

# Transformative Experiences, Rational Decisions and Shark Attacks

Marc-Kevin Daoust, Harvard University (mk.daoust@live.ca)

Forthcoming in *Inquiry* (DOI: [10.1080/0020174X.2021.1953578](https://doi.org/10.1080/0020174X.2021.1953578)). Please cite the published version.

**Abstract:** How can we make rational decisions that involve transformative experiences, that is, experiences that can radically change our core preferences? L. A. Paul (2014) has argued that many decisions involving transformative experiences cannot be rational. However, Paul acknowledges that some traumatic events can be transformative experiences, but are nevertheless not an obstacle to rational decision-making. For instance, being attacked by hungry sharks would be a transformative experience, and yet, deciding not to swim with hungry sharks is rational. Paul has tried to explain why decisions involving “sharky” outcomes are an exception to the rule. However, her putative explanation has been criticized by Campbell and Mosquera (2020). In this paper, I offer a different solution to this problem. Roughly, I argue that transformative experiences give rise a problem for rational decision-making only if the decision can lead to satisfying some of our (new) core preferences, but can also frustrate other (new) core preferences. I also argue that agents can *partially* project what traumatic transformative experiences are like.

**Keywords:** transformative experience, rational choice, expected value maximization, preferences, the Shark Problem

I look out my office window. It is a sunny winter day. It snowed during the night. In light of this, I decide to go for a walk later. Why? Because I prefer to spend my afternoon bathing in the sun, enjoying the fresh winter air, and admiring snow-covered gardens. My decision is guided by my preferences and what is likely to satisfy them.

Most mundane decisions are like this. We try to do what will meet our preferences. But what are we supposed to do when our decisions will radically change our core preferences? Should we try to satisfy our present preferences, or our future ones? And how are we supposed to know what our future preferences will be? These are some of the problems pertaining to transformative experiences, that is, experiences that can radically change our core preferences.<sup>1</sup>

---

<sup>1</sup> See Paul (2014) on transformative experiences. See Pettigrew (2019) for an overview of decision-theoretic principles for changing selves. See also Ullmann-Margalit (2006) for an earlier take on the rationality of transformative decisions. See Bykvist (2006) on the prudential ought and changing selves.

In an influential book, L. A. Paul (2014) has roughly argued that many decisions involving transformative experiences cannot be rational.<sup>2</sup> The gist of Paul’s claim is this. In order to decide rationally, you need to evaluate the various outcomes of a decision (and their likelihood). You want to know which outcomes are valuable, or preferable for you. However, transformative experiences might change your core preferences, and you cannot project what it is like to have a transformative experience. For Paul, this means that you cannot know if you will value some of the possible outcomes resulting from your decision. This puts you in an epistemically problematic position for rationally deciding what to do.

However, Paul acknowledges that some decisions involving transformative experiences can be rational. Suppose I could decide to swim with hungry white sharks. I know what a white shark is, and I have heard of shark attacks before. A reliable informant tells me that the sharks have not eaten for several weeks, and that they will attack me if I swim with them. It would be odd for me to think: “Ah, I cannot rationally decide not to swim with these sharks. After all, being attacked by sharks would change my core preferences, and I have no idea what this experience will be like.” Paul (2014, 26-7) has tried to explain why these types of scenarios are not an obstacle to rational decision-making. However, Campbell and Mosquera (2020) have argued that her putative explanation leads to a serious problem. This is what they call the Shark Problem.

In this paper, I offer a solution to the Shark Problem that is compatible with Paul’s central commitments concerning transformative experiences. Roughly, I argue that we can *partially* know what it is like to have some transformative experiences, like being attacked by sharks. I also argue that transformative experiences give rise to a problem for rational decision-making only if the decision can lead to satisfying some of our (new) core preferences, but can also frustrate other (new) core preferences. This condition excludes decisions like swimming with sharks. In section 1, I present Paul’s thesis and the Shark Problem more clearly. In section 2, I argue that agents can partially project what traumatic transformative experiences are like. This leads me, in section 3, to formulate and defend my solution to the problem.

---

2 See Paul (2014). See Pettigrew (2019, chap. 10) and Talbott (2016) for replies to Paul. See Campbell and Mosquera (2020) on the shark problem for Paul’s view. Dougherty, Horowitz and Sliwa (2015) argue that decisions involving transformative experiences can be rational, because transformative experiences are not an obstacle to knowing how valuable an experience is. Paul’s view concerning transformative experiences also matter for a host of related debates. See, e.g., Paul (2017) and Cappelen and Dever (2017) on indexicality and the first personal point of view.

# 1. Transformative Experiences and the Shark Problem

## 1.1. Personally Transformative Experiences

Some experiences change who we are. A personally transformative experience “changes you enough to substantially change your point of view, thus substantially revising your core preferences or revising how you experience being yourself” (Paul 2014, 16). Paul gives some examples of personally transformative experiences, such as:

1. *Radical life changes.* Some life changes, such as having a child or choosing a career, can be personally transformative. When one deliberates on whether to have a child, one does not know in advance what it is like to be a parent, and what kinds of preferences one will have as a parent. The experience of having a child can radically change your preferences. In order to know your preferences as a parent, one has to make this choice and live the experience of having a child.
2. *Gaining a new sensory ability.* Some changes in the ability to hear, feel, taste, etc. can be personally transformative. When a deaf person deliberates on whether to have a cochlear implant, he or she cannot know in advance what it is like to hear, and what it is like not to be part of the deaf community. The experience of hearing can radically change his or her preferences.
3. *Having entirely new experiences.* Some new experiences, such as eating durian or Vegemite, can be personally transformative. Durian has a unique and radically different taste from everything else. One’s future self might hate or love durian, but in order to figure out what one’s future self will enjoy, one has to try durian.
4. *Having a traumatic experience.* Some tragedies, such as having your limbs amputated without anaesthesia, or being attacked by a shark, can be personally transformative. These traumatic events can radically change our core preferences, desires and goals.

## 1.2. Rational Decisions and Transformative Experiences

Paul (2014) thinks that transformative experiences are a serious problem for rational decision-making. For Paul, rational decisions involve comparing outcomes based on what one prefers. However, in transformative experiences, one’s experiences will radically

change one's core preferences. This is a serious obstacle for deciding rationally what we should do.

First, what does Paul mean by rational? Paul says that rational agents satisfy the requirements of decision theory (Paul 2014, 19).<sup>3</sup> For instance, she thinks that rational agents maximize expected value (ibid., 21). Accordingly, rational agents often make decisions that are sensitive to their preferences. For Paul, your preferences are an important part of expected value maximization, because “what counts as valuable for you when making a particular choice will depend on your dispositions, beliefs, and desires” (ibid., 22).

Not everyone thinks that rationality consists in satisfying the principles of decision theory. Some think that rationality is mere coherence (J. Broome 2013; Worsnip 2018a, 2018b). Others think that rationality consists in responding to the reasons one has (Kiesewetter 2018; Lord 2018). For simplicity, I will stick to Paul's use of this term. Those who accept a different theory of rationality can, in the foregoing, replace “rational” with something like “optimize expected value,” and “rational deliberation” with “deliberation aimed at optimizing expected value.”

Here is why transformative experiences are an obstacle to rational decision-making. Consider the decision to have a child. Based on what one currently prefers, one might (or might not) be inclined to have a child. However, one's rational deliberation should also factor what one's future self will enjoy and dislike. After all, it is your future self that will live with the consequences of your decision. So, one's rational deliberation should be sensitive to the evolution of one's core preferences after making the decision to have (or not to have) a child. The problem is that the decision to have a child is a transformative experience. One does not know in advance what it is like to be a parent, and what kinds of preferences one will have as a parent. This kind of knowledge is only accessible to those who are familiar with the phenomenology of being a parent. Paul says:

When a newborn is produced, both parents experience dramatic hormonal changes and enter other new physiological states, all of which help to create the physical realizers for the intensely emotional phenomenology associated with the birth. These experiences contribute to the forming and strengthening of the attachment relation. Further characteristics of the nature of the attachment manifested between you and your child, ones which can have a dramatic effect

---

<sup>3</sup> Paul also refers to Weirich (2004), who is concerned with realistic requirements of decision theory (i.e., requirements of decision theory that can be useful and relevant for imperfect agents like us).

on the experience of being a parent, are determined by the particular properties of the actual child you produce. All of this generates the unique experience of having one's first child. (Paul 2014, 78)

So, for decisions such as having children, it is difficult to deliberate based on one's future core preferences.<sup>4</sup>

However, Paul acknowledges that some decisions involving transformative experiences are not an obstacle to rational decision-making. For instance, suppose you had fine-grained information about how *you*, personally, will respond to the transformative experience. For instance, suppose you had fine-grained information confirming that you will be a happy and fulfilled parent, or that deciding to have a child will satisfy all your future core preference. Then, perhaps you could rationally decide to have a child (ibid., 126). However, this kind of fine-grained information is often unavailable to us. As Paul says:

Empirical research, at least in the near future, isn't going to exhaustively *determine* whether you should decide to have retinal surgery or not, or whether you'd prefer being a violinist to being a doctor.... [This] prevents you from casually replacing your first personal evaluations with empirical data. (ibid.)

So, if this kind of information were available to us, transformative experiences would not be an obstacle to rational decision-making. But as it happens, this information is often unavailable to us. So transformative experiences are often an obstacle to rational decision-making.<sup>5</sup>

Here is another type of cases involving transformative experiences that is not an obstacle to rational decision-making. Having your limbs amputated without anaesthesia, or being attacked by a hungry white shark, are transformative experiences. Yet, you can rationally determine that swimming with hungry white sharks (or refusing anaesthesia) is a bad decision, even if it would involve a transformative experience.<sup>6</sup> But then, how can we distinguish "sharky" transformative experiences from regular ones? Why do these transformative experiences pose no problem for rationality?

---

4 Paul (2014, 38-40) discusses how decisions involving transformative experiences can have revelatory value, (i.e., the value of having new experiences). Some decisions involving transformative experiences can be rational because we want to have new experiences. But the problem remains if one does not prefer having new experiences.

5 See Pettigrew (2019, chap. 10) and Dougherty, Horowitz and Sliwa (2015) for discussion of this point.

6 Of course, there is a remote chance that swimming with hungry white sharks will be perfectly safe. I will come back to this complication in section 3.

### *1.3. Paul's Way of Dealing with Sharky Transformative Experiences*

Paul gives an explanation of why “sharky” transformative experiences are different from regular ones. Her explanation involves cognitive modeling.

According to Paul, when we deliberate about what we should do, we “mentally stop” and evaluate the various options available to us (Paul 2014, 26). Our goal is to select the decision that maximizes expected value (ibid.). In order to figure this out, we project forward different decisions and outcomes. Specifically, we project “experiential outcomes” that are described by “what it’s like” sentences (ibid.). So, when we deliberate, we ask ourselves what it would be like to make decision A, and what it would be like to make decision B, and so forth. This is what Paul calls cognitive modeling (ibid.). As she says:

When you are considering your options, you evaluate each possible act and its experiential outcomes by imagining or running a mental simulation of what it would be like, should you act, for each relevant possible outcome of each relevant act. You simulate the relevant possible outcomes for yourself, that is, you simulate what it would be like for you to have each of these experiences and what your life would be like after these experiences. (ibid.)

The cognitive simulations allow us to rank various options and decisions, from most to least preferable. However, when our decisions involve transformative experiences, we cannot run these cognitive simulations. We are not in a position to project experiential outcomes, because the decisions involve outcomes that are radically different than everything we have experienced before, and might radically change our core preferences.

However, Paul notes that some rational decisions are possible without running cognitive simulations. For Paul, swimming with hungry white sharks is one of them. One knows that such a decision is bad and irrational even without running cognitive simulations. So, one can rationally decide not to swim with sharks without figuring out precisely what it’s like to be attacked by sharks (ibid., 27).

Thus, the reason why decisions like “I will not swim with hungry white sharks” can be rational is that they do not require cognitive modeling. These decisions are bad, and we know this without running cognitive simulations.

#### 1.4. A Worry for Paul's Explanation and the Discontinuity Objection

A worry with Paul's explanation is this. On the one hand, Paul says that decisions involving transformative experiences are not rational, because (i) we cannot project the experiential outcomes of transformative experiences, a necessary condition for cognitive modeling, and (ii) we do not know how our core preferences will evolve after we go through a transformative experience. On the other hand, she says that we can rationally decide not to swim with sharks. For most of us, swimming with sharks would be a transformative experience. We cannot project the experiential outcomes of this decision.<sup>7</sup> Paul says this is fine, because we do not need cognitive modeling to make such rational decisions. But then, do we really need cognitive modeling for other, non-sharky decisions? Couldn't we say that cognitive modeling is *optional* for rational decision-making?

Campbell and Mosquera (2020) discuss an escape route on Paul's behalf, and argue that it fails. Here it is. Perhaps Paul thinks that there are two types of experiential outcomes. There are *normal outcomes*, such as getting a cochlear implant or becoming a parent. In order to make rational decisions that involve normal outcomes, one has to be familiar with the phenomenology of the outcomes (ibid., 3550-1). Paul's claims in terms of experiential outcomes and cognitive modeling apply to these outcomes. Then, there are *sharky outcomes*, such as being attacked by sharks or being hit by a bus (ibid.). These outcomes can be evaluated without cognitive modeling (i.e., projecting experiential outcomes). This escape route would, if successful, allow us to solve the tension described above. Paul's view would be restricted to decisions that involve *normal outcomes* only. It would not apply to decisions that involve *sharky outcomes*.

However, Campbell and Mosquera argue that that the putative distinction between normal outcomes and sharky outcomes raises a problem. They note that "a normal outcome can be gradually transformed into a sharky outcome" (ibid., 3552), which creates a discontinuity problem for evaluating outcomes. This is what they call the discontinuity objection.

Here is the gist of their argument. Suppose agents have to make a decision between different lives. All of their options are transformative experiences. These options could be: Become a parent, acquire a new sensory ability, become a vampire, etc. (for simplicity, we

---

<sup>7</sup> Or at least: We cannot fully project this. I will come back to this point in section 3 (this is part of my solution to the problem).

will call each of these options an “experientially different life”). Their options are located on a spectrum. At one end of the spectrum, agents will experience a very small amount of intense pain. Then, each option involves a small increase in the amount of intense pain they will experience. At the other end of the spectrum, agents experience the maximal amount of intense pain (ibid.). For instance:

O<sub>1</sub>: Choose experientially different life 1, which involves one second of intense pain per day.

O<sub>2</sub>: Choose experientially different life 2, which involves two seconds of intense pain per day.

...

O<sub>n</sub>: Choose experientially different life n. This life involves permanent intense pain.

The options on one end of the spectrum are normal outcomes (they are transformative, but they are not that bad). However, the options on the other end of the spectrum are sharky outcomes (they are transformative *and* really bad) (ibid., 3553).

Then, there must be a boundary or a threshold between the range of sharky outcomes and the range of normal outcomes (ibid.). The boundary could be sharp or vague. Campbell and Mosquera first assume that it is sharp. For instance, perhaps O<sub>5000</sub> is a normal outcome, but O<sub>5001</sub> is a sharky outcome. In this example, the sharp boundary would be O<sub>5000</sub> (later in their paper, Campbell and Mosquera discuss various generalizations of their arguments with vague thresholds, and epistemicist, supervenientist and ontic interpretations of vagueness. See pp. 3555-3560).

The discontinuity objection is this. Agents cannot compare various outcomes on the spectrum with each other, because they involve experientially different lives. For instance, agents cannot compare O<sub>1</sub> with O<sub>5000</sub>, because they cannot grasp what these lives are like unless they experience them. However, agents can conclude that O<sub>1</sub> is better than O<sub>5001</sub>, or that O<sub>5000</sub> is better than O<sub>5001</sub>, even if they (still) cannot grasp what these lives are like. As Campbell and Mosquera say, “we find it implausible that the agent’s ability to evaluate an outcome and compare it to O<sub>1</sub> vanishes due to the absence of just one fewer second of pain” (ibid., 3553).

Thus, introducing a distinction between normal outcomes and sharky outcomes creates a strange discontinuity in the way agents can evaluate outcomes. So, this escape route is unavailable to Paul. This means that we have not solved the puzzle discussed at the



beginning of this section: If cognitive modeling is optional for skanky outcomes, could we also conclude that that it is optional for *all* outcomes? And if cognitive modeling is optional for all outcomes, how can Paul argue that transformative experiences compromise rational decision-making?

## 2. An Escape Route

### 2.1. Core Preferences, Cognitive Modeling and Transformative Experiences

I agree with Campbell and Mosquera that Paul's explanation raises worries. Solving the Shark Problem with a distinction between types of experiential outcomes raises a serious problem. So, how are we going to draw the line between decisions such as becoming a parent and decisions such as swimming with sharks?

Here is what we are trying to do:

- (1) We want to accommodate the claim that some decisions involving transformative experiences, such as choosing between swimming with sharks and not swimming with sharks, can be rational.
- (2) We want to accommodate Paul's thesis, and in particular, her central claim that many decisions involving transformative experiences do not square well with rational decision-making.<sup>8</sup>
- (3) In accordance with Campbell and Mosquera's claim, we should not make sense of (1) and (2) by appealing to a distinction between normal outcomes that require cognitive modeling, and sharky outcomes that do not require cognitive modeling.

Roughly, my solution is this: First, there can be partial cognitive modeling. Second, transformative experiences give rise a problem for rational decision-making only if the decision can lead to satisfying some of our (new) core preferences, but can also frustrate other (new) core preferences. Decisions like swimming with sharks do not satisfy this condition, because they can only frustrate our core preferences. So, even though these decisions are transformative, they are not an obstacle to rational decision-making.

---

<sup>8</sup> There are worries for Paul's claim. See, e.g., Dougherty, Horowitz and Sliwa (2015) and Pettigrew (2015, chap. 10). However, I wish to remain neutral on whether Paul is right. So, I am trying to find a solution to the problem of sharky outcomes that is compatible with Paul's thesis.

In accordance with (1), my solution can explain why deciding not to swim with sharks is rational. In accordance with (2), my solution is compatible with thinking that transformative experiences are often an obstacle to rational decision-making. Finally, my solution does not rely on a problematic distinction between outcomes that require cognitive modeling and outcomes that do not, which tallies well with (3).

In this section, I explain why there can be partial cognitive modeling. In the next section, I explain why, relative to partial cognitive modeling, decisions like swimming with sharks do not pose a threat for Paul's argument.

## 2.2. *Sharks Attacks and Partial Cognitive Modeling*

For Paul, cognitive modeling is essential for rationally evaluating outcomes. Or more precisely: She thinks that, in the absence of fine-grained information about how one will personally respond to an experience, cognitive modeling is necessary for rational decision-making.<sup>9</sup> However, we cannot project what it is like to have a transformative experience. So, how can we know that decisions like swimming with hungry white sharks are bad?

In response to this worry, I will first argue that the claim that we cannot project what it is like to have a transformative experience is too strong. That is, I will argue that transformative experiences allow for *partial* cognitive modeling. We cannot *fully* grasp what it is like to be attacked by sharks. But we can grasp some important features of all the possible outcomes of this decision, and that is enough to warrant rational decision.<sup>10</sup>

How can agents partially project what it is like to be attacked by sharks? First, agents can foresee that deciding to swim with hungry white sharks will likely involve a great degree of pain. Agents can project what pain is like. Second, they can arguably know that this would be a traumatic event. Agents can project what traumatic events are like. Third, agents can figure out that the recovery from a shark attack will take time, and that if you lose a limb, you might need to adapt to your new reality, which can take months or years. Agents can project what it is like to adapt to a new, more complex reality. There is a number of things that agents can project when they are thinking of shark attacks.

---

<sup>9</sup> See section 1.2.

<sup>10</sup> This possibility is briefly mentioned, but not really explored, in Paul (2017, 199). See Dougherty, Horowitz and Sliwa (2015) on a related notion of partial knowledge.

This means that they can project *some* central aspects of decisions like swimming with sharks. Of course, perhaps agents cannot project what it is like to endure the specific psychological and physical trauma of being attacked by sharks. And they might not know how, exactly, their preferences will evolve after such a traumatic event. But their knowledge of pain and traumas in general give them a partial idea of what it is like to be attacked by sharks.

So, for decisions such as swimming with sharks, cognitive modeling can be helpful to realize that all outcomes are bad. Of course, you cannot know exactly what it is like to live such a traumatic experience. But you know what pain feels like, you know what traumas feel like, and so you can have a rough but sufficiently informative picture of what it is like to be attacked by sharks. This is part of the explanation why you can rationally decide not to swim with sharks.

Note, however, that partial cognitive modeling is sometimes insufficient for making rational decisions. Consider the decision to taste Durian. Suppose I am told by a reliable informant that ripe Durian has a creamy texture, like custard. The experience of creamy food is familiar to me. Accordingly, I can partially project what eating Durian is like. However, in this example, partial cognitive modeling would not solve the problem described by Paul for rational decision-making. For example, the smell and taste of Durian are radically different from everything I have experienced before, and eating Durian could change my culinary preferences. For Paul, these are obstacles to rationally deciding to eat Durian. The mere fact that I can project what the texture of Durian is like does not solve the problem.

For now, I merely want to argue that partial cognitive modeling is possible. I will come back to the rationality of decisions based on partial cognitive modeling in section 3.

### 2.3. *Is Partial Cognitive Modeling Ad Hoc?*

The claim that cognitive modeling can be partial, or can come in degrees, is not *ad hoc*. Many ordinary experiences suggest that we can partially project what it is like to have some radically different experiences. This happens when we are empathetic towards people with radically different lives.<sup>11</sup> Consider the following case:

---

<sup>11</sup> See DeTurk (2001) on definitions of empathy, and on the relationship between intercultural dialogue and the ability to “put yourself in the shoes” of members of other cultures. For DeTurk, there are serious

**Discrimination.** John is a white man. Diane is a black woman. Diane often describes the discrimination and racism she experiences to John. He is not in a position to fully grasp what it is like to be a black woman. But he gets what it is like to suffer, to feel in danger, to feel excluded, and so forth. So while he cannot fully project what it is like to be a black woman, he can have empathy, and he has enough information to know that Diane's experiences are bad and frustrate her preferences.

Cases like Discrimination suggest that partial cognitive modeling is common. People like John cannot fully grasp some experiences (like racism, sexism, homophobia, transphobia, and so forth). But arguably, John can partially project some of the things Diane experiences based on her testimony. This suggests that partial cognitive modeling is possible.

Here is another way to put it. Suppose Diane tells John that she earns less than her white male colleagues (at the same level of education, experience, etc.). Diane also tells John that her boss frequently makes sexist and racist comments in front of her. Then, suppose John says the following:

“But... is this bad for you? I really have to ask. You see, I cannot project what it is like to experience these things. I cannot imagine what you are going through. I would have to be in your shoes to know what this is like. But I cannot, since I am a white male. So tell me: Is it bad for a woman to hear sexist comments all the time, or to be paid less than everyone else?”

Here, John's cluelessness is surprising. Of course, John does not fully grasp what it is like to experience racism or sexism. But John can imagine what it is like when people do not accept who you are, or when you do not get what you deserve. So, John is in a position to know that violations of pay equity and racist comments frustrate Diane's preferences. Only partial cognitive modeling can explain this. John cannot *fully* project Diane's situation to himself. However, if John were *totally unable* to project Diane's situation to himself, he would not be able to empathize. Therefore, there must be some sort of intermediate projection that John has access to.

Thus, we cannot *fully* project transformative experience through cognitive modeling. However, we can have a *partial* picture of transformative experiences through cognitive modeling. And for decisions like swimming with sharks, partial cognitive modeling might be enough to make rational decisions. We do not fully grasp what it is like

---

obstacles to thinking that “members of one social group [can] understand the experiences of another, particularly across divisions of power” (ibid., 377). However, she mentions various studies and examples in which these obstacles are not unsurmountable (ibid., 377-81). Compare with Bailey (2018) and B. J. Broome (2017). See Paul (2017) on empathy and the metaphysics of first personal perspectives, and Cappelen and Dever (2017) for a reply.

to be attacked by sharks, but we might be in a position to rationally decide not to swim with sharks.

However, one could object that the kind of partial cognitive modeling involved in a case like Discrimination differs importantly from the kind of partial cognitive modeling involved in shark attacks. Specifically, empathy is central in cases like Discrimination, but seems absent from the partial cognitive modeling of shark attacks. So, why think that the observations pertaining to cases like Discrimination matter for the partial cognitive modeling of traumatic events?<sup>12</sup>

My response to this worry goes as follows. I think there are important similarities between the two cases. Specifically, empathy can play a similar role for partial cognitive modeling of shark attacks. For example, suppose John meets some shark attack survivors (or watches a documentary about shark attack survivors). The victims talk about the phantom pain they are dealing with, the activities they can no longer engage in, and so on. It would be surprising if John were clueless and asked questions like this:

“But... Do you enjoy being in pain and not having the opportunity to engage in certain activities? I cannot imagine what you are going through. I would have to be in your shoes to know what this is like. But I cannot, since I am not a shark attack survivor. So tell me: Is it bad for a shark attack survivor to suffer, or to have less capabilities?”

John’s cluelessness is as surprising as before. And that is because John is able to empathize. John cannot *fully* project a shark attack survivor’s situation to himself. But he can figure out what it is like to suffer, or to have less capabilities.

### **3. Core Preferences and the Rationality of Deciding not to Swim with Sharks**

Partial cognitive modeling is plausible. It is a useful source of information for rational decision-making. Even if it does not allow us to fully grasp what traumatic transformative experiences are like, they help us see what is bad about them.

But then, partial cognitive modeling can also be a source of information for other decisions, such as having a child. We still need to account for the difference between these two types of decisions. If we can rationally decide not to swim with sharks based on partial

---

12 I thank a referee for inviting me to discuss this objection.

cognitive modeling, why can't we make rational decisions pertaining to parenthood on the same ground?

In this section, I argue that transformative experiences give rise a problem for rational decision-making only if the decision can lead to satisfying some of our (new) core preferences, but can also frustrate other (new) core preferences. Based on partial cognitive modeling, agents can know that swimming with sharks violates this condition. However, the same is not true for decisions like having a child.

### *3.1. Shark Attacks and Rational Decision-Making*

We do not know exactly how our core preferences will evolve after a traumatic event, such as a shark attack. But we know that some core preferences will not go away, such as: having capabilities, avoiding pain, etc. Decisions like swimming with sharks, or having your limbs amputated without anaesthesia, can change some of one's core preferences. Yet, suffering or losing capabilities goes against one's future core preferences, and one knows, based on partial cognitive modeling, that swimming with sharks can hardly contribute to satisfying our (new) core preferences. One does not need to fully grasp what these experiences are like to have that kind of knowledge. So, one has enough information, from partial cognitive modeling, to rationally exclude this decision.

For concreteness, suppose that, after being attacked by sharks, one's preferences could evolve as follows:<sup>13</sup>

New Core Preferences 1: Avoiding pain, learn first-aid techniques.

New Core Preferences 2: Having capabilities, lobby for the use of shark nets or the increased monitoring of shores.

One does not know what one's core preferences will be after being attacked by sharks (since there are two possible sets of future core preferences, and they are different from each other). But one can see that being attacked by sharks will only frustrate one's core preferences. Nothing in these possible future preferences structures will be served by a shark attack. So, one cannot know exactly what the pain or the loss of capabilities resulting

---

<sup>13</sup> These putative new core preferences are inspired by the testimonies of shark attack survivors. See, e.g., Adolph (2019) and Carlsen (2012).

from shark attacks is like, but one can rationally decide, based on their information and their possible future preferences, not to swim with hungry sharks.<sup>14</sup>

Thus, even if swimming with sharks is a transformative experience, one knows that swimming with sharks can only frustrate one's preferences. This is why we can rationally rule out this decision.

### *3.2. Having a Child and Obstacles to Rational Decision-Making*

Paul claims that many decisions involving transformative experiences, such as having a child, are a problem for rational decision-making. My solution is compatible with this possibility. Transformative experiences involve a radical change in one's preferences. They are a problem for rational decision-making only if the decision will lead to satisfying some of our (new) core preferences, but will also frustrate other (new) core preferences. This condition is satisfied when one decides to have a child. After having a child, one's preferences could evolve as follows:

New Core Preferences 1: Seeing friends, sleeping well, have no responsibilities, etc.

New Core Preferences 2: Unconditional love from a child, continuation of the family line, etc.

With respect to New Core Preferences 1, the decision to have a child will frustrate one's core preferences. However, with respect to New Core Preferences 2, the decision to have a child will satisfy one's core preferences. This is why the rational evaluation of our decision is so hard: We do not know exactly how our preferences will evolve over time.

Naturally, some aspects of parenthood can run counter to our core preferences. For example, we have basic needs, such as eating and sleeping. These basic needs are part of our core preferences. Having a child can seriously affect the quantity and quality of our sleep. But this inconvenience pertaining to one of our central preferences can be compensated for by satisfying new, future core preferences. For example, the experience of parenthood can bring out new core preferences, such as the unconditional love from family members, or the continuation of the family line. And one's future self might tolerate sleep deprivation if this is compensated for by the satisfaction of such other core preferences.

---

<sup>14</sup> Also, one knows that being attacked by sharks will not contribute to the satisfaction of one's other preferences (e.g., it will not help one advocate for the rights of people with disabilities, it will not help one lobby against the protection of white sharks, etc.).

This is an important dissymmetry with decisions like swimming with sharks. Being attacked by sharks will lead to intense, long-lasting physical and psychological pain. Avoiding pain is one of our core preferences. We know this is not going to change. Also, we know, through partial cognitive modeling, that this great inconvenience will not be compensated for by the satisfaction of new core preferences.

### *3.3. Core Preferences and the Putative Distinction Between Normal Outcomes and Sharky Outcomes*

The solution I offer is compatible with denying that there is an essential distinction between types of outcomes. Recall that Campbell and Mosquera discuss the following escape route on Paul's behalf: The reason why decisions like not swimming with sharks are rational is that decisions involving sharky outcomes do not require cognitive modeling, while decisions involving normal outcomes require cognitive modeling. Campbell and Mosquera raise a serious worry for this distinction—namely, that the putative distinction between normal outcomes and sharky outcomes lead to a discontinuity problem in rational evaluation.

My solution is that it is hard to make rational decisions involving transformative experiences when the transformative experiences can lead to satisfying some of our (new) core preferences, but can also frustrate other (new) core preferences. This is compatible with thinking that all rational decisions involve partial or full cognitive modeling. Or, if you prefer: The solution is compatible with denying the problematic distinction between types of decisions that require cognitive modeling and types of decisions that do not require cognitive modeling. So, my solution does not face Campbell and Mosquera's worry.

One could object that my solution falls prey to another version of Campbell and Mosquera's worry. That is, one could argue that, on my solution, there is a discontinuity between the outcomes for which one can apply partial cognitive modeling, and other outcomes for which one cannot apply partial cognitive modeling.<sup>15</sup>

However, I think that my solution does not face this problem. For illustration purposes, consider the options on a spectrum discussed in section 1:

---

<sup>15</sup> Another formulation of the worry could be this: There is a discontinuity between the outcomes for which partial cognitive modeling is sufficient for rational evaluation, and other outcomes for which partial cognitive modeling is insufficient for rational evaluation. I also address this possibility in the next paragraphs.



O<sub>1</sub>: Choose experientially different life 1, which involves one second of intense pain per day.

O<sub>2</sub>: Choose experientially different life 2, which involves two seconds of intense pain per day.

...

O<sub>n</sub>: Choose experientially different life n. This life involves permanent intense pain.

My view says that there are *some things* we can evaluate in each of these options. Partial cognitive modeling is possible for each of these options. So, there is no discontinuity between the outcomes for which one can apply partial cognitive modeling, and other outcomes for which one cannot apply partial cognitive modeling. All the above outcomes allow for partial cognitive modeling.

For instance, we can *at least* evaluate the amount of intense pain involved in each of these options. We know that the pain involved with each option is getting worse and worse, and we can partially evaluate each option on such grounds. So, we can make some partial comparisons between O<sub>1</sub>, O<sub>2</sub>,..., and O<sub>5001</sub>. This is an important difference between my solution and Paul's. We are not left with the problematic result that, although we can compare O<sub>5000</sub> and O<sub>5001</sub>, we cannot compare O<sub>1</sub> and O<sub>5000</sub>.<sup>16</sup>

At some point, the option might involve enough pain, so that a rational agent will never choose this option. Suppose this is option O<sub>5001</sub>. On my view, there is no discontinuity between this option and the previous one. If there is too much pain involved in O<sub>5001</sub>, you will also think that there is a serious problem with O<sub>5000</sub>. That is, in O<sub>5000</sub>, pain is *almost* sufficient to rationally exclude this option. Pain can be evaluated in each of these options, and it is a *growing* concern in each of them.

More generally, it is implausible that my view will face a discontinuity objection. This type of objection targets views that sharply distinguish types of outcomes (like: the ones that do not require cognitive modeling, and the ones that can be evaluated without cognitive modeling). But my view is compatible with denying such sharp distinctions. The reason why rational agents can ignore options involving sharky outcomes is not that there is

---

16 This is Campbell and Mosquera's objection. Suppose that O<sub>5000</sub> is a normal outcome while O<sub>5001</sub> is a sharky outcome. Then, the crux of Campbell and Mosquera's discontinuity objection is this: Agents cannot compare O<sub>1</sub> with O<sub>5000</sub>, because they cannot grasp what these lives are like unless they experience them. However, agents can conclude that O<sub>5000</sub> is better than O<sub>5001</sub>. This discontinuity in the evaluations agents can make is odd.

a sharp distinction between sharky outcomes and other ones. We can treat all outcomes “equally,” with a graded notion of cognitive modeling.

### *3.4. Other Cases: Heavenly Outcomes*

Finally, the solution I offer is compatible with thinking that other types of decisions involving transformative experiences are not problematic. So far, most of our discussion has been revolving around bad decisions, such as swimming with hungry sharks. These decisions involve bad transformative experiences, and thus can be rationally dismissed. The argument should also apply to corresponding good transformative experiences. For instance, consider the following case:

**Talent Pill.** You are a professional musician with a precarious career. You are offered a “talent pill” that will give you the ability to be one of the best performers in the world. For instance, the pill will give you the will and the ability to concentrate and improve, so that you become one of the best performers in the world. Surely, you do not know what it is like to be one of the best performers in the world. This would be a whole new experience to you. However, you know that this decision will not frustrate your future preferences.

In Talent Pill, the decision has “heavenly” outcomes (Campbell and Mosquera 2020, 3552; Sullivan 2018). You may not know exactly what the outcomes will be like, and how your future preferences will be satisfied, but you know that the outcomes will be good and will satisfy your preferences. Plausibly, you can rationally choose to take the pill. The fact that this decision involves a transformative experience does not seem to be an obstacle to rational decision-making. My solution is compatible with this possibility, because the decision to take the pill can lead to satisfying some of your (new) core preferences, but cannot frustrate other core preferences.

### *3.5. The Objection from Remote Possibilities*

Here is another objection against my solution. There is a remote possibility that deciding to swim with hungry sharks will satisfy one’s core preferences. Perhaps the sharks will not attack you. Or perhaps your future core preferences will evolve in an unpredictable way, and your decision to swim with sharks will contribute to satisfying these preferences. So, how can we be *sure* that swimming with sharks can only frustrate our preferences?

This kind of objection has to do with the contentious issue of relevant alternatives in rational decision-making. It is an open question how rational agents should deal with

alternatives or possibilities that are highly unlikely. However, this issue is not specific to solution I have offered to the Shark Problem. This is a tangential problem pertaining to rational belief and decision-making *in general*.

Compare: In epistemology, philosophers often assume that, if rational agents have no evidence that they are deceived, their deliberation can ignore or exclude remote possibilities (i.e., events or alternatives that are highly unlikely) such as “I am a deceived brain in a vat” (e.g., Worsnip 2019). In ethics, philosophers often assume that rational agents can make a decision based on a set of *relevant* alternatives and outcomes (e.g., Wedgwood 2017). Suppose that ignoring such remote possibilities is correct. Then, just as one can ignore remote possibilities like “I am a deceived brain in a vat,” one can ignore remote possibilities like “Being attacked by a shark will satisfy my preferences.” In both cases, we are dealing with the same phenomenon—namely, events or possibilities that are highly unlikely.

#### **4. Conclusion**

Transformative experiences could sometimes be an obstacle to rational decision-making. Yet, not all decisions involving transformative experiences raise a problem for rationality. The decision to avoid swimming with hungry white sharks is one of them. How can we explain this? For Paul (2014), the reason why decisions involving sharky outcomes are not an obstacle to rational decision-making is that this type of decision does not require cognitive modeling. Campbell and Mosquera (2020) have argued that introducing a distinction between normal and sharky outcomes leads to a discontinuity in the rational evaluation of outcomes. I agree with Campbell and Mosquera that we should not make a distinction between types of outcomes. My solution to this problem is that some decisions involving transformative experiences allow for partial cognitive modeling. This allows us to see that some decisions, like swimming with hungry white sharks, can only frustrate our preferences. Thus, rational agents can dismiss such options.

**Acknowledgments.** This research was supported by the Social Sciences and Humanities Research Council (grant #756-2019-0133).

## References

- Adolph, F. (2019). Shark attack: five survivors on what it's like when a great white tries to eat you. *The Guardian*, January 2 2019.
- Bailey, O. (2018). Empathy and Testimonial Trust. *Royal Institute of Philosophy Supplements* 84: 139-160.
- Bykvist, K. (2006). Prudence for changing selves. *Utilitas* 18(3): 264-283.
- Broome, B. J. (2017). Intercultural Empathy. *The International Encyclopedia of Intercultural Communication*, 1-6. DOI: 10.1002/9781118783665.ieicc0054
- Broome, J. (2013). *Rationality Through Reasoning*. Oxford: John Wiley & Sons.
- Campbell, T., & Mosquera, J. (2020). Transformative Experience and the Shark Problem. *Philosophical Studies* 177: 3549–3565.
- Cappelen, H., & Dever, J. (2017). Empathy and Transformative Experience Without the First Person Point of View (a reply to LA Paul). *Inquiry* 60(3): 315-336.
- Carlsen, Jack. 2012. How I survived a shark attack and learned some valuable lessons. *The Conversation*, January 17 2012.
- DeTurk, S. (2001). Intercultural Empathy: Myth, Competency, or Possibility for Alliance Building? *Communication Education* 50(4): 374-384.
- Dougherty, T., Horowitz, S., & Sliwa, P. (2015). Expecting the Unexpected. *Res Philosophica* 92(2): 301-321.
- Kiesewetter, B. (2017). *The Normativity of Rationality*. Oxford: Oxford University Press.
- Lord, E. (2018). *The Importance of Being Rational*. Oxford: Oxford University Press.
- Paul, L. A. (2014). *Transformative Experience*. Oxford: Oxford University Press.
- Paul, L. A. (2017). First Personal Modes of Presentation and the Structure of Empathy. *Inquiry* 60(3): 189-207.
- Pettigrew, R. (2019). *Choosing for Changing Selves*. Oxford: Oxford University Press
- Sullivan, M. (2018). *Time Biases: A Theory of Rational Planning and Personal Persistence*. Oxford: Oxford University Press.
- Talbott, W. (2016). Transformative Experience. *Analysis* 76(3): 380-388.
- Ullmann-Margalit, E. (2006). Big Decisions: Opting, Converting, Drifting. *Royal Institute of Philosophy Supplement* 81(58): 157-172.
- Wedgwood, R. (2017). The Predicament of Choice. *Oxford Studies in Metaethics* 12: 294-313.
- Weirich, P. (2004). *Realistic Decision Theory: Rules for Nonideal Agents in Nonideal Circumstances*. Oxford: Oxford University Press.
- Worsnip, A. (2018a). The Conflict of Evidence and Coherence. *Philosophy and Phenomenological Research* 96(1): 3-44.
- Worsnip, A. (2018b). What is (In)coherence?. *Oxford Studies in Metaethics* 13: 184-206.
- Worsnip, A. (2019). What to Believe about your Belief that you're in the Good Case. *Oxford Studies in Epistemology* 6: 206-233.