Neophobia

John Collins

Abstract: L. A. Paul argues that epistemically transformative choice poses a special problem for standard theories of decision: when values of outcomes cannot be known in advance, deliberation cannot even get started. A standard response to this is to represent ignorance of the nature of an experience as uncertainty about its utility. Assign subjective probabilities over the range of possible utilities it may have, and an expected utility for the outcome can be figured despite the agent's ignorance of its nature. But this response to Paul's challenge seems inadequate. Decision theory should leave conceptual room for rational neophobia. A decision theory like Isaac Levi's, which allows for indeterminacy in utility, might accomodate the phenomenon. Levi's discussion of indeterminate utility has focused on examples of risk aversion like the Allais problem and on situations in which there are conflicts of value. Cases of unknowable value arising in transformative choice problems might be handled similarly.

L. A. Paul defines a transformative experience to be one which is both epistemically transformative and personally transformative. An experience is epistemically transformative when there is no way of knowing in advance what the experience will be like, because actually having the experience is the only way of coming to know what it is like. An experience is personally transformative when it is "life-changing in that it changes what it is like to be you, that is, it changes your point of view, and by extension, your personal or subjective preferences" (2015, 16).

Paul argues that such experiences "constitute a class of experiences that raise a special problem for rational decision-making" (2015, 17). And in fact this seems straightforwardly to be the case. Suppose that one is deliberating about whether or not to undergo a transformative experience. Following Paul, let's call such a decision problem a transformative choice problem. Then you are deliberating about what sort of person to become in the future, and in particular you are deliberating about what sort of preferences your future self should have. But some of these possible future preferences might be quite different from your present preferences. They might disagree with your present preferences in various ways. In fact they

Res Philosophica, *Vol. 92, No. 2, April 2015, pp. 283–300*
http://dx.doi.org/10.11612/resphil.2015.92.2.6

might conflict with your present preferences on the very question of whether it is good to be, or to become, a person with such preferences. In such a case, where there is a clash between prior- and envisaged post-choice preferences, it is far from clear which should rationally prevail. That is, it doesn't seem right to say that such conversions can always be justified *ex post facto* from one's transformed point of view. After all, the person one has become might have views that from one's former standpoint seem completely reprehensible. But neither does it seem correct to say that one's prior preferences should always win out either. Mightn't there be genuine cases of enlightenment where your later self thinks quite rightly: I'm a better person now for having undergone that change? And mightn't one add: And I'm better in ways that I simply wouldn't have appreciated beforehand?

For those reasons I think that Paul is absolutely correct in thinking that standard accounts of rational decision-making have a deep difficulty in accounting for choices concerning personally transformative experience.

My interest here, however, is with a parallel problem for decision theory that Paul sees as arising in decision problems involving options that are merely epistemically transformative, like, for example, the decision whether or not to try a new and unfamiliar type of food. This forms one of the major threads running through Chapter 2 of her book, the chapter entitled: "Transformative Choice." I find Paul's argument curiously compelling but also quite elusive. It is my aim in the present paper to explain what I find difficult about Paul's line of thought about epistemically transformative decision problems, and also to attempt to explain why, even when certain distracting side issues are cleared up, there remains a significant core truth here that Paul is sensitive to. I would like to try to display that truth in a way that is free from what to my mind are the distracting side issues.

Paul writes:

> The key to understanding the problem that transformative experience raises is to recognize that the standard models for ignorance can only function if they can represent the structure of the value space of the outcomes for a decision problem. . . .
>
> As a result, in order to use these models for a decision made under conditions of ignorance, *you must be able to know the values of the of the relevant outcomes*. You do not need to know the probabilities that the outcomes, given the acts, will occur, but you do need to know how to value the relevant outcomes. A way to put this is that you must be able to describe the state space of your outcomes, and you must have a suitably defined value function for these outcomes. If you cannot know the values of the relevant outcome or if the values are not yet determined, so that you cannot describe the state space or assign values that

> will remain constant to outcomes, you do not have the information you need to use these types of models to represent your decision. For without an adequate description of the space and without a suitable defined value function for the outcomes, you cannot know if the structure of any particular model adequately represents the structure of the actual situation. (2015, 30–31)

We might summarize the line of argument like this. Deliberation cannot even get started unless the decision maker knows the values of the possible outcomes. When options are epistemically transformative, their values cannot be known in advance. Hence in epistemically transformative choice problems deliberation cannot get started.

There's a picture of decision making behind all this that we might call the simulation model of deliberation. In other passages Paul is quite explicit about this picture:

> 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.
>
> After you run each cognitive simulation, you assign each outcome a subjective value. . . . [O]nce you've determined the overall subjective value of each outcome, you can compare the expected values of different possible acts to determine which one you should perform. (2015, 26–27)

This simulation model of deliberation assumes what Philip Pettit has called the idea of decision theory as a *calculus* for decision making (1991). In order to understand this idea, we shall need to focus a little on the details of the standard theory.

Common to all the standard accounts of decision theory is the idea that rational choice is choice that maximizes expected value. The agent is supposed to have a subjective probability function that assigns credences to all the various possible states of the world, and a subjective utility function that assigns real number values to possible outcomes. This utility function is only unique up to positive affine transformation, in other words both the choice of unit size, and the location of the zero point are arbitrary. (This kind of scale dependence is familiar to us from the case of temperature measurement. Degrees Fahrenheit can be obtained from degrees Celsius by the following affine transformation: multiply by 9/5 and add 32.)

Then the expected value of each of the agent's options can be calculated as a credence-weighted average of the utilities of each of the possible

outcomes of that option. The picture of decision theory as a calculus for decision-making is the natural idea that the process of deliberation mimics this formalism; it is the idea that when a rational decision-maker deliberates, she engages in something like this calculation of expected utilities as subjective probability weighted averages of the utilities of the possible outcomes, where the utilities of the individual outcomes have been arrived at prior to all this by the method of mental simulation.

If this is one's picture of rational deliberation, then it is difficult not to agree with Paul's claim that deliberation cannot even get started unless the decision-maker already knows the values of outcomes.

This picture of rational deliberation goes hand-in-hand with a *psychological realism* about utility and credence. (See, e.g., Buchak 2013, 17.) According to the psychological realist, utility and credence are real mental states. Think of them as degrees of desire and degrees of belief respectively.

For present purposes I'm happy to assume this realist picture. But it will be useful for our purposes to follow Jamie Dreier in drawing a further distinction between two kinds of psychological realism about utility and credence. (See Dreier 1996 and Buchak 2013, 17–18.) Let's focus on the case of utility. Dreier distinguishes between a *constructive* and a *non-constructive realism* about utility. At issue is whether or not facts about an agent's utilities go beyond the facts about the agent's preferences between options. The constructive realist is someone who believes that they do not. According to the constructive realist, all the facts about the agent's utility function supervene on the facts about her preferences. So, for example: the fact that outcome y lies exactly halfway between outcomes x and z on the agent's utility scale is simply the fact that the agent is indifferent between y and a gamble that gives her a fifty percent chance of outcome x and a fifty percent chance of outcome z. For the constructive realist, such facts about preference are constitutive of what it is to have a particular utility function.

A non-constructive realist, on the other hand, thinks that it is possible, in principle, for there to be facts about the utility function that outstrip the facts about what the agent prefers. So according to a non-constructive realist about utility, it might be possible, for example, to access the facts about one's own utility function by direct introspection, or, perhaps, by the method of mental simulation of outcomes that Paul describes.

Now Paul notes that that "for simplicity" she is assuming that "values or utilities assigned to outcomes are psychologically real for the agent (even if, for example, utilities turn out to be partially constituted by their role in preferences)" (2015, 21, fn. 25). But it seems to me that it is only on the assumption of a non-constructive realism about utility that her "deliberation cannot even get started" argument can be made to seem at all plausible.

Suppose that one adopts a constructive realism about utility. Then the whole idea of direct access to one's utilities for outcomes via introspection and mental simulation will seem completely implausible. In fact the whole

idea of decision theory as a calculus for decision making will seem misguided. For the constructive realist, decision theory will be better viewed as what Pettit calls as a *canon* rather than a calculus for good decision making. For the constructive realist, preference is conceptually prior to utility. Any agent whose preferences are coherent in the sense that they satisfy the axioms of formal decision theory can be seen as choosing rationally so as to maximize expected utility, that is, so as to best serve her desires according to her beliefs, where those desires and beliefs, construed as admitting of degree, are quite real, but are nothing over and above that coherent pattern of preference to which she is disposed.

From this viewpoint it seems quite clear what the decision theorist should say about cases of epistemically transformative choice. Since it is impossible to know what an outcome of such a choice will be like in advance of actually having made the choice and experienced the outcome, the method of simulation is unavailable. But so what? In such a situation an "outcome" will in turn be a risky prospect that delivers, with subject probabilities determined by the agent's coherent preferences, various possible utilities if the world turns out to be one way, or another, with respect to how it would turn out to feel like to be the agent experiencing that outcome.

That the precise phenomenological character of each of these "refined" outcomes cannot be anticipated is neither here nor there. Remember: we are working in a decision theoretic framework according to which all that is relevant to the rationality of an agent's choices are the utilities she assigns to outcomes and the credences she gives to possible states of the world. Nothing else is relevant. In particular: further facts about the particular phenomenological character of the outcomes are not relevant. Once one gives up on the non-constructive realist idea that utility is conceptually prior to preference, and thinks instead of the utility function as contructed out of facts about coherent preference, there is nothing at all paradoxical or puzzling about this picture of things: in-principle ignorance as to the precise value of an outcome of an epistemically transformative choice problem simply gets represented, in the usual and obvious way, as a gamble that might yield any one of a range of possible utility values, depending on how things things turn out to be.

So far this all sounds as though I am unsympathetic to Paul's claim that epistemically transformative choice poses a problem for standard decision theory. But that's actually not the case. As I said earlier, I think there's a core of truth to what Paul is claiming. The rest of the paper will be devoted to explaining one way of starting to make good on this claim. It's offered as a friendly amendment to the argument of the second chapter of Paul's book, and she is welcome to accept it or reject it as she sees fit.

Here's the rough idea. Various critics of standard decision theory have argued that decision theory is lacking in that it allows no room for a rational aversion to risk. Similarly—I think—reflection on Paul's epistemically transformative choice examples might lead one to think that the standard

theory is impoverished in another important respect. It's impoverished in that it leaves no conceptual room for what one might call rational *neophobia*. Then in so far as neophobia should not be seen as irrational-in-principle, it will follow that Paul's examples do offer a new and serious challenge to standard accounts of rational choice.

'Neophobia' is a term used in the psychological literature to refer to an abnormal fear of anything new. (Sometimes this is referred to instead as *cainophobia* or *cainotophobia*.) One particularly common form is food neophobia, as many parents of young children well know. Here I will use the term in a neutral way that is not intended to suggest that there is anything abnormal, or pathological, or irrational about this kind of preference structure. Neither do I want to suggest that neophobia is either more or less common than, or more or less reasonable than, the opposing tendency: neophilia (nor, for that matter, to a ceteris paribus indifference toward outcomes that are new and alien in Paul's sense of being epistemically transformative).

Now let's consider what the preferences of a neophobic agent might look like.

In particular, let's consider situations in which a person is confronted with a choice problem in which one of the options has outcomes with which she is experientially unacquainted. For the sake of simplicity I will focus on just the kind of example that Paul introduces: a situation in which an available option is to try some sort of food of a kind that the agent has never previously tasted and in which it might be reasonable to think that the experience of trying it for the first time might be radically unlike any kind of taste experience the agent has ever had in the past. To be definite: let's imagine that the agent, having never previously eaten durian, is now faced with a choice situation in which one of the options is to taste it for the very first time. For the uninitiated: durian is a kind of fruit native to Southeast Asia. Reported opinions about it vary wildly. It has a distinctive smell that some find pleasant, while others find completely disgusting. All agree, however, that the distinctive aroma and taste of the durian fruit are impossible to convey to someone who has never experienced eating it.

Part of the reason for choosing this kind of example is the fact that it seems fairly safe to say, with Paul, that opting for such an outcome will be epistemically transformative for the agent without being personally transformative. Once I've tasted durian for the first time I'll have learned something that I could not possibly have learned in any way other than by actually having had the experience. But at the same time it seems fairly safe to say, in advance, that whatever that experience turns out to be like, it's not going to change in any deep, or important, or fundamental way, the kind of person that I am. It's not, for example, going to result in any change to my core values or preferences, and this is something which, in turn, I can be fairly sure of ahead of arriving at a decision.

So there's the radically unknown, and unknowable option of durian, say, available on the menu. How, according to a standard theory of choice, is the agent supposed to evaluate this option?

The standard proposal, rehearsed earlier, is to represent the agent's ignorance about what the experience of tasting durian will be like as ignorance over a range of possible outcomes in which the experience of tasting and smelling the fruit turns out to be more or less pleasurable (or unpleasant). Now even though the particular felt qualities of the possible experiences in this range cannot be described or anticipated in advance, the idea is that that should not matter, because all of that unattainable information is going to be filtered through the lens of the agent's utility function anyway. Ultimately—so the orthodox story goes—all that is going to end up mattering to the theory of rational choice are the utilities that the agent would assign to each of those possible experiential scenarios were they to turn out to be actual. If that is correct, then the particular, and ungraspable, felt quality of various of those experiences simply falls out of the picture. The adequacy of the standard theory is defended by representing all of that ignorance as simply ignorance as to what the utility of the experience will actually turn out to be.

Now another reason for favoring an illustrative example of this fairly trivial sort is that it also seems fairly safe to say at this point that whatever the experience of tasting the fruit turns out to be like, its utility can be anticipated to fall within a certain range of possible values, so the agent can confidently place upper and lower bounds on how good or bad the experience will turn out to be. So let's assume that we have good evidence that enables us to set aside, for example, such possibilities as that the fruit will turn out to be poisonous, or that it will send the agent into anaphylactic shock, or that it will trigger some other kind of allergic reaction. Similarly, at the other end of the scale, let's suppose that the agent can safely assume in advance that the experience is not going to be *so good* that it will turn out to be "off the charts" in the sense of being better, and of course in an unanticipatible way, than some value set in advance as the maximum possible utility.

Once we have this upper and lower bound to the possible utility of the unknowable experience set, then the idea will be that we can, in principle, go about the task of constructing a kind of *synthetic lottery* over a range of quite familiar outcomes, a synthetic lottery that can then go proxy for the outcome that involves the epistemically transformative experience.

We need not suppose that this lottery have a continuum of possible prizes corresponding to all of the real numbers that are the possible utilities in the interval between the minimum and maximum values. We may suppose that what I'm calling the synthetic lottery has only some finite number of outcomes or prizes. The important thing, however, is that all of those outcomes must involve experiences that are quite familiar to the agent, and that the known utility of each outcome must lie somewhere on the closed

interval between the upper and lower bounds, and that those utilities be sufficiently well distributed, or uniformly spread, over the interval so that whatever the epistemically transformative experience turns out to be like, it will also turn out to have a utility, for the agent, that is very close to the utility of one of the prizes in what I'm calling the synthetic lottery. Again: what one means here by "very close" can simply be adjusted, if required, by increasing the finite number of prizes.

We are now in a position to see what the preferences of what I'm calling a neophobic agent might be like.

Suppose that an agent confronts a decision problem in which some option A is epistemically transformative. Construct a synthetic lottery corresponding to the experientially unknowable option A so that:

(1) For any possible utility value x that the epistemically transformative experience may turn out to have for the agent, there is a possible outcome to the lottery that is both (a) experientially familiar to the agent and (b) has a utility that is (arbitrarily) close to x.
(2) The chances of the various possible outcomes to the lottery are weighted so as to correspond to the agent's subjective probability distribution over the range of possible utilities that the epistemically transformative option A may turn out to have, whatever that subjective probability distribution happens to be.

Then such a synthetic lottery will have, for the agent, an expected utility that is equal to the agent's expected utility for option A.

But now suppose that, despite this equality in expected utilities, the agent nevertheless prefers the prospect of the synthetic lottery to the epistemically transformative option A.

If competing explanations of the pattern have been ruled out, then the remaining preference for the synthetic lottery over the prospect of the epistemically transformative experience may be taken, I think, as an indication that the agent is neophobic. And, of course, the opposite preference pattern, that is, a preference for the radically unfamilar option over the corresponding synthetic lottery constructed so as to have the same expected utility, would be an indication of neophilia.

My feeling is that there need be nothing at all irrational about either of these possibilities. We should have a normative theory of decision liberal enough to allow for cases of rational neophobia. And of course the same goes for the opposed phenomenon of rational neophilia.

It will be helpful here, I think, to compare what I'm calling neophobia with other patterns of preference that orthodox decision theory cannot accommodate and yet which seem perfectly rationally permissible. The first kind of example I have in mind involves an agent who is averse toward risk. Just as many have argued that a theory of decision making should allow for the rationality of various attitudes other than indifference toward risk,

so—it might be argued—such a theory should be just as permissive when it comes to attitudes other than indifference towards what is new.

So let's approach this task by first reviewing the problem that risk poses for standard accounts of decision theory.

> **Example 1: The Allais Problem.** The agent is to win a prize determined by drawing a ticket in a fair lottery with one hundred tickets. Consider the four options A_1 to A_4 displayed in the table below.

	0.01	0.10	0.89
	Ticket 1	Tickets 2-11	Tickets 12-100
A_1	\$1M	\$1M	\$1M
A_2	\$0	\$5M	\$1M
A_3	\$1M	\$1M	\$0
A_4	\$0	\$5M	\$0

Option A_1 guarantees the agent one million dollars no matter which ticket is drawn. Option A_2 is somewhat riskier: it yields five million dollars instead of a million if a ticket numbered 2 through 11 is drawn, but it also leaves the agent with a one percent chance of getting nothing at all. Faced with a choice between these first two options, many agents report a preference for A_1 over A_2. Now this might be taken as evidence that such an agent has a diminishing marginal utility for money: getting the first million dollars makes a lot more difference than getting the next four million dollars would. And, in fact, if the utility difference for the agent between the outcomes Win \$1M and Win \$0 is more than ten times the utility difference between Win \$5M and Win \$1M then a preference for A_1 over A_2 is exactly what expected utility theory prescribes.

The problem is, however, that many of those same agents—apparently perfectly rational people, I'm one of them—also report a preference for A_4 over A_3. That is, they prefer a ten percent chance of five million dollars to an eleven percent chance of one million.

But now agents like us have run foul of standard expected utility theory. For there are simply no utilities that may be assigned to the three outcomes \$0, \$1M, \$5M that can rationalize that pair of preferences as maximizing expected utility. The agent's preferences are in violation of Savage's Sure-Thing Principle, one of the axioms of the standard theory. If you cover over the third column of the table, the pattern of outcomes on what remains is the same for A_1 and A_2 as it is for A_3 and A_4, so the Sure-Thing Principle requires that an agent's preference for comparison for the first pair match that for the second.

What is going on here?

Many decision theorists, going back to Allais himself, have taken this example to be a reductio of any normative theory of choice which rules out as irrational the kind of aversion to risk that characterizes the preference

for A_1 over A_2 and for A_4 over A_3. If these preferences express a perfectly rational attitude toward risk, then standard expected utility theory will have to be liberalized in some way to yield a more reasonable set of norms.

But how might the standard theory be adjusted to accommodate the possibility of rational risk aversion? I will describe two possible answers to that question in what follows. The first of these is a particularly well-worked-out and elegant proposal due to Lara Buchak, developed and defended in her recent book *Risk and Rationality*. I'll approach Buchak's account via an example of the sort she uses to motivate the project in the first chapter of the book. (The version of the example I present here is due to Rachael Briggs.)

> **Example 2: The Pizza Problem.** Confronted with a choice between the following two options:
> (A) One pizza for sure.
> (B) A gamble that yields two pizzas if the toss of a fair coin lands heads and nothing if the coin lands tails.
> My friend and I share a preference for (A) over (B).

But now let's stipulate that the explanation of my preference for (A) over (B) differs from that of my friend's preference for (A) over (B). In particular, let's suppose that I prefer the certainty of one pizza to a toss-up between two pizzas and nothing, because one pizza is just about all that I can eat. I'm full after a single pizza, and as a result, the value I assign to getting a single pizza lies more than half way along the interval on my utility scale from no pizza to two pizzas. As a result of the fact that I have this kind of diminishing marginal utility for pizza, I prefer (A) to (B).

Things are quite different, on the other hand, in my friends's case. My friend, let's suppose, is insatiable. For him, the utility of the second pizza is undiminished by the fact that he has already eaten the first. So for my friend:

$$U(\text{two pizzas}) - U(\text{one pizza}) = U(\text{one pizza}) - U(\text{no pizza})$$

Yet my friend, like me, prefers (A) to (B). Why? Because he is risk averse. He simply does not want to take the chance of getting nothing.

There's another possibility here too, which I will only mention and then set aside. An agent with an insatiable appetite for pizza might prefer (A) to (B) out of *pessimism* rather than risk aversion. That is, the agent might judge that the probability of the fair coin landing heads is less than one half when his dinner depends on the outcome of the toss.

But let's set that further possibility aside. Let's suppose that we are satisfied that my friend assigns subjective probability $\frac{1}{2}$ to the coin's landing heads, whether or not his dinner depends on the outcome, and let's suppose further that we are satisfied that, for him, the utility of one pizza is exactly half way between the utilities he asssigns to two pizzas and that he assigns to nothing. Then by the lights of standard decision theory, my friend's

preference for (A) over (B) is irrational, since, for him, the expected utilities of (A) and (B) are equal to one another.

Yet—to many of us at least—this seems to be the wrong thing to say about my friend's preferences. To many of us it seems as though it is perfectly rationally permissible to be averse to risk taking in this kind of way. From this viewpoint, standard expected utility theory seems unduly harsh or over-restrictive for deeming such patterns of preference irrational.

But perhaps one ought to be suspicious of what I have stipulated above in setting out the details of this second example. By stipulating that we are satisfied, somehow, in a way that is independent of his preference for (A) over (B), that my friend's subjective probability for the coin landing heads is $\frac{1}{2}$ and that his utility gain from the second pizza is equal to the utility gain from the first, we might be thought to be committing ourselves to a non-constructive realism about utility and begging the question against the constructive realist.

Now, certainly, if there *are* further features of my friend's psychological state that we can point to and identify as those psychological features that ground the facts about his utility function stipulated in the second example, then that would demonstrate the inadequacy of any theory that left no room for that possibility. But suppose that there are no such further features to be found. Then a defender of the standard theory might simply reply that the apparent distinction stipulated in Example 2 between my friend's situation and mine is really a distinction without a difference. That is, the claim a defender of the standard theory might make is that this apparent distinction between my friend's situation and mine is precisely the consequence of that incorrect, non-constructive, conception of utility.

This still seems wrong to me. However if the defender of the standard theory adopts this strategy the mistake now seems to be not that a certain rationally permissible set of preferences is being ruled out incorrectly as irrational, but rather that the standard theory is leaving no room at all for a preference structure that in fact is perfectly possible.

Buchak develops and defends a theory of *risk-weighted expected utility* in which the choice-worthiness of an act is determined by three factors, not two. In this risk-weighted theory, the traditional roles of subjective probability and utility are augmented by a third factor, namely a *risk function*

$$r : [0, 1] \longrightarrow [0, 1]$$

that is non-decreasing and such that $r(0) = 0$ and $r(1) = 1$. The function r is intended to capture the facts about an agent's attitude to risk, and, crucially, does so in a way that can be elicited from a pattern of preferences that is coherent in an appropriate technical sense quite independently of the elicitation of probability and utility.

In order to see how this tripartite risk-sensitive scheme works it will help first of all to reformulate the standard account of expected utility in a kind

of stepwise fashion that proceeds from an initial monotonoic rank-ordering of outcomes from worst to best.

The most general case need not detain us here. The basic idea can be grasped by looking at a simple case where there are two possible states of the world s and t and two possibles outcomes x and y ordered so that the latter is at least as good as the former.

In that case the standard expression for the expected utility of an option $f = \{s,x\ ;t,y\}$, that is, of the act that delivers outcome x in state s and outcome y in state t is:

$$\mathrm{SEU}(f) = p(s).U(x) + p(t).U(y)$$

which, since x,y have been listed in order of increasing goodness, can be re-written in stepwise fashion as:

$$\mathrm{SEU}(f) = U(x) + p(t)(U(y) - U(x))$$

Now that we have this equivalent step-wise reformulation of standard expected utility, we can adjust it, via the risk function as follows to obtain Buchak's risk-weighted expected utility REU.

$$\mathrm{REU}(f) = U(x) + r(p(t)).(U(y) - U(x))$$

To get a sense of how this works, let's see how it might be applied to make sense of the distinction between my attitude and my friends's attitude toward pizza in Example 2 above.

Here the two relevant states of the world are H and T, the two possible results of the toss of the fair coin, and the outcomes, ranked for both of us in order from worst to best are no pizza, one pizza, two pizzas.

Then the previously mentioned distinction between my friend's risk aversion and insatiable desire for pizza, and my own risk neutrality and diminishing marginal utility for pizza can be captured by, for example, the assumption that my utility function for pizza is U_1 where

$$U_1(n) = \sqrt{(2n)}/2$$

where n is the number of pizzas received, while my friend's utility function is

$$U_2(n) = n$$

And, furthermore, my risk function r_1 is the identity function

$$r_1(x) = x$$

while my friend's risk function is

$$r_2(x) = x^2$$

Note that for both of us:

$$p(H) = p(T) = 1/2$$

since he and I agree that the coin is a fair one.

Plugging these utility and risk functions into the expression for risk-weighted expected utility we see that for me the value of the gamble g that delivers nothing on heads and two pizzas on tails is:

$$\mathrm{REU}(g) = 0 + r_1(p(T)).(U_1(\text{two pizzas}) - U_1(\text{no pizza}))$$

in other words:

$$\mathrm{REU}(g) = 0 + 1/2.(1 - 0) = 1/2$$

and this is less than the utility I assign to receiving a single pizza, that is,

$$U_1(1) = \sqrt{2}/2 \approx 0.707.$$

For my friend, on the other hand:

$$\mathrm{REU}(g) = 0 + r_2(p(T)).(U_2(\text{two pizzas}) - U_2(\text{no pizza}))$$

and so for him:

$$\mathrm{REU}(g) = 0 + 1/4.(2 - 0) = 1/2$$

which is less than the utility he assigns to getting a single pizza, that is, $U_2(1) = 1$.

This indeed has the required result that both of us prefer one pizza for sure to the gamble that gives us a 50% chance of two and a 50% chance of nothing. But that pattern of preferences has a quite different explanation in his case, where it is due to risk aversion, and in my case, where it stems from my diminishing marginal utility for pizza.

An appropiately chosen risk function can similarly rationalize the characteristic pattern of preferences in the Allais problem.

However, there is another kind of example that raises a similar challenge to standard decision theory, and which also cannot be accommodated in Buchak's system. The problem is due to Daniel Ellsberg and it turns out, I think, to be even more helpful to us than the first two examples in seeing how the possibility of rational neophobia might be treated formally (1961).

> **Example 3: The Ellsberg Problem.** An urn contains balls of three colors: red, black, and yellow. You know that it contains exactly thirty red balls and that there are an additional sixty balls which are either black or yellow, but in a ratio that is not known to you. You are asked to compare first the pair of options E_1 and E_2 the outcomes of which are determined by the color of a ball drawn at random from the urn, as specified in the table below.

	Red	Black	Yellow
E_1	\$100	\$0	\$0
E_2	\$0	\$100	\$0
E_3	\$100	\$0	\$100
E_4	\$0	\$100	\$100

Then you are asked to compare option E_3 to option E_4. As was the case in the Allais example above, many apparently perfectly rational agents express a preference for E_1 over E_2, and for E_4 over E_3, despite the fact that there is no standard expected utility representation of that pair of preferences. The situation is strikingly similar to the Allais case in that once again we have a violation of Savage's Sure-Thing Principle: cover over the third column of outcomes on "Yellow," and the pattern of outcomes on what remains is the same for E_1 and E_2 as it is for E_3 and E_4.

But there the similarities end. Strikingly, the risk-weighted utility theory of Buchak cannot accommodate the rationality of the Ellsberg preferences, although as we have seen her account can deal perfectly well with the Allais phenomenon. This difference arises because Buchak drops the Sure-Thing Principle in her axiomatization of preference; part of its work gets down by an axiom she calls Strong Comparative Probability, and it is the Strong Comparative Probability axiom that separates the Allais and the Ellsberg problems. The Allais preferences satisfy it; the Ellsberg preferences do not. (For details see Buchak 2013, 98–100, and Machina and Schmeidler 1992, 762–763.)

The moral of all this seems to be that the pattern of preferences commonly elicited by the Ellsberg example should be seen as an expression not of an aversion to risk, but rather of an aversion to what Ellsberg called *ambiguity*. It seems as though what leads to the choice of E_1 over E_2, and the choice of E_4 over E_3, is a preference for gambling on options where the outcomes have known objective probabilities, rather than options where the situation is "ambiguous" in the sense that the agent does not know what the objective probabilities are.

Now Isaac Levi is a prominent example of a decision theorist who has argued that the Ellsberg preferences should be regarded as perfectly rationally permissible, and that the way to accommodate them in a formal theory of decision is to allow that an agent's subjective probabilities, that is, her degrees of belief, may be *indeterminate* (1986).

In Levi's account, an indeterminate belief state is represented not by a single sharp subjective probability function, but by a convex set P of probability functions. (To say that the set is "convex" is to say that whenever p and q are probability functions in P, then every mixture $\alpha.p + (1-\alpha).q$, where $0 < \alpha < 1$, is also a probability function in P.)

There are various different ways in which such indeterminate probabilities might figure in a formal decision rule. Here we will follow Levi's suggestion that the agent first reduce the set of available options to those that are *E-admissible*.

> **Definition:** If an agent's utility function is u and her indeterminate belief state is represented by the convex set P of probability functions, then an option A is *E-admissible* for the agent if and only if there exists a probability function

> $p \in P$ such that A has maximal expected utility among all her options when those expected utilities are calculated using p and u.

In the Ellsberg example the agent's indeterminate belief state is represented by the set of all probability functions that assign probability 1/3 to Red, probability x to Black where $0 \leq x \leq 2/3$ (and a multiple of 1/60), and probability $2/3 - x$ to Yellow. With these indeterminate degrees of belief, both elements of the option set $\{E_1, E_2\}$ are E-admissible in Levi's sense. If p is chosen from P so that $x = p(\text{Black}) \leq 1/3$ then option E_1 has maximal expected value. For any other choice of p the option E_2 achieves the maximum. So either may be chosen. We can see similarly that both elements of the option set $\{E_3, E_4\}$ are E-admissible.

We could leave it at that, or we could follow Levi in allowing that some second-round rule of choice be applied to further winnow down the options that have survived the first-round test of E-admissibility. For example, if the agent adopts the rule of choosing the option from the E-admissible set that has the highest "security level," that is, the maximin expected utility over all $p \in P$, then the agent will indeed choose E_1 over E_2 and E_4 over E_3. The security levels for the four options E_1-E_4 in that order are $100/3, 0, 100/3$, and $200/3$ respectively (taking the utility of money for the agent to be given by function $u(\$n) = n$.)

Now Levi also maintains that an agent's utilities might also be indeterminate, and this allows him to give a similar account of the rational permissibility of the Allais preferences.

We allow, that is, that an agent's utilities for outcomes be given by a convex set U of determinate utility functions. Since there is already a "choice of scale" indeterminacy in measuring utility—we noted earlier the fact that utilities, like temperatures, will only ever be unique up to a choice of zero point and unit—let's assume that there is a pair of options x, y between which the agent is not determinately indifferent and that are ranked in the same order, y preferred to x say, by every utility function in the agent's set U. Then we may "normalize" the set U by choosing the scale for each of its elements u so that $u(x) = 0$ and $u(y) = 1$

The earlier definition of E-admissibility is then naturally extended to this system that allows indeterminacy in both probability and utility:

> **Definition:** If an agent's indeterminate belief state is represented by the convex set P of probability functions, and her indeterminate value state by a normalized convex set of utility functions, then an option A is *E-admissible* for the agent if and only if there exists some probability function $p \in P$ and some utility function $u \in U$ such that A has maximal expected utility among all her options when those expected utilities are calculated using p and u.

The application of this idea to the Allais problem is quite straightforward. The agent determinately ranks \$0 below \$1M, which is in turn ranked below \$5M. We may choose \$0 and \$1M as the outcomes with respect to which all the utility functions in the set U are normalized, by setting $u(\$0) = 0$ and $u(\$1M) = 1$ for all $u \in U$. Suppose that the agent's value state is then represented by a convex set U of utility functions such that for some $u \in U : u(\$5M) < 1.1$ and for some other $u' \in U : u'(\$5M) > 1.1$, Then for such an agent the characteristic Allais preferences will be rationally permitted, since each of A_1, A_2 will be E-admissible choices from the set $\{A_1, A_2\}$ and each of A_3, A_4 will be E-admissible choices from the set $\{A_3, A_4\}$. And an agent who adopts, for example, the second-round rule of choosing from among the E-admissible options the one whose second-worst outcome is best, will consider the characteristic Allais preferences to be the uniquely rational ones.

I think we should accommodate the possibility of rational neophobia in exactly the same way that Levi treats the Allais problem. That is, I think we should approach it as a phenomenon that can arise when an agent has indeterminate utilities for certain outcomes. Faced with a choice problem involving an epistemically transformative option, an agent can find herself with no determinate attitude toward the goodness of that outcome, with no determinate utility for it. The situation is not one which resolves itself into an uncertainty over which of some set of more fine-grained sub-outcomes is true. It's simply a matter of a basic and irresolvable indeterminacy. That's why the orthodox decision theorist's suggestion that we elicit her utility for the transformative outcome by the method of constructing a synthetic lottery need not always work. It's not possible to elicit a sharp determinate value for the utility of an outcome when it is just a fact that no such unique value exists. The synthetic lottery may yield some unique number, but so what? It's providing an answer to a different question.

If the agent simply has no determinate utility for an outcome X because she is phenomenologically unacquainted with outcomes of that type, then she may recognize that both the outcome X and its synthetic lottery "equivalent" are E-admissible options. And then she might rationally opt for the synthetic lottery over the unknown outcome because she adopts a second-round rule of preferring the familar to the unknown. This is a neophobic preference structure, and it should not be ruled out by a normative theory of choice as irrational. So we should admit indeterminacy in utility, and we should allow for the possibility of rational neophobia.

Indeterminacy of utility can arise in various ways. One variety in which Levi has been particularly interested throughout his career is the kind of indeterminacy that stems from a conflict in values. An agent may recognize that two different and perhaps competing features of an outcome are relevant to establishing its utility. The agent may know that the utility of the outcome is to be figured as a tradeoff between these competing criteria—as some weighted mixture of the simple determinate utilities that would be

arrived at if only one or the other of the two factors were relevant. And yet the agent may be forced to admit that there is no fact of matter as to how the weighting of that mixture should get done. In such a situation the agent will assign no determinate utility to the outcome. The best she may be able to do is to assign it some interval of real-number values parametrized by the possible values of the weighting factor.

In some of the most fascinating, and elusive, passages of Chapter 2 of her book, L. A. Paul seems to be pushing just this kind of point. I have in mind those passages in which, for example, she stresses the richness and multi-dimensionality of the notion of value. To take that kind of criticism seriously might seem to be to reject the standard decision theoretic framework in a rather drastic and fundamental way. It might seem to require rejecting the very idea that rationality of choice could depend simply on facts about expected utility. I've previously resisted that idea strongly and argued it at length with L. A. Paul. But it now seems to me that the required revision to the standard theory need not be so drastic, and that the means for handling her cases of epistemically transformative choice are already well known from the work of Isaac Levi and others and might already be required to handle other well-known problems. That's how I now read those fascinating and elusive passages of the second chapter of Paul's book. I've come to see her discussion of epistemically transformative choice problems as identifying a new and very important role for the theory of indeterminate utility. It's one more reason to be grateful to Paul for having written such a rich and interesting book.

John Collins
E-mail : john.collins@columbia.edu

References:

Buchak, Lara. 2013. *Risk and Rationality*. Oxford: Oxford University Press.

Dreier, James. 1996. "Rational Preference: Decision Theory as a Theory of Practical Rationality." *Theory and Decision* 40: 249–276. http://dx.doi.org/10.1007/BF00134210.

Ellsberg, Daniel. 1961. "Risk, Ambiguity, and the Savage Axioms." *Quarterly Journal of Economics* 75 (4): 643–669. http://dx.doi.org/10.2307/1884324.

Levi, Isaac. 1986. "The Paradoxes of Allais and Ellberg." *Economics and Philosophy* 2: 23–53. http://dx.doi.org/10.1017/S026626710000078X.

Machina, Mark and David Schmeidler. 1992. "A More Robust Definition of Subjective Probability." *Econometrica* 60 (4): 745–780. http://dx.doi.org/10.2307/2951565.

Paul, L. A. 2015. "What You Can't Expect When You're Expecting." *Res Philosophica* 92 (2): 149–170. http://dx.doi.org/10.11612/resphil.2015.92.2.1.

Acknowledgements An earlier version of this paper was presented to the *Res Philosophica* Conference on Transformative Experience held at Saint Louis University on September 19-20, 2014. Thanks to Jon Jacobs and the other organizers of that conference, and to all the participants. I owe a special debt both to Laurie Paul for many helpful discussions of these matters, and to Sophie Horowitz, whose acute comments on the presented version of this paper led me to revise it extensively. Thanks also to two anonymous referees for helpful suggestions.

Pettit, Philip. 1991. "Decision Theory and Folk Psychology." In *Foundations of Decision Theory: Issues and Advances*, edited by Michael Bacarach and Susan Hurley, 147–175. Oxford: Blackwell.