.A. Paul: "The transformative experience"*. Edge. Retrieved February 13, 2021 from [https://www.edge.org/conversation/la\\_paul-la-paul-the-transformative-experience](https://www.edge.org/conversation/la_paul-la-paul-the-transformative-experience).

- Paul, L. A. (2015a). Transformative experience: Replies to Pettigrew, Barnes and Campbell. *Philosophy and Phenomenological Research*, 91(3), 794–813. <https://doi.org/10.1111/phpr.12250>
- Paul, L. A. (2015b). What you can't expect when you're expecting. *Res Philosophica*, 92(2), 149–170. <https://doi.org/10.11612/resphil.2015.92.2.1>
- Paul, L. A. (2015). Transformative choice: Discussion and replies. *Res Philosophica*, 92(2), 473–545. <https://doi.org/10.11612/resphil.2015.92.2.15>
- Paul, L. A. (Undated). *Teaching Guide for Transformative Experience*. L. A. Paul personal website. Retrieved February 13, 2021, from <https://lapaul.org/papers/teaching-guide-for-transformative-experience.pdf>.
- Pettigrew, R. (2015). Transformative experience and decision theory. *Philosophy and Phenomenological Research*, 91(3), 766–774. <https://doi.org/10.1111/phpr.12240>
- Pettigrew, R. (2016). Risk, rationality and expected utility theory. *Canadian Journal of Philosophy*, 45(5–6), 798–826. <https://doi.org/10.1080/00455091.2015.1119610>
- Pettigrew, R. (2016). Transformative experience, by L. A. Paul. *Mind*, 125(499), 927–935. <https://doi.org/10.1093/mind/fzw014>
- Pettigrew, R. (2019). *Choosing for changing selves*. Oxford University Press.
- Ramsey, F. R. (1926). Truth and probability. In R. B. Braithwaite (Ed.), *The foundations of mathematics and other logical essays*. Kegan, Paul, Trench Trubner & Co.
- Sartre, J. P. (1946/1948). *Existentialism is a humanism* (P. Mairet, Trans.). Methuen.
- Savage, L. J. (1954). *The foundations of statistics*. John Wiley.
- Suppes, P. (1974). The measurement of belief. *Journal of the Royal Statistical Society. Series B (Methodological)*, 36(2), 160–191.
- Van Leeuwen, N. (2016, April 25). *Partially Doxastic Imaginings*. Symposium on Bence Nanay, “The Role of Imagination in Decision-Making”. Retrieved February 13, 2021, from <https://philosophyofbrains.com/2016/04/25/symposium-on-bence-nanay-the-role-of-imagination-in-decision-making.aspx>.
- Varga, S., & Guignon, C. (Spring 2020). *Authenticity*. The Stanford Encyclopedia of Philosophy. Retrieved February 13, 2021, from <https://plato.stanford.edu/archives/spr2020/entries/authenticity/>.
- Von Neumann, J., & Morgenstern, O. (1953). *Theory of games and economic behavior*. Princeton University Press.
- Zynda, L. (2000). Representation theorems and realism about degrees of belief. *Philosophy of Science*, 67(1), 45–69. <https://doi.org/10.1086/392761>

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.