CAMBRIDGE JOURNALS

Proof Delivery Form

Episteme

Date of delivery: 04-06-2018

Journal and vol/article ref: epi epi1800028

Number of pages (not including this page): 20

This proof is sent to you on behalf of Cambridge University Press. Please print out the file and check the proofs carefully. Please ensure you answer all queries.

Please EMAIL your corrections within 3 days of receipt to:

Lesley Bennun
lesley.bennun@btinternet.com

Please clearly indicate any corrections required by page, column and line reference.

Authors are strongly advised to read these proofs thoroughly because any errors missed may appear in the final published paper. This will be your ONLY chance to correct your proof. Once published, either online or in print, no further changes can be made.

NOTE: If you have no corrections to make, please also email to authorise publication.

- The proof is sent to you for correction of typographical errors only. Revision of the substance of the text is not permitted, unless discussed with the editor of the journal. Only **one** set of corrections are permitted.
- Please answer carefully any author queries.
- Corrections which do NOT follow journal style will not be accepted.
- A new copy of a figure must be provided if correction of anything other than a typographical error introduced by the typesetter is required.

- If you have problems with the file please email **epiproduction@cambridge.org**

Please note that this pdf is for proof checking purposes only. It should not be distributed to third parties and may not represent the final published version.

Important: you must return any forms included with your proof. We cannot publish your article if you have not returned your signed copyright form

Please do not reply to this email

NOTE - for further information about **Journals Production** please consult our **FAQs** at http://journals.cambridge.org/production_faqs

Author Queries

Journal: Episteme

Manuscript: S174236001800028Xjra

Q1 The distinction between surnames can be ambiguous, therefore to ensure accurate tagging for indexing purposes online (eg for PubMed entries), please check that the highlighted surnames have been correctly identified, that all names are in the correct order and spelt correctly.

Q2 If the second author would like their email address to be included, please provide it.

Q3 Kahneman (2011) and Kahneman and Tversky references are missing from the list.

Q4 Gigerenzer 2008 is not cited in the text.

Typesetter Query:

1. The following author, year is listed in References, whereas it is not cited in text. Please cite or delete: Buchak (2017).

Episteme (2018) page 1 of 20 © Cambridge University Press
doi:10.1017/epi.2018.28

REAL WORLD PROBLEMS

Q1 L. A. PAUL AND JOHN QUIGGIN
Q2 lapaul@unc.edu

ABSTRACT

In the real world, there can be constraints on rational decision-making: there can be limitations on what I can know and on what you can know. There can also be constraints on my ability to deliberate or on your ability to deliberate. It is useful to know what the norms of rational deliberation should be in ideal contexts, for fully informed agents, in an ideal world. But it is also useful to know what the norms of rational deliberation should be in the actual world, in non-ideal contexts, for imperfectly informed agents, especially for big, life-changing decisions. That is, we want to know how to deliberate as best we can, given the real-world limitations on what we can know, and given real-world limitations on how we are able to deliberate. In this paper, our concern is with the norms of rational deliberation in certain, important, non-ideal contexts, where the reasoning occurs from the agent's first person, subjective point of view. The norms governing the process of deliberation for real people in the sorts of non-ideal contexts we'll consider need to reflect the way that real agents, with an incomplete grasp on the facts and an imperfect ability to deliberate, can be expected to proceed. Our central contention is that framing the deliberative process from the first person perspective allows us to uncover and explore important, real-world constraints on boundedly rational agents deliberating from the subjective perspective.

INTRODUCTION

When undertaking a big decision, I want to make the best possible decision. To do that, I want to deliberate in the best possible way. I want to deliberate as sensibly and effectively as possible, taking proper account of my needs and preferences.

When advising me, you may want to advise me in the best possible way. To give me the best possible advice, you need to deliberate as sensibly and effectively as possible, taking proper account of my needs and preferences. In both situations, then, it is important to identify the proper norms for the type of deliberation in question.

The ideal of deliberation is rational deliberation, reasoning from the available information to the best available conclusion. In ideal situations, we may expect ideal deliberation.

In the real world, however, the situation is rarely ideal. There may be constraints on rational decision-making: limitations on what I can know and limitations on what you can know. There may also be constraints on my ability to deliberate or on your ability to deliberate. It is useful to know what the norms of rational deliberation should be in ideal contexts, for fully informed agents, in an ideal world. But, it is at least as useful to know what the norms of rational deliberation should be in the actual world, in

non-ideal contexts, for imperfectly informed agents making their own decision, especially when those agents are making big, life-changing decisions. That is, we want to know how to deliberate as best we can, given the real-world limitations on what we can know, and given real-world limitations on how we are able to deliberate. In this paper, we focus on this issue: we are concerned with non-ideal agents attempting to rationally deliberate, from their subjective perspective, in certain distinctive kinds of important, non-ideal contexts.

Our project, then, concerns *bounded rationality*. However, it's a distinctive kind of bounded rationality: we are interested in how to assess decision-making from an imperfectly formed first-person perspective. Ordinarily, explorations of the norms of decision-making under bounded rationality assume that the observed patterns of a boundedly rational choice should be assessed from a perfectly informed third-person perspective. That is, ordinarily, a standard analysis of the norms of decision-making under bounded rationality considers the problem from an idealized, quasi-observational, unbounded perspective that we can describe as "objective." In contrast, we'll consider the problem from a non-idealized, first-personal, bounded perspective that we'll describe as "subjective."

On a standard analysis taken from the objective perspective, the focus of the analysis is on the errors made by boundedly rational agents, and on the way in which those agents' decisions might be improved to meet the standard of the perfect, i.e., unbounded, objective, rational deliberator. The objective deliberator has access to all the facts about the state of the world and the capacity to reason to the optimal decision.

For example, a prominent and influential approach is the 'Nudge' analysis of Thaler and Sunstein (2008), which focuses primarily on the case of boundedly rational agents. Despite having access to the information required for an optimal decision, in the sense prescribed by objectivism, such agents often make mistakes. The 'nudge' idea is that a policy-maker (implicitly assumed to be unboundedly objective) can help them by framing choice problems in such a way as to make the optimal choice more salient, while leaving the agent with the freedom to make mistakes if they wish.

Another example comes from the work of Kahneman and Tversky, who present evidence that experimental subjects and real-world decision-makers frequently make suboptimal inferences and choices, even when they have all the information required for an optimal choice. A typical example is the 'Linda' problem, in which subjects, given a description of a young woman, assign a higher probability to the conjunction "Linda is a bank teller active in the feminist movement" than to the single premise "Linda is a bank teller." Kahneman (2011) suggests possible correctives. Again, the focus is on the way in which those agents' decisions might be improved to meet the standard of the perfect, i.e., unboundedly rational, objective deliberator."

We also focus on the boundedly rational agent. However, rather than analyzing the way in which those agents' decisions might be improved to meet the standard of the unbounded objective deliberator, we argue that, for a certain class of cases, such an approach is not relevant. (For these cases, we think the standard for the objective deliberator cannot be met in any realistic sense.)

So in contrast to approaches to bounded rationality that address the problem from the objective perspective, we will consider decision-making under bounded rationality *from a non-ideal, or "subjective," perspective*. Our analysis of the norms of decision-making under bounded rationality will consider the problem from the subjectively limited, first-personal perspective, in order to explore how observed patterns of a boundedly

rational choice should be assessed by a non-ideal agent who, for principled reasons, cannot have a complete grasp on all the facts from her subjective point of view. (The philosopher Bernard Williams (1981), among others, has explored decisions taken based on an agent's reasons, but without the same first-person focus that we want to take.)

We think a normative approach to the problem of decision-making from the subjective point of view in non-ideal contexts is well worth exploring, because it captures a certain element of real-world decision-making. To determine the best way to deliberate in these real-world situations for choices made from the first-person perspective, we need to know what norms, in principle, real agents can be expected to follow. (Our normative account could also be used as a frame for more descriptive, empirically based approaches to the problem of decision-making from the subjective point of view.) In pursuit of our normative goal, we will focus on identifying and exploring a set of in-principle problems for agents who are imperfectly informed, imperfect deliberators attempting to make high-stakes, big decisions from their subjective point of view. We'll ask: What are the challenges for rational deliberation, when such deliberation occurs from the subjective perspective, for real agents considering real-world problems? How might we respond to these challenges? Questions of this kind lie at the intersection of philosophy, economics, and psychology.

The paper is organized as follows. Section 1 introduces the rational decision project. Section 2 presents our decision-theoretic approach to choice under uncertainty. Section 3 defines our central concepts, including bounded rationality, objectivism, and subjectivism, and identifies our target: the boundedly subjective agent. Section 4 describes, in detail, the concept of rational security, and shows how it fails as a result of bounded awareness and transformative experience. Section 5 describes the possibility of an alternative approach, described as 'choosing reasonably,' and 'choosing with confidence,' and proposes the notion of 'confidence' as an alternative to the notion of 'security.' Section 6 returns to the motivating example with which we began the paper and discusses the possibility of making reasonable choices when the requirements of the rational choice project are not satisfied. Some concluding comments are offered.

1. INTRODUCTION TO RATIONAL DELIBERATION

We'll start by introducing and framing the dominant approach to rational deliberation, which involves what can be described as 'rational choice.'

The central issues addressed are

(A) How should we characterize rational choice and its requirements?

(B) What decision procedures will guarantee consistency with the requirements of rational choice?

We can describe the project of providing an answer to these questions as the 'rational decision project.'

An answer to question (A) typically takes the form of a set of axioms and a 'representation theorem,' that is, a demonstration that choices will satisfy a given set of rationality axioms if and only if they may be represented by the maximization of a suitable function.

An answer to question (B) consists of a set of explicitly specified decision rules for rational decision-making that provide a purported basis for what we will call 'security.' Briefly, security is an assurance that the agent can make decisions that are the best possible in those circumstances, conditional on their preferences and the information available to them. Rational security for an agent making a choice can be verified by a subject with full access to the revealed preferences and beliefs of the agent. As such, (B) is implicitly designed to capture the norms of *ideal* deliberation.

A principled and systematic failure to satisfy the conditions of rational choice set out as in (A) is usually described as constituting 'bounded rationality.' We motivate our view by showing how, in addition, in contexts of subjective decision-making, the standard answer given to (B) also fails: in non-ideal contexts, the security promised by the rational decision project is illusory. We will describe three contexts in which security can fail, the contexts of *bounded awareness*, *transformative experience*, and *causal failure*.

These three potential failures of security reflect three problems with the rational decision project. The first problem is with listing all the possible consequences. The second problem concerns assigning utilities to consequences. The third problem concerns the assumption of constancy in the chooser with respect to the optimal course of agency. In this context, we'll argue that there are important real-life cases where the agent is unable to be aware of all propositions about acts available to the agent and states that may arise conditional on those acts, where she cannot understand the consequences (conjunctions of acts and states) that differ with respect to preferability, or where she lacks the capacity to predict her own preferences over those consequences, should they be realized.

2. ACTS, STATES, AND CONSEQUENCES

Choices may be represented in many different ways. More formally, decision theory commonly involves acts, states and consequences that may be represented in different, but fundamentally equivalent ways.

The usual convention in economic decision theory is to distinguish between states of the world and consequences, and to represent acts as mappings from a set of states to a set of consequences. Probabilities and utilities are derived from preferences by way of a representation theorem, as in Savage (1954), who draws on the earlier work of Ramsey (1926).

In philosophy, one common approach follows the epistemic decision theory of Jeffrey (1990). We will formulate our position in these terms. Jeffrey (1990) treats both acts and states as propositions about the world, and consequences as conjunctions of act and state propositions. Also following Jeffrey, we will take propositions to be linguistic, truth evaluable entities. Facts are true propositions.[1] Credences and utilities are derived from evaluations of the probability and desirability of these propositions. As Jeffrey (1990: 59) observes, this is '*unified* in the sense that it attributes probabilities and desirabilities to the same objects' (emphasis in original). By contrast, a central feature of Savage's approach is a separation between probability (beliefs about states of the world) and desirability (preferences about consequences).

1 We adopt this definition of "fact" simply for expedience. In other work, Paul, takes facts and propositions to be non-linguistic entities, as opposed to linguistic entities.

Adopting the exposition in Pettigrew (2015), we have:
The primitives are:

* A set of *acts* A, where each $A \in \mathsf{A}$ is a proposition that describes a possible act that an agent might perform and stated that she does in fact perform that act.
* A set of *states* S where each $S \in \mathsf{S}$ is a proposition that describes a possible state of the world; they are mutually exclusive and exhaustive.
* Conjunctions of the form $A\text{^}S$ are called outcomes or *consequences* relative to A and S.

Agents are normally assumed to have complete preferences over acts denoted by an ordering $\succcurlyeq$. Preferences may be represented in terms of probabilities (or credences) and desires (utilities or valuations).

Beliefs are represented by probabilities or credences $P(S|A)$ where $P(S|A)$ is the agent' s credence that the world is in state S under the subjunctive supposition that she performs act A .

In this model, utilities are given by a real-valued function $u(A\text{^}S)$ representing the utility obtained from the consequence $A\text{^}S$ arising from state S, conditional on having performed action a. Utilities represent preferences over consequences in the sense that $u(A\text{^}S) > u(A'\text{^}S')$ if and only if $A\text{^}S$ is preferred to $A'\text{^}S'$.[2]

We may define the set of utilities and credences potentially arising from action A as $\{u(A\text{^}S), P(S|A){:}S \in \mathsf{S}\}$, and represent an evaluation of those utilities and credences by a valuation function V. The canonical case is the expected utility function

$$V\{u(A\text{^}S), P(S|A) : S \in \mathsf{S}\} = \Sigma_{S\in\mathsf{S}}\, u(A\text{^}S)P(S|A)$$

We will define events as sets of states and observe that conditional on the occurrence of a given event, an act gives rise to a probability distribution over consequences, which may be computed using Bayes' theorem.

A variety of alternatives to, and generalizations of, expected utility theory, consistent with this general framework, have been proposed. Among the most widely used are the rank-dependent family (Quiggin 1982; Buchak 2013) and models of choice with ambiguous probabilities, inspired by Ellsberg (1961). The arguments in this paper are equally applicable to expected utility theory and to more general theories of rational choice, such as rank-dependent and ambiguity models.

3. DEFINING TERMS

Rational choice models are standardly formulated from metaphorically speaking, an "observational" or quasi-scientific point of view. We will not adopt this standard

2 We do not in fact take utility values to be real numbers, nor do we think the issue is easily resolved. The utility of one consequence is meaningful only relative to others. The standard expected utility model does not assist us here, since the utility values it generates are unique only up to an affine transformation. In particular, this makes comparisons between agents, or between present and future selves, problematic, an issue that has concerned economists ever since Robbins (1938).

formulation. To clarify our preferred formulation of the problems, we'll need to define some terms:

(i) The *unboundedly rational agent* knows all the possible states, and all the possible actions, and can evaluate all the possible consequences. For this reason, the unbounded rational agent is the best possible reasoner. She is a "perfect deliberator."
(ii) The *boundedly rational agent* may not be aware of all the possible states, or of the possible actions, or able to evaluate all the possible consequences.[3] For this reason, the bounded rational agent is not the best possible reasoner. She is an "imperfect deliberator."

We will also want definitions of 'objective' and 'subjective.' Loosely building on the definitions of objectivism and subjectivism given by Kolodny and MacFarlane (2010), we can start with:

*Objectivism**: X ought to *A* iff performing A maximizes X's objective utility, that is, if performing A maximizes X's utility with regard to all the facts (and truths).
and
*Subjectivism**: X ought (at t) to *A* iff performing A maximizes X's subjective utility, that is, if performing A maximizes X's utility with regard to all the facts (and truths) that X knows and grasps (at t).[4]

These definitions capture a basic, intuitive distinction, but we'll need to modify them to fit with our decision theoretic approach. Take a complete proposition to fully describe a possible state. If X knows and fully grasps all the facts, X knows and grasps the complete proposition that describes the actual state, that is, X knows and grasps the complete true proposition. With this in hand, we can define our preferred versions of 'objectivism' and 'subjectivism':

Objectivism: X objectively ought to *A* iff performing A maximizes X's utility with regard to all the facts (and truths) given by the complete proposition that describes the actual state, that is, given the actual state of the world. Here, we are concerned with the best act *A^S* that is available, given the actual state of the world.
Subjectivism: X subjectively ought (at t) to *A* iff performing A maximizes X's utility with regard to all the facts (and truths) given by the propositions that X knows and grasps (at t). Here, we are concerned with the best act *A^S* that is available, given the possible states of the world consistent with what X knows and grasps (at t).

An agent is *objective* (or has an objective perspective) if she knows and grasps the complete proposition that describes the actual state of the world. An agent is *subjective* (or has

3 We are using the inclusive "or" here.
4 Kolodny and MacFarlane (2010) don't refer to utility. They propose the following definitions:
Objectivism: *S* ought to φ iff φing is the best choice available to *S* in light of all the facts, known and unknown.
Subjectivism: *S* ought to φ iff φing is the best choice available to *S* in light of what *S* knows at *t*.

a subjective perspective) if she is unable to know and grasp the complete proposition that describes the actual state of the world.

Again, agents are normally assumed to have complete preferences over acts denoted by an ordering $\succcurlyeq$. Preferences may be represented in terms of probabilities (or credences) and desires (utilities or valuations).

Now that we've defined our terms, we can describe four kinds of agent deliberation.

(a) When X is an *unboundedly rational objective agent*, X is a perfectly informed perfect deliberator.
(b) When X is a *boundedly rational objective agent*, X is a perfectly informed imperfect deliberator.
(c) When X is an *unboundedly rational subjective agent*, X is an imperfectly informed perfect deliberator.
(d) When X is a *boundedly rational subjective agent*, X is an imperfectly informed imperfect deliberator.

Our focus is on (d), boundedly rational subjective agents. As we have observed, analysis of boundedly rational agents commonly analyze the situation and develop decision norms from the perspective of an unboundedly rational objective observer, with access to all the facts about the state of the world and the capacity to reason to the optimal decision. In contrast, our proposed decision norms will assume the choices are made from the perspective of a boundedly rational subjective observer, one who lacks access to all the facts and the capacity to reason to the optimal decision. We'll show, by exploring a certain class of extremely important, real-world choices, that the normative standard derived from the perspective of the unboundedly rational subjective deliberator cannot be met by the individual chooser in any realistic sense. As a result, current treatments of boundedly rational agents need to be supplemented with an account of the way non-ideal agents should attempt to rationally deliberate in our distinctive kinds of non-ideal contexts.

4. RATIONAL SECURITY AND HOW IT FAILS

We'll begin our argument by defining a notion we term "rational security" and showing how it fails in non-ideal, real-world contexts involving bounded awareness or epistemic and personal transformation.

The central (implicit) claim of rational decision theory is that, properly applied, the expected utility procedure (or some variant) yields *rational security*: provided the explicitly specified decision rules defined by the maximization of expected utility are followed, an agent can make decisions that will select her most preferred action from any possible choice set.

The concept of rational security is applicable in the case of an unboundedly rational subjective agent. Recall that the rational decision model begins with the assumption that agents possess, and have the capacity to solve, a complete state-act-consequence model for the decision under consideration. This requires that the agent has:

(1) Awareness of all propositions about acts available to the agent and states that may arise conditional on those acts.

(2) Understanding of the consequences (conjunctions of acts and states) that differ with respect to preferability.
(3) The capacity to predict the agent's own preferences over those consequences, should they be realized.

If these conditions are met and the agent's preferences satisfy the consistency requirements of (possibly generalized) expected utility, the decision model provides principles of choice that give an agent rational security. Rational security thus implies the following *principles of choice*:

(i) Any two rational agents with the same preferences and credences (or degrees of belief) would make the same choice (or be indifferent between choices).
(ii) Any rational agent, informed of the preferences and information for another agent, would suggest the same choice as optimal.
(iii) *Ex post*, after the outcome was observed, all rational agents would agree that the choice made *ex ante* was the correct one, conditional on the information and known preferences at that time.
(iv) For a finite problem, it is possible to specify an algorithm, implementable by an autonomous agent, that will generate optimal choices for any decision problem given a specification of the chooser's preferences and available information.

As these features indicate, rational security requires a full specification of feasible acts, states of the world, and agent preferences over consequences at the time of choice.

This model fails for important types of real-world decision-making involving bounded awareness and epistemic and personal transformation. The problems of bounded awareness and transformative experience, discussed in the examples below, show how rational security cannot be preserved when making decisions from an ordinary, subjectively bounded perspective.

In particular, rational security fails in decision contexts where the agent is *unaware*: contexts in which the agent fails to meet condition (1), for she lacks awareness of all propositions about available acts and states of the world that jointly determine the consequences of those acts.

Rational security also fails in decision contexts where the agent has a *transformative experience*: contexts in which the agent fails to meet conditions (1–3), for she lacks the capacity to predict her own preferences (or assign utilities) over the consequences that could be realized. (She fails to meet conditions (1) and (2), depending on what one requires "knowledge" to involve: on our view, she can fail to know certain propositions because testimony is not available, or even when testimony is available, she can fail to know the *de se* truths these propositions concern, that is, she does not grasp the propositions in the right way, the way she needs to in order to decide and act. See Paul (2018) for discussion.)

In the remainder of this section, we will show, more formally, how security fails in the presence of bounded awareness and transformative experience.

4.1 Bounded awareness

Until recently, problems of bounded awareness received little attention in the decision-theoretic literature. Early work such as that of Fagin and Halpern (1987) and Modica and Rustichini (1994) illustrates some of the difficulties that needed to be addressed.

As we discussed above, the rational decision project starts with the assumption that the agent is aware of all acts and states and can evaluate the consequences arising from the conjunction of given acts and states.

Thinking about my decision to move (or not) to the city shows that this framework is problematic. The question of whether or not I will find work will obviously occur to me. But, given the understanding of the world that I have acquired in my small community, I can't be aware of the variety of jobs that might, or might not, be available to me.

In some cases, this is a matter of lacking fine-grained distinctions between possible jobs. I might, for example, anticipate the possibility of finding work as a laborer, an occupation familiar from village life, but not have much idea of the kinds of work done by urban laborers.

On the other hand, there may be possibilities of which I am completely unaware. For example, I might end up as a dog-walker in New York City. But the thought that this is a way of making a living is a possibility that would never have occurred to me. Where I grew up, dogs ran free or were kept in enclosed yards. The idea that a rich person would pay someone to walk their dog would seem ridiculous to anyone from my community.

In formal terms, it's useful to distinguish between coarse awareness and restricted awareness. Grant and Quiggin (2013*a*) develop this distinction in the context of an extensive form model of inductive reasoning about unawareness.

With *coarse awareness*, consequences and states of the world are partitioned in a coarse grained way. The example of laboring jobs involves coarse awareness about consequences.

Problems of coarse awareness also affect agents who fail to represent relevant distinctions between different states of the world. For example, I may not be aware that the chances of getting a job differ according to the time of year, or according to the state of the economy. More fundamentally, the concept of 'the state of the economy,' familiar to nearly everyone in an industrial economy, may be unavailable to me, as someone coming from an agrarian society where productive activity depends mainly on the state of the weather. If I lack the concept, I can't have knowledge of this state of the world.

An agent who fails to draw a distinction between two different states of the world can represent an act correctly only if the consequences of that act are the same in each state. The problem is not limited to awareness of consequences. I may be unaware, or incompletely aware, of some possible states of the world.

If I stay in my small community, and decide on what crop I should plant to provide for my subsistence needs, the state of the urban economy will not affect the outcome of my decision. But if the consequences of an action differ between 'boom' and 'bust' states, and I fail to distinguish between these possibilities, then, whatever consequence I assign to the action, I must be mistaken in at least one of these states.

It's not clear how this problem should be handled. One interpretation is that the agent implicitly assumes that the consequence associated with one state, say the 'boom' state, will be realized, and disregards the other possible consequence. The effect of this implicit assumption is that the other 'bust' state is disregarded.

This leads us to models of *restricted awareness*, in which the agent is completely unaware of some states of the world, and of consequences that arise only in those states. We've already considered the possibility of unforeseen consequences such as becoming a dogwalker. The case of unconsidered states is also important.

We may suppose that the agent correctly perceives the consequences of acts in states of which she is aware. However, the restricted nature of the state space means that the

probability distribution over consequences associated with a given act is misrepresented. As an example, the villager may be unaware of potentially fatal diseases prevalent in the city. This will lead to an underestimation of the probability of the consequence 'Death' associated with a move.

We may also consider bounded awareness in terms of the subjective and objective perspectives we've discussed, using vision as a metaphor. In the objective perspective of rational decision theory, the birds-eye view is implicitly assumed to have available arbitrarily fine resolution and an arbitrarily broad field of view. By contrast, the subjective perspective involves the limitations of human vision.

4.2 *Transformative experience*

Cases of transformative experience raise several problems for the rational decision project. An especially thorny issue concerns the assumption of constancy in the chooser with respect to the optimal course of agency.

As we have already noted, in the standard model, agents are assumed to be able to make consistent probability judgements of the form

$$\Pr\{c; a\} = p,$$

which may be interpreted as 'If I choose act a, then I will experience consequence c with probability p'.

Once the consequences have been listed, to fill out the model, the agent must assign utilities to consequences, with the more preferred consequence receiving the higher utility. Here, we interpret 'more preferred,' where A is more preferred than B, in terms of a psychologically real preference for A over B. In this situation, we understand the psychological preference to involve a desire or another (psychologically real) conscious state of the agent. Alternatively, we may interpret 'more preferred' as 'judged to be more valuable,' where the mental state of the chooser involves a representation and assignment of value or utility.

Decision contexts where the agent has a transformative experience are contexts in which the agent fails to meet conditions (1–3) for rational security, for she lacks the capacity to represent and predict her own preferences (or assign utilities) over the consequences that could be realized. Moreover, since the science is incomplete, there is no further (suitably reliable) source of information. That is, by assumption, there is no omniscient observer or sufficiently expert scientist to tell her what her preferences should be in this situation. She cannot substitute testimony for her ignorance, for she has no such testimony available to her. Further, as we'll discuss below, the consequences are transformative with respect to her preferences, creating an *ex ante/ex post* conflict.

As a result, she cannot assess the desirability of all of the possible consequences rationally. In rational choice theory, the act that yields the most preferred probability distribution over consequences is the one that should be chosen; that is, the agent is to follow the expected utility rule: Determine the probability of each state and attach a utility number to each consequence, then choose the act that maximizes expected utility. However, the agent *cannot* follow the expected utility rule if her preferences (or utilities) for possible consequences are undefined at the time of the choice.

The immediate consequence is that the three conditions of the rational decision project cannot be met, and all four of our principles of choice (i–iv) fail.

Recall that Principle (i) stated that any two rational agents with the same preferences and credences (or degrees of belief) would make the same choice (or be indifferent between choices). However, because at the time of decision, preferences (utilities) are undefined, there is no meaningful sense in which the preferences and credences (or degrees of belief) of two rational agents can be compared.

Principles (ii–iv) fail for the same reason: there is no basis for a rational agent to defend a transformative choice as optimal, and no meaningful way to compare preferences *ex ante* and *ex post*.

There are other interesting problems that arise as the result of the agent's failure to represent defined preferences and utilities with respect to the possible consequences. In our approach, the agent's preferences (utilities) are undefined when she deliberates and chooses. However, if she chooses to undergo an experience that transforms her both epistemically and personally, she may form defined preferences (utilities) *ex post* in response to the experience.

In this situation, Principles (iii–iv) fail for an interesting reason: the agent's preferences *ex ante* are inconsistent or incommensurable with her preferences *ex post*. We find ourselves with a version of a Kuhnian conceptual revolution from the individual's perspective: from the agent's subjective point of view, she's undergone a preference revolution. If her preferences are incommensurable across her person stages (the temporal stages with preferences and choice behavior that make up a persisting person over time), there may be no way to define a consistent algorithm extending *ex ante* to *ex post* that can be implemented by an autonomous, model-utility based intelligent agent.

Consider the case when I'm a villager moving to the city. When I'm deliberating about my move to the city, I'm uncertain as to what kind of job I might find, if any. The rational choice model tells me to estimate the utility of being employed in different kinds of jobs and of being unemployed.

Immediately, however, we run up against a problem. Because I've never left my tiny village before (I grew up here), I know nothing of what it's like to live elsewhere. How am I, *ex ante*, supposed to determine the utilities of these different possible outcomes? Until I arrive on the scene and take up the job in question, I lack any deeper understanding of the nature and character of the life I'm about to undertake. The lack of deeper understanding stems from my conceptual and imaginative impoverishment, due to my lack of experience. In an essential sense, I lack the ability to assign value to the different consequences of these different possible jobs in different possible cities. I lack this understanding even if I have access to descriptions, from friends of mine who have already moved, of all the different jobs I could get and all the ways this could change my life.

The problem arises because taking up a new job in a new city, for someone with my lack of experience and exposure to the larger world, would be both epistemically and personally transformative. Following Jackson (1982), Lewis (1986), and Paul (2015), perceptual states that involve the valuing and representation of new experiential kinds for the decision-maker are, in an important sense, epistemically inaccessible from the *ex ante* perspective.

An illustrative example can help. Consider the way the nature of perceptual states concerning what it is like to see are epistemically inaccessible to a congenitally blind chooser. Such a blind chooser, at the subjective *ex ante* position, cannot first-personally represent

and evaluate consequences concerning what it would be like for him to see colors or other visual properties when deciding whether to have a type of retinal surgery that, while painful, could endow him with a limited capacity for ordinary vision.[5]

On this picture, background knowledge of the kind of experience involved is necessary for the agent to have the ability to imagine, represent, and thus assign value to the possible consequences. Put in slightly more formal terms, under the standard model, when I am deliberating over whether to move to some particular city, and if I move, whether to take up some particular job, I am reflecting about and comparing different courses of action, with different consequences.

This representation of the decision situation describes the decision in terms of a state-act-consequence framework. I am uncertain about which branch of a decision tree to follow, in virtue of my uncertainty about which outcome is best. This uncertainty models, in a very rough and intuitive sense, my psychological reflections as I decide. But the reality is that I, as an epistemically impoverished chooser, cannot specify, *ex ante*, an essential element of this model. Without the right background experience, I cannot represent the values and assign the utilities to the consequences in question.

The problem is normative: it isn't that the conceptual resources required are too complex, or that humans are just bad at forecasting how they'll respond to various situations. The problem is that, given how human brains work, humans require experience of the relevant kinds in order to have the epistemic capacity to represent and value possible consequences involving experiences of that kind. Description and testimony lack the requisite expressive power. Even in cases where the individual does have reliable testimony about the consequences in question (a speculative, and much stronger assumption than we are making here), he may not be able to represent the crucial *de se* truths involving those facts. (He may be told how he is likely to respond to the consequences, e.g., with pain of such and such intensity, or with joy or confusion, but still be unable to represent the nature of these experiences in himself in order to form and represent the needed utilities.)

Once the chooser has the new experience, she is epistemically transformed. The experience gives her the epistemic capacity to imaginatively represent the nature of future or possible new experiences of that kind. For example, once the blind adult gains the ability to see, and sees color for the first time, his conceptual and imaginative resources are enriched in ways that allow him to assign value to what it is like for him to see, and in particular, what it is like to experience ordinary sight and the life consequences that flow from this change in the nature of his lived experience.[6]

Epistemic impoverishment leads to a second problem. Having the experience of moving to a city and taking up a new job isn't merely epistemically transformative: the epistemic change can be so dramatic that it scales up or otherwise causes a change in what I care about, that is, what I prefer. If the experience of moving to the city is life-changing in a way that changes my core personal preferences, I'll be personally transformed. (This possibility is also reflected in our example of the blind chooser.) We'll define an experience that both epistemically and personally transforms a chooser as a "transformative experience."

5 See Paul (2018) for an in-depth discussion of how the blind chooser discovers *de se* truths when he becomes sighted.

6 This is not mere speculative musing. Individuals often have great difficulty adjusting to their new lives after such changes, in particular because the testimony they received beforehand was so inadequate for preparing them to understand and represent the nature of the changes they'd experience.

The second problem for the rational decision project stems from the transformation in my preferences: my preferences *ex ante* are inconsistent with my preferences *ex post*. Before I move to the city, my preferences are undefined. When I move to the city, I undergo an epistemic and personal transformation, creating preferences (values) in response to the nature of the new experiences I have. I discover. I adapt. I revise. After I move to the city, my preferences are constituted by my response to the lived experience of having a new job in a new city (Paul 2016, 2018; Paul and Healy 2018).

This means that preference change stemming from new experiences can create problems for choosers. If the states involving the new experience (and resulting from the new experience) are epistemically inaccessible to the chooser before she has the new experience, preference changes resulting from the new experience will be epistemically inaccessible to her before she undertakes the act that she chooses to perform. As a result, she cannot assume her preferences will remain constant, and any preference changes she might experience cannot be anticipated in the usual way.

Given the possibility of epistemic and personal transformation, I cannot make my decision rationally, by prospectively assessing the states of the world and the feasible acts and the consequences arising from their conjunction, in the way that the rational decision project requires. In sum:

On the standard rational choice model, I need to determine which act, moving to the city or staying at home, has the greatest expected utility. However, I cannot determine the best outcome if, *ex ante*, my utilities for the consequences are undefined. In addition, on the standard model, my preferences are assumed to be constant with respect to the optimal course of agency with respect to my decision. However, if the new experience creates preferences *ex post* which are incommensurable with my preferences *ex ante*, this assumption also fails. As a result, I cannot use the standard model to determine the best possible decision for myself in the circumstances.

4.3 Summary

The standard approach to rational decision-making takes an unbounded, objective approach, representing the decision problem under consideration from a 'bird's eye' or all-seeing view. The approach assumes that the terrain of the problem can be fully viewed (or represented), that the decision-maker is somehow independent from the world, and that the challenge is merely to calculate the best choice to make, that is, to determine the norms for perfect deliberation under these circumstances. These assumptions are necessary for security.

But as we've shown, for an important class of real-life cases, the rational decision project promises security but can't actually deliver it. Many decisions are not, cannot, and should not be taken from the unbounded, objective perspective. They are, instead, taken from the bounded, subjective perspective.

When we make decisions from a bounded subjective perspective, we cannot always see the entire terrain, nor deliberate perfectly. We explored the subjective limitations on the agent in our discussion of bounded awareness and epistemically transformative experience. When we make decisions from this sort of subjective perspective, we are embedded in the world in a way that we may change ourselves as we change the world, preventing us from being perfect deliberators. The failure of act-state independence also creates additional subjectivism, because it prevents us from understanding the decision problem

from a context-free, purely observational perspective. To put this in terms of the act-state-consequence representation, when act-state independence fails and preferences *ex ante* are inconsistent with preferences *ex post*, the agent's preferences cannot be made consistent by the agent's taking some sort of "perspective-free" approach.

For such cases, we need an alternative.

5. CHOOSING REASONABLY

If rational choice procedures fail to deliver the promise of security, how should we make choices, and what advice can we give to others who must choose? We think decision theory needs to supplement its rational choice procedure guaranteed to yield security in ideal conditions with *reasonable* choice procedures designed to provide us with as much *confidence* as possible in non-ideal conditions.

We do not take 'rational' and 'reasonable' to be synonymous. As we have seen, what we've called 'rational' choice proceeds by way of deductive logic from normatively compelling axioms to optimal choices. By contrast, what we'll call 'reasonable' choice merely requires that we choose in accordance with our best principles, without a guarantee of making the best choice.

5.1 What does 'reasonable' mean?

Etymologically, 'rational' is derived directly from the Latin 'ratio,' and ultimately from 'reri' ('consider') while reasonable is derived from the same root via French. A typical dictionary definition of 'rational' is 'consistent with or based on or using reason.'

In ordinary English usage and in theoretical discussions of choice and related issues, the two terms have different connotations. 'Rational' choice connotes the adoption of methodical procedures, most commonly involving the application of deductive logic to derive conclusions from a set of axioms and known 'primitive' preferences. 'Reasonable' in ordinary language encompasses a range of connotations, including

(a) sensible/fair (as judged by an impartial observer);
(b) based on a process of reasoning; and
(c) derived from reasons.

We will draw on all of these, without, we hope, committing ourselves to any particular interpretation of philosophical terms of art such as 'reason' and 'reasoning.' A *reasonable agent* conforms to (a–c) when making a decision.

5.2 What does 'confidence' mean?

For important classes of decisions, rational security is unattainable for the reasons we have set out. Yet decisions must be made, and we want to make better decisions rather than worse. How can we do this in the absence of rational security?

We suggest the more modest goal of increasing 'confidence' in our decisions. In this view, the goal of decision theory is to provide reasoning tools that enable us to make better use of the information and cognitive capacities available to us, without seeking the illusory security that is built into the rational decision project.

We will begin by describing some specific features that a notion of confidence should have. A notion of confidence based on reasonable choice should be

* amenable to arguments for and against confidence in a particular judgement/choice;
* consistent with the ordinary language meaning of confidence;
* increasing with increasing grounds for confidence (for example, independent arguments for the same choice);
* positively associated with (current and anticipated) wellbeing, in the sense that choices in which we have confidence should generally yield better outcomes than alternative actions in which we do not have confidence;
* be sensitive to relevant empirical work in psychological, cognitive, and computer science.

It's also important to draw contrasts with the idea of security derived from rational choice. First, confidence is most naturally interpreted in qualitative terms, although in some special cases it may be expressed in terms of numerical probabilities or credences.

Second, confidence is most naturally interpreted in terms of subjective, first-personal reasoning rather than objective, third-personal reasoning. Sources for confidence differ from person to person and this difference can't be reduced to revealed preferences and beliefs. Different individuals bring different histories and different capacities to the table. Temperamental differences may also matter. Some people may be highly confident in intuitive judgements about (say) transformative experiences, while others may gain more confidence from induction or (in limited domains) probabilistic inference.

5.3 Confidence as a qualitative partial order

In formal terms, we propose to represent confidence about beliefs in terms of a partial order over propositions and choices. This allows for a much richer interpretation of confidence than one restricted to numerical probabilities and credences. First, and most simply, it admits the case where confidence is described in qualitative rather than quantitative terms. So, confidence might be represented by a Likert-scale with a finite number of elements such as 'highly confident,' 'somewhat confident,' and 'not very confident'. Second, it allows for incommensurable notions of confidence: for example, we can distinguish between confidence derived from a theoretical model and confidence derived from empirical regularities.

In related work, Shear and Quiggin (2017) develop a modal logic of confidence based on justifications, and show that the logic is sound and complete with respect to an appropriately designed class of Kripke–Fitting frames.

5.4 Procedures and principles

Principles of reasonable choice are not, in general, universally applicable as is assumed for rational choice. However, they may be 'ecologically rational' in particular environments. Ecological rationality appears when the structure of boundedly rational decision mechanisms matches the structure of information in the environment (Todd and Gigerenzer 2012). Examples of ecologically rational procedures, for appropriate environments, are

the 'precautionary principle' (in the interpretation of Grant and Quiggin 2013*b*) and a directly opposed approach which may be called the 'exploratory principle.'

The precautionary principle begins with the assumption that we have available a 'status quo' action in which we have high/complete confidence in propositions about the state of the world and the state-contingent consequences of the action in question. Now consider an alternative action that, conditional on some proposition p will yield improved consequences, Suppose that p is believed with moderate, but not high confidence, and that there is little confidence about the consequences of the action if p does not hold. Then the Strong Precautionary Principle proposed by Grant and Quiggin (2013*b*) calls for the alternative action to be rejected in favor of the status quo.

A canonical example is a proposal to undertake an industrial or agricultural development in a previously unstudied area. Let proposition p, held with moderate confidence, be 'the area has no unique environmental values or particular vulnerabilities' and suppose that, in the absence of environmental damage, the development would yield positive benefits. The Precautionary Principle would require that the proposal be rejected, or deferred until further study was undertaken.

Directly opposed to the Precautionary Principle is the 'Discovery principle,' which treats the existence of poorly understood consequences as a reason in favor of adopting a particular course of action.

Which of these principles we might wish to adopt depends in part on the kind of environment in which we are making choices and also on our conception of ourselves and our attitude to transformative experience.

5.5 What we lose when we lose security, what we gain when we gain confidence

In abandoning security, we gave up the central goal of the rational decision project.

Recall that rational security implies:

S.1 Any two rational agents with the same preferences, prior beliefs and information would make the same choice (or be indifferent between choices).

S.2 Any rational agent, informed of the preferences and information for another agent, would suggest the same choice as optimal.

S.3 *Ex post*, after the outcome was observed, all rational agents would agree that the choice made *ex ante* was the correct one, conditional on the information and known preferences at that time.

S.4 For a finite problem, it is possible to specify an algorithm that may be implemented by an autonomous, model-utility based intelligent agent that will generate optimal choices for any decision problem given a specification of the chooser's preferences and available information.

However, as we have shown, the conditions of rational security S.1–S.4 cannot be realized by boundedly rational subjective agents facing real-world problems. Rather, real-world problems are characterized by the following modifications of these conditions.

S*.1 In general, boundedly rational subjective choosers will make different choices, and these differences cannot be fully accounted for by differences in information and preferences. Two agents, confident in their own judgements and choices may agree to differ.

This is impossible in the standard rational choice model as shown by Aumann (1976). Moreover, because deliberation is itself a transformative experience, in which preferences are formed rather than merely being discovered, the process of choosing reasonably allows us to improve our decision procedure by aligning our preferences with our chosen outcomes.

S*.2 The second-person problem of choosing on behalf of others may be fundamentally different from first-person choice.

S*.3 From an *ex post* or external perspective, some choice procedures (in cases of bounded rational agents) may be judged as ecologically rational (well adapted to the environment in which the choices are made) or not. *Ex ante*, boundedly rational subjective agents cannot make this judgement about their own subjective choices.

S*.4 Algorithms, like human agents, are finite and bounded, more capable in some respects than humans but less capable in others. For decisions of the complexity typically found in life, algorithms are incapable of providing security.

We've shown how boundedly rational subjective choosers cannot make choices with rational security. We propose that, even if we recognize that boundedly rational subjective choosers cannot choose with rational security, that is, cannot choose rationally, *they can choose reasonably and thereby enhance their confidence*. The term confidence encompasses the following claims:

C.1 Reasonable agents with similar preferences and information will recognise each others' choices as reasonable, though they may not make the same choices.

C.2 Reasonable agents, with some understanding of the choice problem faced by other agents, and the preferences and beliefs of those agents can give useful advice which may lead to improved choices.

C.3 *Ex post*, after the outcome is observed, reasonable agents would agree that the choice made *ex ante* was a reasonable one, conditional on the information and known preferences at that time. However, they may conclude that, given the same information, and a chance to reconsider, they would make a different choice.

C.4 It is possible to specify algorithms that narrow the set of choices that might be considered reasonable, for example by testing for dominance and applying transitivity.

Thus, in shifting from security to confidence, we lose S.1–4 but gain C.1–4.

6. RECAP: BRIGHT LIGHTS, BIG CITY

Return to the reflection on whether to move to the city. I'm considering moving from the rural village in which I have always lived, to a major city, possibly in another country. I'm deliberating about what I should do, and in an ordinary sense I'm uncertain about how to act. I'm uncertain about whether to move to a city and, if I decide to move, which city to move to. I don't know for sure what will happen if I do move, or how I will feel about the outcomes when they occur. I'm a boundedly rational subjective chooser.

Can I make my decision rationally, by prospectively assessing the relevant external circumstances, the possible consequences, and the functional relationship between them, in the way that the rational decision project requires? Using a standard approach to decision

theory, I might try to represent my decision problem in terms of rational uncertainty. If I were to try to use a standard model, as I deliberate about where I might move to, I reflect on the possible consequences of my actions, and the likelihood that these consequences will be realized contingent on my actions. To each action-consequence pair, I assign a probability and a utility. I then choose the action with the highest expected utility.

For example, considering the possible consequences of a move to the city, I might be uncertain as to what kind of job I might find, if any. An ideal rational choice procedure would entail, first, estimating the utility of being employed in different kinds of jobs and of being unemployed (which might vary depending on the city in which these outcomes were realized). Then, determine in which states of the world (for example, those where other people from my home village can help me find work) I would be employed and with what kind of job, and in which states of the world I would be unemployed. According to standard decision theory, I should estimate the probabilities of these states using the calibration procedure described above. Finally, I should calculate the expected utilities for each choice, being employed (for each kind of job), and being unemployed, and pick the action that maximizes my expected utility.

As we have seen, however, the proposed calculations involve unreasonable assumptions about my capacity to envisage and evaluate the future. I am, after all, a bounded, subjective agent! I am a boundedly rational subjective chooser. So, let's consider a reasonable choice procedure instead.

In reasoning about a move to the city, I might begin with some qualitative judgements about the world, and consideration of my personal dispositions. An initial step would be to consider whether the world (or at least the part of it I am considering) is characterized by unfavorable surprises, in which case I might apply some version of the Precautionary Principle, or by favorable surprises, in which case I might apply the Discovery Principle. I should also consider personal dispositions, in particular, those regarding transformative experiences that I might encounter in the city.

Suppose, for illustration, that a naïve application of expected utility theory would favor a move to the city. That is, based on the possibilities I have considered explicitly, the probability of an outcome yielding enhanced utility (a higher paying job and a nice house) is greater than the probability of an outcome yielding reduced utility (unemployment). Depending on my view of the world and personal dispositions, reasonable choice procedures might either confirm or reject this assessment.

Clearly, an optimistic view of the world as characterized by favorable surprises, combined with a positive desire for personal transformation, would reinforce the decision to move to the city.[7]

Suppose, by contrast, that I have a view of the world as characterized by unfavorable prizes and am inclined to adopt the Precautionary Principle as the basis of my reasoning procedure. Strong versions of the Precautionary Principle would require me to rule out the poorly understood choice of moving to the city, and instead to stay at home.

However, if I adopt a modified version of the Precautionary Principle, as suggested by Grant and Quiggin (2013*b*), I might want to consider moving to the city, but retaining the option of returning to the village if things turned out badly. My attitude to this option will

7 Classic literary representations of migrants to the city, going back at least as far as Dick Whittington, exemplify these characteristics.

in turn depend on how I reason about transformative experience. I might fear that, having lived in the city, I would become unhappy with village life, even if I recognized it as offering me better living standards. Alternatively, I might welcome the transformative experience and feel that having had an adventure, even one that turned out badly, would remove some of my existing discontent with village life. The first of these judgements would imply staying at home while the second would favor moving to the city.

7. CONCLUDING COMMENTS

The rational decision project has made substantial contributions to our understanding of the way people make choices and to providing tools for formal reasoning about choices. Nevertheless, it has proved inadequate as a complete model, both descriptively and normatively. It has long been evident that people do not, in practice, satisfy the axiomatic requirements of expected utility theory, the core model of the rational decision project. Despite a proliferation of generalizations of the basic model, there remains no generally accepted model capable of offering a satisfactory description of observed choices.

More problematic perhaps is that large classes of decisions, including the most important choices people make in their lives, remain outside the scope of rational choice theory. Most real-world problems are simply too complex to allow for a comprehensive representation in the state-act-consequence framework, without which the tools of the rational choice project cannot be relied upon to yield good outcomes. And the problem of transformative experience means that the consequences of possible choices cannot be represented and evaluated in the way required.

We therefore propose to supplement the model with the more modest objective of reasonable choice and, in place of the (often illusory) security offered by the rational decision project suggest a goal of choosing with confidence. Choosing reasonably involves recognizing the limits of the situation, selecting the best rule for the situation, and applying it. We can hope for a good result (choosing reasonably gives us the most confidence we could have that we might get a good result), but there is no guarantee that utility will be maximized in the ordinary sense.

By abandoning the goal of a comprehensive model of rational choice, applicable to all people and in all circumstances, we open up the possibility of developing tools and procedures that can enhance confidence in particular kinds of decisions, made in environments to which these tools and procedures are adapted. The better goal is to find reasoning tools that enable us to make our best use of the information and cognitive capacities available to us, without requiring the security enshrined by the rational decision project.[8]

REFERENCES

Aumann, R. J. 1976. 'Agreeing to disagree', *Annals of Statistics*, 4: 1236–9.
Buchak, L. 2017. *Risk and Rationality*. New York, NY: Oxford University Press.

8 Thanks are due to Branden Fitelson, Ram Neta and Richard Pettigrew for extremely helpful comments and discussion.

Fagin, R. and Halpern, J. Y. 1987. 'Belief, Awareness, and Limited Reasoning.' *Artificial Intelligence*, 34: 39–76.
Ellsberg, D. 1961. 'Risk, Ambiguity and the Savage Axioms.' *Quarterly Journal of Economics*, 75: 643–69.
Gigerenzer, G. 2008. *Rationality for Mortals: How People Cope With Uncertainty*. Oxford: Oxford University Press.
Grant, S. and Quiggin, J. 2013*a*. 'Inductive Reasoning about Unawareness.' *Economic Theory* 54: 717–55.
—— 2013*b*. 'Bounded Awareness, Heuristics and the Precautionary Principle.' *Journal of Economic Behavior and Organization*, 93: 17–31.
Jackson, F. 1982. 'Epiphenomenal Qualia.' *Philosophical Quarterly*, 32: 127–36.
Jeffrey, R. 1990. '*Formal Logic: Its Scope and Limits*.' New York, NY: McGraw-Hill.
Kolodny, N. and MacFarlane, J. 2010. 'Ifs and Oughts.' *Journal of Philosophy*, 107: 115–43.
Lewis, D. 1986. 'What Experience Teaches.' Reprinted in *Papers in Metaphysics and Epistemology*, pp. 262–90. Cambridge: Cambridge University Press.
Modica, S. and Rustichini, A. 1994. 'Awareness and Partitional Information Structures.' *Theory and Decision*, 37: 107–24.
Paul, L. A. 2015. *Transformative Experience*. Oxford: Oxford University Press.
—— 2016. 'The Subjectively Enduring Self.' In I. Phillips (ed.), *Routledge Handbook of the Philosophy of Temporal Experience*. London: Routledge.
—— 2018. '*De se* preferences and empathy for future selves.' *Philosophical Perspectives*. doi.org/10.1111/phpe.12090.
—— **and Healy, K.** 2018. 'Transformative Treatments.' *Noûs*, 52: 320–35.
Pettigrew, R. 2015. 'Transformative Experience and Decision Theory', *Philosophy and Phenomenological Research*, 91: 766–74.
Quiggin, J. 1982. 'A Theory of Anticipated Utility.' *Journal of Economic Behavior and Organization*, 3: 323–43.
Ramsey, F. P. 1926. 'Truth and Probability.' In R. B. Braithwaite (ed.), *Ramsey, F. P., 1931, The Foundations of Mathematics and other Logical Essays*, Ch. VII, pp. 156–98. London: Kegan, Paul, Trench, Trubner & Co.; New York: Harcourt, Brace and Company.
Robbins, L. 1938. 'Interpersonal Comparisons of Utility: A Comment.' *Economic Journal*, 48: 635–41.
Savage, L. J. 1954. *Foundations of Statistics*. New York, NY: Wiley.
Shear, E. and Quiggin, J. 2017. 'Justification Logic with Confidence.' Working Paper, University of Queensland.
Thaler, R. H. and Sunstein, C. R. 2008. *Nudge: Improving Decisions About Health, Wealth, and Happiness*. New Haven, CT: Yale University Press.
Todd, P. M. and Gigerenzer, G. 2012. *Ecological Rationality: Intelligence in the World*. Oxford: Oxford University Press.
Williams, B. 1981. 'Internal and External Reasons.' In *Moral Luck*, pp. 101–13. Cambridge: Cambridge University Press.

L. A. PAUL is Eugene Falk Distinguished Professor of Philosophy at the University of North Carolina at Chapel Hill and Professorial Fellow at St Andrews University in Scotland. She is the author of *Transformative Experience* (Oxford University Press, 2014) and is co-author, with Ned Hall, of *Causation: A User's Guide* (Oxford University Press, 2013). She is currently writing a book on transformative experience and the nature of the self to be published by Farrar, Straus, and Giroux.

JOHN QUIGGIN is a Vice-Chancellor's Senior Fellow in Economics at the University of Queensland. He is prominent both as a research economist and as a commentator on Australian and international economic policy. His book, *Zombie Economics: How Dead Ideas Still Walk Among Us*, was released in 2010 by Princeton University Press, and has been translated into eight languages.