CrossMark

A preference model for choice subject to surprise

Simon Grant · John Quiggin

Abstract Grant and Quiggin (Econ Theo 54:717–755, 2013a, J Econ Behav Organ 93:17–31, 2013b) suggest that agents employ heuristics to constrain the set of acts under consideration before applying standard decision theory, based on their restricted model of the world to the remaining acts. The aim of this paper is to provide an axiomatic foundation, and an associated representation theorem, for the preference model proposed by Grant and Quiggin. The unawareness of the agent is encoded both in the specification of the states and in an elaboration of the set of consequences. We illustrate how the representation can be applied to rationalize two notions of the precautionary principle that were proposed and discussed in Grant and Quiggin (J Econ Behav Organ 93:17–31, 2013b).

1 Introduction

Von Neumann and Morgenstern Neumann and Morgenstern (1947) and Savage (1954) provided the foundations for the formal theory of decision under uncertainty. Von Neumann and Morgenstern presented a simple set of axioms under which preferences

S. Grant (✉) · J. Quiggin
The University of Queensland, Brisbane, Australia
e-mail: simongrant1962@gmail.com

S. Grant
The Australian National University, Canberra, Australia

over risky prospects (lotteries allocating prizes with known objective probabilities) can be represented by an expected utility functional. Savage addressed the more general setting of representing preferences over acts, that is, mappings from a set of possible states of the nature, without known objective probabilities, to a set of conceivable consequences.

From the 1980s onwards, a variety of alternatives to, and generalizations of, expected utility theory were proposed, including rank-dependent utility theory, weighted utility theory, cumulative prospect theory, and maxmin expected utility theory. These theories were motivated by the observation of violations of the predictions of expected utility theory, such as the Allais problem (Allais 1953) and the Ellsberg problem (Ellsberg 1961).

From the normative perspective of expected utility (EU) theory, the EU axioms are implications of rationality, and non-EU representations of choice involve violations of rationality. Nevertheless, in common with EU, all the generalized EU models discussed above require unbounded rationality, or at least unbounded reasoning capacity. Individuals are presumed to consider all possible states of nature, and arbitrarily large sets of acts.

In the Twenty-first century, research in decision theory has focused more and more on the case of bounded rationality. Among the most important examples of bounded rationality is the case where agents have only partial awareness of the set of states of nature. A large literature has developed concerning the representation of unawareness. Notable contributions include Halpern and Rêgo (2009) and Heifetz et al. (2006). Schipper (2014) provides an extensive bibliography.

By contrast with the focus on epistemological problems, relatively little attention has been paid to problems of decision under bounded awareness, as in Grant and Quiggin (2013a).[1] For some representations of unawareness, the problem is trivial. If agents are naively unaware of their own unawareness, their preferences may be represented by expected utility theory or one of its generalizations, applied to the restricted representation of the world available to them.

More difficult problems arise when individuals are aware (for example, on the basis of inductive reasoning) of their own unawareness. In this case, the naive approach of maximizing expected utility with respect to a restricted representation of the world is unappealing, and vulnerable to manipulation by other, more aware, agents. On the other hand, any attempt to take unconsidered possibilities into account appears to run into an insoluble paradox—if possibilities are taken into account in decision making, they cannot be said to be unconsidered.

Grant and Quiggin (2013a, b) suggest that agents employ heuristics to constrain the set of acts under consideration before applying standard decision theory, based on their restricted model of the world to the remaining acts. An example, developed in Grant and Quiggin (2013b) is the precautionary principle, commonly advocated as a requirement for decision making in relation to environmental hazards.

The aim of this paper is to provide an axiomatic foundation, and an associated representation theorem, for the preference model proposed by Grant and Quiggin.

---

[1] Other attempts to axiomatize preferences of a decision-makers facing unawareness include Li (2008), Schipper (2014), Karni and Vierø (2014) .

The unawareness of the agent is encoded both in the specification of the states and in an elaboration of the set of consequences. The state space includes both a standard Savage state space and a set of 'surprise' states. The 'standard' set of consequences (taken to be a closed bounded interval of 'monetary' consequences) is augmented by two unforeseen consequences, one unfavorable and ranked below the worst possible monetary consequence and the other favorable and ranked above the best possible monetary consequence. Unforeseen consequences can arise only in surprise states.

The formal specification of the above framework appears in Sect. 2. We then posit in Sect. 3 that the preferences restricted to those acts involving no unforeseen consequences admit a Gul and Pesendorfer (2014) expected uncertain utility (EUU) representation.[2] We extend this to the entire set of acts by means of a three-cell partition of the set of acts. The first cell of the partition comprises those acts the decision-maker perceives to be subject to an unforeseen unfavorable surprise. The second cell comprises those acts the decision-maker perceives to be subject to an unforeseen favorable surprise but not subject to an unforeseen unfavorable surprise. The third cell consists of acts for which the possibility of an unforeseen consequence may be neglected. Our axioms ensure that all acts in the first (respectively, second) cell are ranked below (respectively, above) those in the third.

To allow for the representation to represent the preferences of pairs of acts that reside within the same cell of our three-cell partition of the acts, we define an operation that maps any act into the set of acts with no unforeseen consequences by replacing any occurrence of the unforeseen unfavorable (respectively, favorable) consequence with the worst (respectively, best) possible monetary consequence. The resulting act is thus one that is a member of the subset of acts involving no unforeseen consequences, which is the domain of the EUU representation. We derive an associated representation and illustrate how the representation can be applied to rationalize two notions of the precautionary principle that were proposed and discussed in Grant and Quiggin (2013b).

Section 4 offers some concluding comments.

2 Framework

We begin with an elaboration of the standard Savage framework by defining a triple $(\Omega, \mathcal{C}, \mathcal{F})$ consisting of a state space Ω, a consequence space $\mathcal{C}$, and a set of acts $\mathcal{F}$. The non-standard features of the model arise from the specification of these elements in a way that allows for both 'surprises' and 'unforeseen' consequences.

We take the state space to be a product space $\Omega = S \times \{0, 1\}$. The first component corresponds to the standard way that Savage encodes everything a decision-maker may be uncertain about. That is, S is a mutually exclusive set of descriptions of the world leaving no relevant aspect undescribed *that can be described*. However, because the decision-maker is cognizant of the fact that his understanding of the world is *incomplete* (which in turn renders incomplete his description of the world as embodied in S) we augment the Savage-style description s with a binary component.

---

[2] This model is closely related to Jaffray's (1989) preference model over belief functions.

This binary component is 0 if the situation is one in which no surprises can arise, and 1 if the situation allows for the possibility of surprise since it falls outside his (necessarily) bounded understanding of the world. We refer to any augmented state $(s, 0)$ as a *surprise-free* state and any state $(s, 1)$ as one that allows for the possibility of surprise, including unforeseen consequences. We denote the set of surprise-free states, $S \times \{0\}$, by Ω_0, and its complement, by Ω_1. (That is, $\Omega_1 = S \times \{1\}$).

The set of consequences is taken to be $\mathcal{C} = X \cup \{\ominus, \oplus\}$, where: $X = [\ell, m] \subseteq \mathbb{R}$, with $\ell < m$, is the set of (known) monetary consequences; and $\ominus$ (respectively, $\oplus$) is a label for an *unfavorable* (respectively, *favorable*) *unforeseen* consequence that may arise in the surprise event. The (strict) inequality ordering $<$ is extended to $X \cup \{\ominus, \oplus\}$ by setting $\ominus < \ell$ and $m < \oplus$.

Let Σ_S denote a σ-algebra of subsets of S. From this, we define the following two σ-algebras of subsets of Ω. The first is the collection of events to which acts will be adapted, and the second will turn out to be the subset of those events for which the decision-maker is able to precisely quantify the likelihood.

Events Let $\mathcal{E}$ denote the smallest σ-algebra that contains all sets of the form $A \times \{0\}$ and $A \times \{1\}$, for some $A \in \Sigma_S$.

M-Events Let Σ denote the smallest σ-algebra that contains all sets of the form $A \times \{0, 1\}$, for some $A \in \Sigma_S$.

We take the set of acts $\mathcal{F}$ to be the subset of the mappings $f : \Omega \rightarrow X \cup \{\ominus, \oplus\}$, that are adapted to $\mathcal{E}$, and for which $f(S \times \{0\}) \subset X$.[3] We associate with any act $f \in \mathcal{F}$, the two functions $f_0 : S \rightarrow X$ and $f_1 : S \rightarrow X \cup \{\ominus, \oplus\}$ defined by setting $f_0(s) := f(s, 0)$ and $f_1(s) := (s, 1)$.

An act $f \in \mathcal{F}$ has *no unforeseen consequences* if $f^{-1}(X) = \Omega$. We denote by $\bar{\mathcal{F}}$ the set of acts with no unforeseen consequences.

An act $f \in \mathcal{F}$ is said to be *surprise-free* if $f(s, 0) = f(s, 1)$, for all $s \in S$. We denote by $\bar{\mathcal{F}}_0$ the set of surprise-free acts.

Observe that $\bar{\mathcal{F}}_0 \subset \bar{\mathcal{F}} \subset \mathcal{F}$.

Following common practice, we will abuse notation and let each $x \in X$ denote the constant act f in $\mathcal{F}$ for which $f(\omega) = x$ for all ω in Ω, as well as the constant mappings $f_0, f_1 : S \rightarrow X$ for which $f_0(s) = f_1(s) = x$ for all s in S.[4]

For any subset $E \in \Omega$ and any pair of acts f and f', let $f_E f'$ denote the act given by

$$f_E f'(\omega) = \begin{cases} f(\omega) & \text{if} \quad \omega \in E \\ f'(\omega) & \text{if} \quad \omega \notin E \end{cases}.$$

---

[3] That is, neither of the two unforeseen consequences, $\ominus$ and $\oplus$, can occur if a surprise-free state $\omega \in S \times \{0\}$ obtains.

[4] Note that, since $f_0^{-1}(X) = S$ there are no constant acts f in $\mathcal{F}$ associated with the unforeseen consequences $\ominus, \oplus$.

3 Preferences

The decision-maker's preferences $\succsim$ are defined on $\mathcal{F}$. We first impose conditions such that these preferences restricted to acts for which there are no unforeseen consequences (that is, to acts in $\bar{\mathcal{F}}$) conform to a special subclass of the Gul–Pesendorfer family of *expected uncertain utility functions*. Moreover, when further restricted to acts for which there are no surprises (that is, to acts in $\bar{\mathcal{F}}_0$) they conform to subjective expected utility theory.

3.1 Foreseen consequences

Gul and Pesendorfer (2014) axiomatize the class of *expected uncertain utility functions*, which are characterized by:

(1) a prior, comprising a countably additive and non-atomic probability measure defined over a σ-algebra of *ideal events*; and,
(2) an interval utility index defined over 'intervals' of consequences.

In Gul and Pesendorfer's set-up, it is assumed the individual can quantify without difficulty the likelihood of an ideal event. Intuitively, this is connected to the defining property of ideal events, namely, they are ones for which Savage's sure thing principle holds.

Definition 1 (*Ideal Events*) A subset $E \in \mathcal{E}$ is *ideal* if for every pair of acts, f and g in $\bar{\mathcal{F}}$, $f \succsim g_E f$ implies $f_E g \succsim g$. Let $\mathcal{I}$ denote the set of ideal events.

Hence, the restriction of her preferences to acts which are *adapted* to ideal events, conforms to the standard model of subjective expected utility. When confronted with acts that are not measurable with respect to ideal events, it is as if the individual forms *ideal* lower and upper bounds, representing the range of possible outcomes implied by the uncertainty that she cannot precisely quantify and then evaluates such an act by its expected *uncertain* utility. As Gul–Pesendorfer demonstrate their model is flexible enough to accommodate Allais-problems as well as Ellsberg-problems.

In our setting, however, the only feature of the uncertainty the individual perceives to be *unquantifiable* is whether or not a 'surprise' will occur. That is, the set of ideal events will be the collection of subsets Σ defined in Sect. 2. Hence for our application of Gul and Pesendorfer's model, the prior is defined on Σ. Thus, for any set $A \in \Sigma_S$, the event $A \times \{0, 1\}$ is an ideal event (and hence admits a precise quantification under the prior in the EUU representation). Neither $A \times \{0\}$ nor $A \times \{1\}$, however, is ideal and so the uncertainty of such events cannot be precisely quantified. Moreover, the events $S \times \{0\} = \Omega_0$ and $S \times \{1\} = \Omega_1$ will turn out to be what Gul and Pesendorfer dub "diffuse" since their likelihoods cannot be bounded by *any* ideal event, except, trivially, by the universal event Ω (and any event the decision-maker judges to be just as likely as the universal event) and by the empty set (and any event the decision-maker views as "null"). Formally:

Definition 2 (*Null Events*) A subset $E \in \mathcal{E}$ is *null* if for every pair of acts, f and g in $\bar{\mathcal{F}}$, $f \sim g_E f$. Let $\mathcal{N}$ denote the set of null events.

Definition 3 (*Diffuse Events*) A subset $D \in \mathcal{E}$ is *diffuse* if for any $E \in \mathcal{I} \backslash \mathcal{N}, E \cap D \neq \emptyset$ and $E \cap (\Omega \backslash D) \neq \emptyset$. Let $\mathcal{D}$ denote the set of diffuse events.

That is, an event is diffuse if no non-null ideal event is contained in either that event or its complement.

The *unquantifiable* uncertainty arising from the possibility of surprise (but precluding at this stage any possibility of unforeseen consequences) is accommodated by an interval utility index. Formally, this is a function $u : I \rightarrow \mathbb{R}$, where $I = \{[\ell', m'] : \ell \leq \ell' \leq m' \leq m\}$, that is, the set of closed intervals that are a subset of $[\ell, m]$. Furthermore, it is monotonic in the sense that $u([\ell', m']) \geq u([\ell'', m''])$, whenever $\ell' \geq \ell''$ and $m' \geq m''$.

Our initial assumption has three parts. The first is that all events in Σ are ideal. The second is that the set of states in which the decision-maker may be surprised is diffuse. The last, is that the set of preferences restricted to acts that do not allow for the possibility of unforeseen outcomes satisfy the six axioms of Gul and Pesendorfer (2014).[5]

Axiom 1 (*EUU over $\bar{\mathcal{F}}$*)

(*i*) Every M-event E in Σ is ideal.
(*ii*) The set Ω_1 is diffuse.
(*iii*) The preferences $\succsim$ restricted to $\bar{\mathcal{F}}$ satisfy Axioms 1–6 of Gul and Pesendorfer (2014).

Our first result is a corollary of Gul and Pesendorfer's (2014) representation theorem.

Theorem 1 *The preferences $\succsim$ satisfy Axiom 1 if and only if there exists a countably additive and non-atomic probability measure μ on Σ and an interval utility u such that $\succsim$ restricted to $\bar{\mathcal{F}}$ admits a representation of the form:*

$$\bar{V}(f) = \int_S u(\min(f_0(s), f_1(s)), \max(f_0(s), f_1(s))) \, \mu_S(ds) ,$$

where μ_S is the marginal distribution on S obtained from μ.

We observe that the restriction of $\succsim$ to the set of surprise-free acts $\bar{\mathcal{F}}_0$ admits a standard subjective expected utility representation.

Lemma 1 *If for each monetary consequence x in $[\ell, m]$, we set $v(x) := u(x, x)$, then it follows that for $\succsim$ restricted to $\bar{\mathcal{F}}_0$, we have $\bar{V}(f) = \int_S v(f_0(s)) \, \mu_S(ds)$.*

Since Ω_1 is a diffuse event, the decision-maker cannot quantify what is the probability a state that allows for the possibility of surprise will or will not obtain. More precisely, if we define for any event A (not necessarily in Σ) the inner probability $\mu_*(A)$ and outer probability $\mu^*(A)$ of A as follows:

$$\mu_*(A) = \sup_{E \in \Sigma, E \subset A} \mu(E) \quad \text{and} \quad \mu^*(A) = \inf_{E \in \Sigma, E \supset A} \mu(E),$$

then we have: $\mu_*(\Omega_1) = \mu(\emptyset) = 0$ and $\mu^*(\Omega_1) = \mu(\Omega) = 1$.

---

[5] For the convenience of the reader we list these axioms in the appendix.

3.2 Unforeseen consequences

Turning to acts that allow for the possibility of unforeseen consequences, our notion that a decision-maker perceives an act to be subject to an unforeseen unfavorable surprise is that the event in which the unfavorable unforeseen consequence $\ominus$ can arise is not null.

Definition 4 An act f in $\mathcal{F}$ is deemed *subject to an unforeseen unfavorable (*uu*) surprise* if $f^{-1}(\ominus) \notin \mathcal{N}$.

Notice that the set $f_1^{-1}(\ominus) \times \{0, 1\}$ is the 'smallest' ideal event (that is, is in Σ) that contains $f_1^{-1}(\ominus)$. So for preferences that satisfy axiom 0 and thus whose restriction to $\bar{\mathcal{F}}$ admits an EUU representation characterized by $((\mu, \Sigma), u)$, we have an act f is subject to an unforeseen unfavorable (*uu*)surprise if and only if $\mu^*\left(f^{-1}(\ominus)\right)\left(= \mu\left(f_1^{-1}(\ominus) \times \{0, 1\}\right)\right) > 0$.

Similarly, for favorable unforeseen surprises we have:

Definition 5 An act f in $\mathcal{F}$ is deemed *subject to a unforeseen favorable (*uf*) surprise* if $f^{-1}(\oplus) \notin \mathcal{N}$.

Just as was the case for acts subject to unforeseen unfavorable surprises, for preferences that satisfy axiom 0 and thus whose restriction to $\bar{\mathcal{F}}$ admits an EUU representation characterized by $((\mu, \Sigma), u)$, we have an act f is subject to an unforeseen favorable surprise (*uf*) iff $\mu^*\left(f^{-1}(\oplus)\right)\left(= \mu\left(f_1^{-1}(\oplus) \times \{0, 1\}\right)\right) > 0$.

Denote by $\mathcal{F}^{uu}$ (respectively, $\mathcal{F}^{uf}) \subset \mathcal{F}$, the subset of acts that are subject to unforeseen unfavorable (respectively, favorable) surprises. As foreshadowed in the introduction above, we assume that any act in $\mathcal{F}^{uu}$ (respectively, $\mathcal{F}^{uf} - \mathcal{F}^{uu}$) is ranked below (respectively, above) the constant act that yields ℓ (respectively, m).

Axiom 2 The preferences $\succsim$ exhibit *a strict aversion to unforeseen unfavorable surprises* if $\ell \succ f$, for every act f in $\mathcal{F}^{uu}$.

Axiom 3 The preferences $\succsim$ exhibit *a preference for unforeseen favorable surprises* if $f \succ m$, for any act f in $\mathcal{F}^{uf} - \mathcal{F}^{uu}$.

To express the final axiom and the representation theorem, it is convenient to define the following operation that maps any act into the set $\bar{\mathcal{F}}$ by replacing any occurrence of the unforeseen unfavorable consequence $\ominus$ with the worst foreseen consequence ℓ and correspondingly, replacing any occurrence of the unforeseen favorable consequence $\oplus$ with the best foreseen consequence m.

Definition 6 For any f in $\mathcal{F}$, denote by $\bar{f}$ the *truncation* of the act in $\bar{\mathcal{F}}$ given by

$$
\bar{f}(\omega) = \begin{cases} f(\omega) & \text{if} \quad f(\omega) \in X \\ \ell & \text{if} \quad f(\omega) = \ominus \\ m & \text{if} \quad f(\omega) = \oplus \end{cases}.
$$

Notice that for any f in $\bar{\mathcal{F}}$, since by definition $f^{-1}(\ominus) = f^{-1}(\oplus) = \emptyset$, it follows by construction that $\bar{f} = f$.

The last axiom in conjunction with the other three axioms, which allows us to interpret the preferences as rationalizing a two-stage decision procedure, in which the first stage involves allocating each act into one of three 'bins.' The first contains those acts not subject to unforeseen consequences; the second contains those acts subject to unforeseen unfavorable surprises; and, the third contains the remaining acts (that is, those subject to unforeseen favorable surprises *and not* subject to unforeseen unfavorable surprises).

Axiom 2 tells us that any act in the first bin is ranked strictly above any act in the second bin. Axiom 3 tells us that any in the third bin is ranked strictly above any act in the first bin (and hence, by transitivity, strictly above any act in the second bin).

Our final axiom says that the preference between any pair of acts allocated to the same 'bin' is governed by the preference between their respective 'projections' into $\bar{\mathcal{F}}$.

Axiom 4 The preferences $\succsim$ permit *comparability of unforeseen surprises* if for any pair of acts f and g in $\mathcal{F}$, any one of $\{f, g\} \subset \mathcal{F} - (\mathcal{F}^{uu} \cup \mathcal{F}^{uf})$, $\{f, g\} \subset \mathcal{F}^{uu}$ or $\{f, g\} \subset \mathcal{F}^{uf} - \mathcal{F}^{uu}$ holds,

$$f \succsim g \Leftrightarrow \bar{f} \succsim \bar{g}.$$

Putting all four axioms together delivers the following representation.

Theorem 2 *(Representation Theorem) Suppose $\succsim$ is a binary relation on the acts $\mathcal{F}$. Then, the following are equivalent.*

 (i) The relation $\succsim$ satisfies Axioms 1–4.
(ii) There exists a countably additive and non-atomic probability measure μ on Σ and an interval utility u, such that the relation $\succsim$ admits a representation of the form:

$$V(f) = \begin{cases} \bar{V}(\bar{f}) & \text{if } \mu^*\left(f^{-1}(\ominus) \cup f^{-1}(\oplus)\right) = 0 \\ \bar{V}(\bar{f}) - V(m) + V(\ell) & \text{if } \mu^*\left(f^{-1}(\ominus)\right) > 0 \\ \bar{V}(\bar{f}) + V(m) - V(\ell) & \text{if } \mu^*\left(f^{-1}(\ominus)\right) = 0 \,\&\, \mu^*\left(f^{-1}(\oplus)\right) > 0, \end{cases}$$

where

$$\bar{V}(f) = \int_S u\left(\min\left(f_0(s), f_1(s)\right), \max\left(f_0(s), f_1(s)\right)\right) \mu_S(ds),$$

where μ^* is the outer measure and μ_S is the marginal measure on S obtained from μ.

The proof of the sufficiency of the axioms for the preferences to admit such a representation appears in the appendix.

For necessity, first notice for a representation of the form given in (ii), any act placed in the first bin is ranked strictly above any act in the second bin and strictly below any

act in the third bin, notice by construction for any act f in $\mathcal{F}$, $\bar{V}\left(\bar{f}\right) \in \left[\bar{V}\left(\ell\right), \bar{V}\left(m\right)\right]$. Hence for any three acts $f, g,$ and h such that $\mu^*\left(f^{-1}\left(\ominus\right) \cup f^{-1}\left(\oplus\right)\right) = 0$, $\mu^*\left(g^{-1}\left(\ominus\right)\right) > 0$, $\mu^*\left(h^{-1}\left(\oplus\right)\right) > 0$ and $\mu^*\left(h^{-1}\left(\ominus\right)\right) = 0$, we have

$$V(f) - V(g) = \left[\bar{V}\left(\bar{f}\right) - \bar{V}\left(\bar{g}\right)\right] + \left[\bar{V}\left(m\right) - \bar{V}\left(\ell\right)\right] > 0,$$
$$V(h) - V(f) = \left[\bar{V}\left(\bar{h}\right) - \bar{V}\left(\bar{f}\right)\right] + \left[\bar{V}\left(m\right) - \bar{V}\left(\ell\right)\right] > 0.$$

Thus, for any act f:

(1) subject to an unforeseen unfavorable surprise (that is, $\mu^*\left(f^{-1}\left(\ominus\right)\right) > 0$) we have $V(f) < V(\ell)$, so Axiom 2 holds;
(2) subject to an unforeseen favorable surprise but not subject to an unforeseen unfavorable surprise (that is, $\mu^*\left(f^{-1}\left(\ominus\right)\right) = 0$ and $\mu^*\left(f^{-1}\left(\oplus\right)\right)$) we have $V(f) > V(m)$, so Axiom 3 holds.

Furthermore, for any pair of acts that reside in the same 'bin', their ranking is determined by the evaluation of their respective truncations, so Axiom 4 holds. And finally, since $V(\cdot)$, restricted to $\overline{\mathcal{F}}$ is by construction an EUU representation, Axiom 1 holds.

3.3 Rationalizing precautionary principles

As an illustration of the scope of the preferences described here, we consider the precautionary principle. This principle, most commonly presented as a guide to environmental policy decisions, is widely used, but controversial and hard to define precisely. While no exact formulation of the principle has achieved unanimous support, the precautionary principle has been widely advocated, the version adopted by the Wingspread Conference (1998) is fairly representative:

> Where an activity raises threats of harm to the environment or human health, precautionary measures should be taken even if some cause and effect relationships are not fully established scientifically. In this context, the proponent of an activity, rather than the public, bears the burden of proof.

This idea has proved difficult to interpret in a decision-theoretic framework. As Sunstein (2003) argues, strong versions of the principle appear to rule out any possible option, since all activities raise some threat of harm. For example, regulation aimed at mitigating health risks will normally involve economic costs and thereby limit the resources available to address other health or environmental concerns. On the other hand, weak versions commonly amount to little more than a restatement of standard theory. For example, Gollier et al. (2000) interpret the precautionary principle in terms the option value associated with waiting for further information. Since a complete risk analysis should incorporate option values, this defence of the precautionary principle amounts to the claim that, in practice, risk analyses are usually incomplete, and that this error can be corrected by adoption of the precautionary principle.

Grant and Quiggin propose Strong and Modified versions of the Precautionary Principle which lie between the extremes described above in the sense that the Strong

version does not preclude all possible choices, while the Modified version rules out some choices that might arise from a standard risk analysis based on the assumption of full awareness.

Grant and Quiggin (2013b) argue that the Precautionary Principle may be understood in terms of aversion to unforeseen unfavorable surprises, and that, in an interactive context, this interpretation naturally gives rise to a burden of proof on the proponents of potentially hazardous activities. This presumption is not supposed to be general. Rather it arises in the context of what Grant and Quiggin call 'domains of unfavorable surprises' such as activities involving poorly understood environmental or health risks. The converse category, 'domains of favorable surprises' is exemplified by scientific research, where the whole rationale of the activity is to discover phenomena that were previously unknown or that provide unexpected insights into poorly understood problems.

In this section, we derive conditions on preferences of the class given by Theorem 2 which obey Strong and Weak versions of the Precautionary Principle. The Strong Form considered here is essentially identical to that considered by Grant and Quiggin (2013b). The Weak Form considered here differs in its formulation from the Modified Form considered by Grant and Quiggin (2013b), which is presented in the context of an extensive form-game and invokes the idea of a 'fallback option.' However, the underlying intuition is the same as for the Weak Form considered here.

Suppose a DM faces the problem of choosing an act from a finite set of available acts $\mathcal{B} \subset \mathcal{F}$. If her choice is governed by a binary relation $\succsim$ represented by a function $V : \mathcal{F} \to \mathbb{R}$, we may express her problem as solving the constrained optimization:

$$\max_{f \in \mathcal{B}} V(f). \tag{1}$$

The precautionary principle is often advocated in settings where not all risks are well understood. In such settings, precaution may be manifested by a heuristic restriction attention to surprise-free choices.

Strong form of the precautionary principle: (SPP)

In choice problems *for which there exists an option not subject to surprises (i.e. $\mathcal{B} \cap \bar{\mathcal{F}}_0 \neq \varnothing$) choose such an option.*

It is useful to consider this principle in relation to the critique offered by Sunstein. Unlike the interpretations of the principle put forward and criticized by Sunstein, SPP does not rule out activities that involve risk, as long as those risks are well defined (in the sense that $f \in \bar{\mathcal{F}}_0$) and the potential consequences are bounded (in our model by the interval $[\ell, m]$). For 'manageable' risks of this kind a standard expected utility analysis may be applied.

We shall consider the following domain of acts which we call the domain of unfavorable surprises. An act in this class has the property that the outcome arising should a surprise occur can be no better than what would have obtained if there was no surprise.

Definition 7 (*Unfavorable Surprises*) The domain of unfavorable surprises is taken to be:

$$\mathcal{F}^{\mathrm{US}} = \{f = (f_0, f_1) \in \mathcal{F} : f_0 \geq f_1\}.$$

Proposition 1 *If $V(.)$ takes the form given in the representation expressed in Theorem 2 (ii), and the choice set is in the domain of unfavorable surprises $\mathcal{B} \subset \mathcal{F}^{\mathrm{US}}$ then* $\mathrm{argmax}_{f \in \mathcal{B}} V(f)$ *conforms to the strong form of the precautionary principle whenever $\mathcal{B} - \mathcal{F}^{uu} \subset \bar{\mathcal{F}}_0$.*

Proof First notice that $\mathcal{B} \cap \mathcal{F}^{uf} = \varnothing$ since $f_0 \geq f_1$ for all $f \in \mathcal{F}^{\mathrm{US}}$. So if $\mathcal{B} \cap \mathcal{F}^{uu} = \varnothing$, then $\mathcal{B} \subset \bar{\mathcal{F}}_0$ and hence $\mathrm{argmax}_{f \in \mathcal{B}} V(f) \subset \bar{\mathcal{F}}_0$, as required. So consider now the case where $\mathcal{B} \cap \mathcal{F}^{uu} \neq \varnothing$. Let $f^0 \in \mathcal{B} \cap \bar{\mathcal{F}}_0 \neq \varnothing$. Applying the expression for the representation in Theorem 2 *(ii)* we have, for any $g \in \mathcal{B} \cap \mathcal{F}^{uu}$,

$$V\left(f^0\right) - V(g) = \bar{V}\left(f^0\right) - \bar{V}(\bar{g}) + V(m) - V(\ell)$$
$$= [V(m) - V(\ell)] - \left[\bar{V}(\bar{g}) - \bar{V}\left(f^0\right)\right]$$
$$> 0, \quad \text{since } [V(m) - V(\ell)] > \left|\bar{V}(\bar{g}) - \bar{V}\left(f^0\right)\right|.$$

As Grant and Quiggin (2013b) note, choices that are too conservative and by precluding all options that are subject to surprises may lead to sub-optimal choices. In particular, suppose that, although an act is subject to surprises, the outcome in the event of a surprise can be bounded below by ℓ, so that the *unforeseen and unfavorable consequence $\ominus$* is precluded. For options subject to surprise, the probability of events contained in the interval $[\ell, m]$ will, in general, be ambiguous. But the representation in Theorem 2 *(ii)* incorporates attitudes to ambiguity, which allow for any degree of ambiguity aversion, up to and including the polar case $u\left(\left[\ell', m'\right]\right) = u\left(\left[\ell', \ell'\right]\right)$. We may define:

Weak form of the precautionary principle: (WPP)

Whenever there exist options without unforeseen consequences (i.e., $\mathcal{B} \cap \bar{\mathcal{F}} \neq \varnothing$) choose such an option.

Proposition 2 *If $V(.)$ takes the form given in the representation expressed in Theorem 2 (ii), and the choice set is in the domain of unfavorable surprises $\mathcal{B} \subset \mathcal{F}^{\mathrm{US}}$ then* $\mathrm{argmax}_{f \in \mathcal{B}} V(f)$ *conforms to the weak form of the precautionary principle whenever $\mathcal{B} \cap \bar{\mathcal{F}} \neq \varnothing$.*

Proof Let $f \in \mathcal{B} \cap \bar{\mathcal{F}} \neq \varnothing$. Applying the expression for the representation in Theorem 2*(ii)* we have, for any $g \in \mathcal{B} \cap \mathcal{F}^{uu}$,

$$V(f) - V(g) = \bar{V}(f) - \bar{V}(\bar{g}) + V(m) - V(\ell)$$
$$= [V(m) - V(\ell)] - \left[\bar{V}(\bar{g}) - \bar{V}(f)\right]$$
$$> 0, \quad \text{since } [V(m) - V(\ell)] > \left|\bar{V}(\bar{g}) - \bar{V}(f)\right|.$$

In the contexts of domains of unfavorable surprises, WPP rules out acts that leave the decision-maker exposed to the unbounded possible losses associated with unforeseen adverse consequences. Any weaker principle would not have this property. Hence, in such domains of unfavorable surprise, it seems sensible to impose the WPP as a heuristic constraint.

4 Concluding comments

Standard models of decision under uncertainty require decision-makers to consider all possible states of nature, and the consequences they yield for the acts under consideration. The starting point is typically a representation theorem showing that given conditions on choices are satisfied, if and only if preferences over all possible acts can be represented by a functional of a particular form. To the extent that the given conditions are considered normatively appealing, the associated choices may be regarded as rationally justified.

In reality, no boundedly rational agent can consider all relevant possibilities, and it is necessary to limit the set of states and consequences that are considered, leaving the unconsidered elements of the problem as potential 'surprises.' It might be hoped that there is some optimal way to choose the trade-off between the risk of surprises and the effort involved in considering additional possibilities. However, this hope is self-contradictory in two ways. First, the problem of choosing an optimal trade-off is a decision problem and is therefore subject to the same problems of bounded rationality, which implies the existence of an infinite regress. Second, and more critically, the only way to determining the optimality of considering additional possibilities is to estimate the consequences of doing so, which in itself requires consideration of the acts in question.

It follows that any choice procedure capable of being implemented by a boundedly rational decision-maker must be imperfect from the perspective of an unboundedly aware outside observer. At most, such procedures may be ecologically rational in the sense of Goldstein and Gigerenzer (2002), Grant and Quiggin (2013a).

Nevertheless, it is possible to define a domain of choices that may be considered by a boundedly rational decision-maker, and to impose conditions on those choices sufficient to yield a representation theorem. In the present paper, we have applied this approach to derive a version of the EUU model of Gul and Pesendorfer with an augmented outcome space to accommodate unforeseen consequences.

Appendix

We list the six axioms presented in Gul and Pesendorfer (2014) adapted to the framework we developed in Sect. 2.

GP 1: The preferences $\succsim$ are complete and transitive.

GP 2: For any pair of acts $f, g \in \bar{\mathcal{F}}$, $f > g \Rightarrow f \succ g$.

GP 3: $y_{E \cap D} x \sim y_{E \cap D'} x$, for any pair of outcomes $x, y \in X$, any ideal event $E \in \mathcal{I}$, and any diffuse events $D, D' \in \mathcal{D}$.

GP 4: For any four outcomes $x > y$ and $w > z$, and any pair of ideal events $E, E' \in \mathcal{I}$ $x_E y \succsim x_{E'} y \Rightarrow w_E z \succsim w_{E'} z$.

GP 5: For any pair of acts $f, g \in \bar{\mathcal{F}}_0$ that are finite-ranged, if $f \succ g$, then there exists a finite partition of ideal events $E_1, \ldots, E_n$ of Ω such that $\ell_{E_i} f \succ m_{E_i} g$ for all i.

GP 6: Fix $g, h, (f^n)_{n=1}^{\infty} \in \bar{\mathcal{F}}$. Suppose $g \succsim f^n \succsim h$ for all n. (*i*) If f^n is finite-ranged for all n and converges pointwise to f then $g \succsim f \succsim h$. (*ii*) If f^n converges uniformly to f then $g \succsim f \succsim h$.

As [Gul and Pesendorfer](2014) note: GP 1 and 2 are standard ordering and monotonicity assumptions; GP 3 embodies the intuition that the decision-maker cannot use (non-null) ideal events to quantify the likelihood of diffuse events; GP 4 is simply Savage's comparative probability axiom (*P4*) applied to ideal events; GP 5 is Savage's small event continuity axiom (*P6*) applied to pairs of acts that are adapted to ideal events, that in conjunction with Axiom 0 ensures that there are no atoms in Σ; and finally, GP 6 ensures both the countable additivity of the prior probability as well as guaranteeing the continuity of the interval utility.

Proof of Theorem 2 **Sufficiency of the axioms.** Axiom 1(i) implies that all acts in $\bar{\mathcal{F}}_0$ are adapted to ideal events. Hence for $\succsim$ restricted to $\bar{\mathcal{F}}_0$, GP 1–6 yields a standard expected utility representation with a countably additive probability measure μ and continuous utility index $v : X \to \mathbb{R}$.

Since by Axiom 1(ii), Ω_1 is diffuse, it follows that for any interval $[x, y] \subset X$ we can set $u(x, y) = v(z)$ where $z \in X$ is such that $y_{\Omega_1} x \sim z$. Axioms GP 2 and 6 ensure that $z \in [x, y]$ exists and therefore that u is well defined.

Fix an act $f \in \bar{\mathcal{F}}$. A greatest lower bound of f that is measurable with respect to Σ is given by: $\lfloor f \rfloor (s, i) = \min(f_0(s), f_1(s))$. Similarly, a least upper bound of f that is measurable with respect to Σ is given by: $\lceil f \rceil (s, i) = \max(f_0(s), f_1(s))$.

Hence, applying the representation result of Gul and Pesendorfer ([2014], Theorem 1, p5), it follows that

$$V(f) = \int u(\lfloor f \rfloor, \lceil f \rceil) \, d\mu$$

$$= \int_S u(\min(f_0(s), f_1(s)), \max(f_0(s), f_1(s))) \, \mu_S(ds),$$

as required. $\qquad \Box$

References

Allais, M. (1953). Le comportement de l'homme rationnel devant le risque: critique des postulats et axiomes de l'école Américaine. *Econometrica*, *21*(4), 503–546.

Ellsberg, D. (1961). Risk, ambiguity and savage axioms. *Quarterly Journal of Economics*, *75*, 643–669.

Gilboa, I., & Schmeidler, D. (1989). Maxmin expected utility with non-uniqe prior. *Journal of Mathematical Economics*, *18*, 141–153.

Goldstein, D., & Gigerenzer, G. (2002). Models of ecological rationality: The recognition heuristic. *Psychological Review*, *109*(1), 75–90.

Gollier, C., Jullien, B., & Treich, N. (2000). Scientific progress and irreversibility: An economic interpretation of the 'Precautionary Principle'. *Journal of Public Economics*, *75*, 229–53.

Grant, S., & Quiggin, J. (2013a). Inductive Reasoning About Unawareness. *Economic Theory*, *54*, 717–755.

Grant, S., & Quiggin, J. (2013b). Bounded awareness, heuristics and the precautionary principle. *Journal of Economic Behavior and Organization*, *93*, 17–31.

Gul, F., & Pesendorfer, W. (2014). Expected uncertain utility theory. *Econometrica*, *82*(1), 1–39.

Halpern, J., & Rêgo, L. (2009). Reasoning about knowledge of unawareness. *Games and Economic Behavior*, *67*(2), 503–25.

Heifetz, A., Meier, M., & Schipper, B. (2006). Interactive unawareness. *Journal of Economic Theory*, *130*(1), 78–94.

Jaffray, J.-Y. (1989). Linear utility theory for belief functions. *Operations Research Letters*, *8*(2), 107–112.

Karni, E. and Vierø M.-L. (2014). 'Awareness of unawareness: A theory of decision-making in the face of ignorance', Queen's Economics Department Working Paper No. 1322. http://ideas.repec.org/p/qed/wpaper/1322.html.

Morgenstern, O., & Von Neumann, J. (1947). *Theory of games and economic behavior* (2nd ed.). Princeton: Princeton University Press.

Li, J. (2008). 'A note on unawareness and zero probability', Penn Institute for Economic Research Working Paper 08–022. http://economics.sas.upenn.edu/pier/working-paper/2008/note-unawareness-and-zero-probability.

Savage, L. J. (1954). *The foundations of statistics. New York (2nd ed.). (1972)*. John Wiley & Sons; New York: Dover Publications.

Schipper, B. (2014). The unawareness bibliography. http://www.econ.ucdavis.edu/faculty/schipper/unaw.htm.

Schmeidler, D. (1989). Subjective probability and expected utility without additivity. *Econometrica*, *57*, 571–587.

Sunstein, C. R. (2003). Beyond the precautionary principle. *University of Pennsylvania Law Review*, *151*(January), 1003–1058.

Walker, O. (2014). "Unawareness with 'Possible' possible worlds", forthcoming Mathematical Social Sciences.

Wingspread Conference. (1998). *Wingspread statement on the precautionary principle*. Racine, Washington, February: Press release.