Measurements of Attitudes toward Unawareness

Edi Karni* — Johns Hopkins University
Marie-Louise Vierø† — Aarhus University

July 10, 2025

Abstract

Decisions under uncertainty may result in new, unanticipated, consequences. Decision makers may be aware of being unaware of possible consequences of their decisions and take it into account when choosing among alternative courses of action. Decision makers' attitudes toward encountering unanticipated consequences is reflected in their choice behavior. This paper proposes, for the first time, measures of the attitudes toward unawareness, thereby filling a lacuna in the literature on decision making under uncertainty and awareness of unawareness.

Keywords: Unawareness, Awareness of Unawareness, Attitude toward Unawareness
JEL classification: D8; D81; D83

*Email: karni@jhu.edu

†Corresponding author. Department of Economics and Business Economics, Aarhus University, Fuglesangs Alle 4, 8210 Aarhus V, Denmark. Email: mlv@econ.au.dk. Vierø gratefully acknowledges financial support from Aarhus University Research Foundation.

1 Introduction

Decisions under uncertainty may result in new, unanticipated, outcomes of whose existence the decision maker was previously unaware, i.e., outcomes that are unknown unknowns. Theories of individual decision making under uncertainty that allow for the possibility of the decision maker being unaware has recently been a topic of much interest. In the context of decision theory, Karni and Vierø (2013) addressed this by expanding the state space and axiomatized a process for updating preferences when awareness grows, dubbed 'reverse Bayesianism'. This approach was further explored and elaborated in Karni and Vierø (2015, 2017), Dominiak and Tserenjigmid (2018, 2022), Karni, Valenzuela-Stookey and Vierø (2021), Chakravarty, Kelsey, and Teitelbaum (2021), and Vierø (2021, 2023).

Karni and Vierø (2017) incorporated the idea that decision makers may anticipate that their universe may expand, even though they do not know in what way. In other words, they may have awareness of their unawareness, and this will affect their choice behavior. For such decision makers, the various models of individual behavior in the face of uncertainty and unawareness incorporate the decision makers' attitudes toward unawareness. However, missing from this literature are *measures* of decision makers' attitudes toward unawareness. Even if two decision makers agree on the degree to which they are aware of their unawareness, they may still exhibit different attitudes towards it. For example, they may exhibit different intensities of awareness aversion.

The main objective of this paper is to fill this lacuna by proposing measures of unawareness attitudes. Analogous to measures of risk aversion, we provide both a utility measure (analogous to the Arrow-Pratt measure of risk aversion) and an unawareness premium (analogous to the risk premium). We formalize what it means for one preference relation to be more unawareness averse than another, thereby allowing for interpersonal comparisons of unawareness attitudes. This allows for tractable analysis of economic problems under awareness of unawareness. We also study some behavioral implications and suggest an experimental design for elicitation of the unawareness premium.

A topical example of the usefulness of the proposed measures are individual decisions of whether to vaccinate against COVID and, if they decide to vaccinate, the choice of which vaccine. Various health consequences of getting COVID and their likelihoods have been discovered since the eruption of the pandemic. However, this is a new virus which raises the fear that additional, possible long term, unanticipated effects are yet to be discovered. At the same time, the vaccines developed are also new. The clinical trials of the new vaccines provide evidence of their effectiveness in preventing infection and possible side effects. However, there is lingering concern that not yet seen, presumably rare, side effects may be discovered. In deciding whether to vaccinate and which vaccine to choose, it is natural to presume that decision makers take into account the possible existence of unanticipated health consequences, and that their attitudes toward such unknown unknowns, as well as their beliefs regarding their likelihoods, affect their choice behavior. For instance, the more unawareness averse is a decision maker, the more likely she is to choose a vaccine that underwent more rigorous clinical trials, or to opt for a vaccine that was produced invoking traditional methods (e.g., injection of weakened COVID virus) when the alternatives are vaccines that were developed by the application of new, untested, techniques.

The study of the behavioral implications of unawareness is also taken up in epistemology and game theory, including Heifetz, Meier, and Schipper, (2006, 2008, 2013), Halpern and Rego (2009, 2013), Piermont (2017), and Grant and Quiggin (2013a,b). Learning under unawareness was explored by Grant, Meneghel, and Tourky (2022). Eichberger and Guerdjikova (2024) propose what may best be described as a case-based model of decision making under ambiguity and unawareness.

A decision-theoretic, non-Bayesian, approach to modeling decision makers' awareness of unawareness of the potential consequences of their actions was recently proposed in Karni (2024). This approach is an adaptation of Ewens (1972) generalization of De Morgan's (1838) formulae of probabilistic predictions of known and unknown outcomes to the context of decision making under uncertainty. Schipper (2022) explores consistency of models of prediction based on partition exchangeability with 'reverse Bayesianism', while Grant and Quiggin (2015) presents an alternative model of growing awareness.

Experimental tests of reverse Bayesianism were conducted by Becker, Melkonyan, Proto, Sofianos, and Trautman (2020). Their study provides support to the 'reverse Bayesianism' hypothesis.

The plan for the rest of the paper is as follows: In the next section we introduce a classification of attitudes toward unawareness. In Section 3 we introduce measures of attitudes toward unawareness that permit interpersonal comparisons of these attitudes. Section 4 discusses behavioral implications of these measures. Concluding remarks and extensions appear in Section 5.

2 Classification of Attitudes toward Unawareness

2.1 Preliminaries

Let C be a finite set of *known consequences*, and assume that $|C| \geq 2$. We assume that, at the point at which a decision must be made, the set of known consequences C is given. Denote by $\hat{x}$ a "place holder" representing the possible existence of new, unanticipated, consequences. Formally, given C, $\hat{x} := \neg C$. Let $\mathcal{C} := C \cup \{\hat{x}\}$ and denote by $\Delta(\mathcal{C})$ the set of probability distributions on $\mathcal{C}$. For the purpose of this paper, $\Delta(\mathcal{C})$ is the *choice set.*

Let $\succcurlyeq$ be a complete and transitive binary relation on $\Delta(\mathcal{C})$, dubbed a *preference relation.* The symbols $\succ$ and $\sim$ denote the asymmetric and symmetric parts of $\succcurlyeq$, respectively. We assume throughout that $\succcurlyeq$ is continuous (in the topology of weak convergence) and monotonic increasing with respect to first-order stochastic dominance (see definition 3 in subsection 3.2).

If the choices are between objective probability distributions on $\mathcal{C}$, the preference relation $\succcurlyeq$ on $\Delta(\mathcal{C})$ is regarded a primitive. If the choices are between acts (i.e., mappings on an underlying state space to $\Delta(\mathcal{C})$), then elements of $\Delta(\mathcal{C})$ are subjective probability distributions, identified with constant acts, and $\succcurlyeq$ on $\Delta(\mathcal{C})$ is the restriction of the preference relation on acts to the subset of constant acts.

Let δ_c denote the Dirac measure on $\mathcal{C}$ concentrated at c. We apply the common abuse of notation of writing $c \succcurlyeq c'$ if $\delta_c \succcurlyeq \delta_{c'}$. Given C and $\succcurlyeq$, let $\bar{c}$ and $\underline{c}$ denote preference-dependent elements of C such that $\bar{c} \succ \underline{c}$ and $\bar{c} \succcurlyeq c \succcurlyeq \underline{c}$ for all $c \in C$. Since C is finite and, by assumption, $\succcurlyeq$ is non-trivial (i.e., $\succ \neq \varnothing$), such best and worst elements of C exist.

A prerequisite for developing measures of *attitudes* toward unawareness is an agreed-upon measure of the *degree* of unawareness. We take this measure to be the probability of discovering unanticipated consequences (i.e., the probability assigned to $\hat{x}$). Depending on the model, this probability may be subjective, as in the 'reverse Bayesianism' model of Karni and Vierø (2017) and in the case-based decision model of Eichberger and Guerdjikova (2024), or objective, as in the formulae of De Morgan (1838) and Ewens (1972).[1] In the former case, the probability of $\hat{x}$ is the subjective probability measure of the event (i.e., a subset of the state space) in which an

[1]See also Karni (2024).

unanticipated consequence may obtain and there is no presumption that different decision makers should agree on the subjective probability assigned to $\hat{x}$. In the latter case, it is one minus the sum of the objective probabilities of the consequences in C. In this case there is an agreement on the value of the probability assigned to $\hat{x}$, which is the measure of the degree of unawareness. In either case, the measures of the attitudes toward unawareness we propose are independent of the specific interpretation, and *do not require that decision makers agree on the probability of* $\hat{x}$.

For each $z_p \in [0,1]$, define

$$\Delta_{z_p}(C) = \{p \in \Delta(\mathcal{C}) \mid \sum_{c \in C} p(c) = z_p\}. \tag{1}$$

Then, $\Delta_{z_p}(C)$ for $z_p \in [0,1)$ are sets of 'defective' probability distributions on C. The elements of $\Delta_{z_p}(C)$ are the distributions in $\Delta(\mathcal{C})$ that assign the same probability, $1 - z_p$, to the unknown outcome, and each corresponds to an element of $\Delta(\mathcal{C})$. Note that $\Delta_1(C)$ is the set of all probability distributions on C, i.e. that only have known consequences in their support. Note also that $\cup_{z_p \in [0,1]} \Delta_{z_p}(C) = \Delta(\mathcal{C})$.

Broadly speaking, a decision maker's attitudes toward unawareness depends on his attitude toward the possibility of encountering an unanticipated consequence as well as on his valuation of the defective probability distributions in $\Delta_{z_p}(C)$. As we note below, the latter aspect incorporates the decision maker's ordinal preferences on C (i.e., his ranking of the elements of C) and his risk attitude.

2.2 Local attitudes toward unawareness

As was mentioned above, decision makers' attitudes toward unawareness depend on the level of unawareness, their valuations of the consequences in C, and their risk attitudes. In other words, decision makers' attitudes toward unawareness reflect their valuations of unanticipated consequences whose nature is, by definition, unknown relative to their valuations of the distribution of the known consequences. In particular, facing a choice between to alternatives, $p, p' \in \Delta(\mathcal{C})$ such that $z_p = z_{p'}$, (i.e., alternatives that involve the same probability of encountering unanticipated consequences) a decision maker may be less inclined to choose the alternative that assigns larger probabilities to less desirable known consequences. Likewise, if a decision maker is faced with the prospects of bad known consequences, he will be more positive towards the unknown than if he is faced with good known consequences. For example, a patient may be more inclined to try an experimental treatment after having learned that the conventional treatment does not work, than he was before. Hence, unawareness attitude must be defined locally.

For each $q \in \Delta_1(C)$, define an equivalence class, P_q, of distributions in $\Delta(\mathcal{C})$ that have the same likelihood ratios of all pairs of known consequences. Formally,

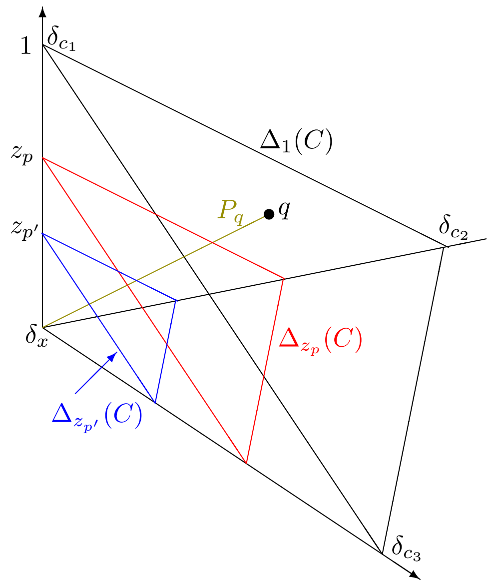
$$P_q = \{p \in \Delta(C) \mid \frac{p(c)}{p(c')} = \frac{q(c)}{q(c')},\ \forall c, c' \in C\}. \tag{2}$$

Equivalently, $P_q = \{\alpha q + (1-\alpha)\delta_{\hat{x}} \mid \alpha \in [0,1]\}$. The equivalence class fixes the quality of the known consequences.[2]

To illustrate diagrammatically an equivalence class P_q, suppose that $C = \{c_1, c_2, c_3\}$. For this example, the sets $\Delta_1(C)$, $\Delta_{z_p}(C), \Delta_{z_{p'}}(C)$ and P_q are depicted in Figure 1 below.

[2]We can view the equivalence class as randomization between an urn with fixed proportions of known colors of

Figure 1: Illustration of defective simplexes and an equivalence class P_q



Definition 1 *Given C and $q \in \Delta_1(C)$, the preference relation $\succcurlyeq$ on $\Delta(\mathcal{C})$*

- *displays unawareness aversion at P_q if $z_p \geq z_{p'}$ implies that $p \succcurlyeq p'$, for all $p, p' \in P_q$;*
- *displays unawareness proclivity at P_q if $z_p \geq z_{p'}$ implies that $p' \succcurlyeq p$, for all $p, p' \in P_q$;*
- *displays unawareness neutrality at P_q if $p' \sim p$, for all $p, p' \in P_q$.*

According to Definition 1, the preference relation displays unawareness aversion (proclivity) at P_q if, within the equivalence class, the decision maker prefers a lower (higher) probability of the unknown. It displays unawareness neutrality at P_q if, within the equivalence class, the decision maker is indifferent between different probabilities of the unknown.

2.3 Global attitudes toward unawareness

The measure of unawareness attitudes is local in the sense that it is defined for P_q. A preference relation displaying the same local attitudes at P_q for all $q \in \Delta_1(C)$, is said to display these attitudes globally.

Definition 2 *The preference relation $\succcurlyeq$ on $\Delta(\mathcal{C})$*

- *displays unawareness aversion if it displays unawareness aversion at P_q for all $q \in \Delta_1(C)$;*
- *displays unawareness proclivity if it displays unawareness proclivity at P_q for all $q \in \Delta_1(C)$;*

balls and the unknown consequence, with varying randomization probability. Later, in Section 4, we will discuss an experiemental design based on this idea.

- *displays unawareness neutrality if it displays unawareness neutrality at P_q for all $q \in \Delta_1(C)$.*

To characterize global unawareness attitudes, we make use of the fact that at any point the set, C, of known consequences is finite and thus best and worst consequences . $\bar{c}$ and $\underline{c}$ exist. Clearly, $u(\bar{c}) \geq \bar{U}_q \geq u(\underline{c})$ for all $q \in \Delta_1(C)$. Therefore, invoking our measure of local unawareness attitudes, we immediately have:

Proposition 1 *A preference relation $\succcurlyeq$ on $\Delta(\mathcal{C})$ displays strict unawareness aversion if and only if $\delta_{\underline{c}} \succ \delta_{\hat{x}}$, and it displays strict unawareness proclivity if and only if $\delta_{\hat{x}} \succ \delta_{\bar{c}}$.*

Proof: If $\delta_{\hat{x}} \succsim \delta_{\underline{c}}$, then $\succsim$ does not display strict unawareness aversion at $P_{\delta_{\underline{c}}}$. Hence the property is not global. For the converse, note that the global property implies the local property at $P_{\delta_{\underline{c}}}$. The argument for strict unawareness proclivity is analogous. ■

3 Measures of Attitudes toward Unawareness

3.1 Utility characterizations

To get measures of unawareness attitude analogous to the Arrow-Pratt measure and the risk premium, we need more structure on preferences. Henceforth, we assume that the preference relation has an expected utility representation, generalized to accommodate awareness of unawareness as in Karni and Vierø (2017).[3] Then, for all $p, p' \in \Delta(\mathcal{C})$, $p \succcurlyeq p'$ if and only if

$$\sum_{c \in C} u(c)p(c) + (1 - z_p)u(\hat{x}) \geq \sum_{c \in C} u(c)p'(c) + (1 - z_{p'})u(\hat{x}), \tag{3}$$

where the function $u : \mathcal{C} \to \mathbb{R}$ is unique up to positive affine transformation. The representation generalizes expected utility to lotteries that include the unknown consequence in their support by assigning a utility index $u(\hat{x})$ to the unknown. Thus, the representation has an extra parameter compared to a standard expected utility representation. Without loss of generality, we normalize u as follows: $u(\bar{c}) = 1$, $u(\underline{c}) = 0$.

Expression (3), is equivalent to

$$z_p \frac{\sum_{c \in C} u(c)p(c)}{z_p} + (1 - z_p)u(\hat{x}) \geq z_{p'} \frac{\sum_{c \in C} u(c)p'(c)}{z_{p'}} + (1 - z_{p'})u(\hat{x}). \tag{4}$$

For all $q \in \Delta_1(C)$, define

$$\bar{U}_q := \sum_{c \in C} u(c)q(c) \tag{5}$$

and note that for all $p, p' \in P_q$,

$$\frac{\sum_{c \in C} u(c)p(c)}{z_p} = \frac{\sum_{c \in C} u(c)p'(c)}{z_{p'}} = \bar{U}_q. \tag{6}$$

Thus, (4) is equivalent to

$$(z_p - z_{p'})(\bar{U}_q - u(\hat{x})) \geq 0. \tag{7}$$

[3] For an axiomatic characterization, see Karni and Vierø (2013, 2017).

Dividing through by $\bar{U}_q$, which is positive given the transformation of u, we obtain that

$$p \succcurlyeq p' \Leftrightarrow (z_p - z_{p'})(1 - \frac{u(\hat{x})}{\bar{U}_q}) \geq 0. \quad (8)$$

In particular, since $z_q = 1$, the statement in (8) implies that, for all $q \in \Delta_1(C)$ and $p' \in P_q$, it holds that $q \succcurlyeq p'$ if and only if $(1 - z_{p'})(1 - \frac{u(\hat{x})}{\bar{U}_q}) \geq 0$.

Without loss of generality, suppose that $z_p - z_{p'} > 0$ and (8) holds. If the expression on the right-hand side of (8) holds with strict inequality then the preference relation displays strict unawareness aversion at P_q, and if it holds with equality it implies unawareness neutrality at P_q. If (8) does not hold, the preference relation displays strict unawareness proclivity at P_q.

Define the utility measure of unawareness attitude displayed by $\succcurlyeq$ at P_q by

$$\phi_q^u := u(\hat{x})/\bar{U}_q. \quad (9)$$

We have that $\succcurlyeq$ displays

- strict unawareness aversion at P_q if and only if $u(\hat{x})/\bar{U}_q < 1$;
- unawareness neutrality at P_q, if and only if $u(\hat{x})/\bar{U}_q = 1$;
- strict unawareness proclivity at P_q if and only if $u(\hat{x})/\bar{U}_q > 1$.

Note that the measure is invariant over the class of positive affine transformations of the utility function, as long as $u(\underline{c}) \geq 0$.[4] The measure's dependence on $\bar{U}_q$ implies that it depends on the support, C, of the distributions being compared and on the risk attitude exhibited by the underlying preference relation. Consequently, the preference relation's attitude toward unawareness is affected by its evaluations of the known consequences and the risk attitude it displays.

Note that, under the normalization $u(\bar{c}) = 1$, $u(\underline{c}) = 0$, $\succcurlyeq$ on $\Delta(\mathcal{C})$ displays strict unawareness aversion (proclivity) if and only if $u(\hat{x})/\bar{U}_q < (>)\, 1$ for all $q \in \Delta_1(C)$. If the preference relation $\succcurlyeq$ on $\Delta(\mathcal{C})$ is non-trivial (i.e., $\succ$ is non-empty), then it cannot display global unawareness neutrality.

3.2 Probability premium

A natural measure of attitude toward unawareness is the sacrifice a decision maker is willing to make to reduce her unawareness exposure. We refer to such measure, expressed in probability terms, as a *probability premium.* There are numerous ways of measuring "the probability sacrifice." Here we invoke one such concept based on the notion of stochastic dominance.[5]

Because the set of consequences is unstructured, we make the following definition of first-order stochastic dominance. Without essential loss of generality, suppose that the elements of C are strictly ranked (i.e., the indifference relation on C is empty).

Definition 3 *For all $p, p' \in \Delta_1(C)$, p first-order stochastically dominates p' if $\Sigma_{\{c \succcurlyeq c'\}}[p'(c') - p(c')] \geq 0$, for all $c \in C$.*

[4]Positive linear transformations of u that change the signs of $\bar{U}_q$ and/or $u(\hat{x})$ would require changing the conditions. For example, if $\bar{U}_q < 0$, by (8), $\succcurlyeq$ displays strict unawareness aversion at P_q if and only if $u(\hat{x})/\bar{U}_q > 1$.

[5]Another concept is described in the concluding section.

With this definition in mind, define

$$\mathcal{P} := \{\eta(t) = t\delta_{\bar{c}} + (1-t)\,\delta_{\underline{c}} \mid t \in \mathbb{R}\}.$$

Then, $\mathcal{P}$ is a linear path in the space $\mathbb{R}^{|C|}$. The intersection of $\mathcal{P}$ with $\Delta(\mathcal{C})$ is the subset of $\mathcal{P}$ where $t \in [0,1]$. It consists of convex combinations of the best and worst Dirac measures on C, such that $\eta(0) = \delta_{\underline{c}}$ and $\eta(1) = \delta_{\bar{c}}$. Clearly, $\eta(t)$ first-order stochastically dominates $\eta(t')$, for all $t > t'$.

Given $\succcurlyeq$ on $\Delta(\mathcal{C})$, define $t^{\succcurlyeq}(p)$ by $\eta\left(t^{\succcurlyeq}(p)\right) \sim p$ for those $p \in \Delta(\mathcal{C})$ for which such indifference exists. Thus, $t^{\succcurlyeq}(p)$ is the weight we have to put on the best Dirac measure to get a convex combination that is indifferent to p. The continuity and monotonicity of $\succcurlyeq$ with respect to first-order stochastic dominance implies that $t^{\succcurlyeq}(p)$ is well defined for all $p \in \Delta_1(C)$. Note that if $\delta_{\bar{c}} \succ \delta_{\hat{x}} \succ \delta_{\underline{c}}$, then $t^{\succcurlyeq}(p) \in [0,1]$ and is well-defined for all $p \in \Delta(\mathcal{C})$. If $\delta_{\hat{x}} \succcurlyeq \delta_{\bar{c}}$ or $\delta_{\underline{c}} \succcurlyeq \delta_{\hat{x}}$, then $t^{\succcurlyeq}(p) \in [0,1]$ for p such that z_p is sufficiently close to 1.

If $p \succ \delta_{\bar{c}}$ (which may be the case if $\delta_{\hat{x}} \succ \delta_{\bar{c}}$), then continuity and monotonicity of $\succcurlyeq$ with respect to first-order stochastic dominance implies that $\delta_{\bar{c}} \sim \alpha p + (1-\alpha)\delta_{\underline{c}}$ for some $\alpha \in (0,1)$. In this case, define $t^{\succcurlyeq}(p) = \frac{1}{\alpha}$.

If $\delta_{\underline{c}} \succ p$ (which may be the case if $\delta_{\underline{c}} \succ \delta_{\hat{x}}$), then continuity and monotonicity of $\succcurlyeq$ with respect to first-order stochastic dominance implies that $\delta_{\underline{c}} \sim \alpha\delta_{\bar{c}} + (1-\alpha)p$ for some $\alpha \in (0,1)$. In this case, define $t^{\succcurlyeq}(p) = \frac{\alpha}{\alpha-1}$.

Invoking the fact that $z_q = 1$, for each $p \in P_q \setminus \{q\}$ we define $\pi^{\succcurlyeq}(P_q, p)$ by the equation

$$\pi^{\succcurlyeq}(P_q; p) := \frac{t^{\succcurlyeq}(q) - t^{\succcurlyeq}(p)}{1 - z_p}. \tag{10}$$

Proposition 2 asserts that $\pi^{\succcurlyeq}(P_q; p)$ is independent of p.

Proposition 2 *For all $q \in \Delta_1(C)$ and $p, p' \in P_q \setminus \{q\}$, $\pi^{\succcurlyeq}(P_q; p) = \pi^{\succcurlyeq}(P_q; p') := \pi^{\succcurlyeq}(P_q)$.*

By Proposition 2, $\pi^{\succcurlyeq}(P_q; p)$ is constant within each equivalence class P_q. Thus, it depends on P_q but is independent of p and, therefore, of the degree, $1 - z_p$, of unawareness.

Proof of Proposition 2: Given any $q \in \Delta_1(C)$, we need to show that

$$\frac{t^{\succcurlyeq}(q) - t^{\succcurlyeq}(p)}{1 - z_p} = \frac{t^{\succcurlyeq}(q) - t^{\succcurlyeq}(p')}{1 - z_{p'}}.$$

for all $p, p' \in P_q$. Without loss of generality, suppose that $z_{p'} < z_p$.

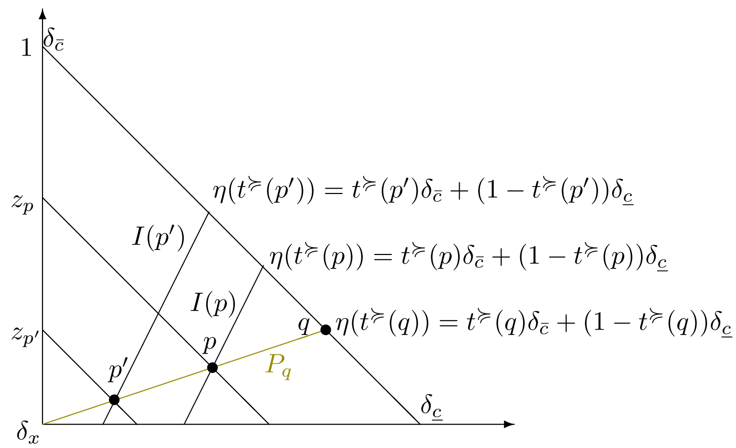
The choice set $\Delta(\mathcal{C})$ is a $(|\mathcal{C}| - 1)$ - dimensional simplex. Given an expected utility preference relation $\succcurlyeq$ on $\Delta(\mathcal{C})$, let $I(p) := \{p' \in \Delta(\mathcal{C}) \mid p' \sim p\}$ denote the indifference class of $p \in \Delta(\mathcal{C})$. For every $p \in \Delta(\mathcal{C})$, $\eta(t^{\succcurlyeq}(p)) = \mathcal{P} \cap I(p)$.

The indifference classes corresponding to $\succcurlyeq$ are the intersections of $\Delta(\mathcal{C})$ with $(|\mathcal{C}| - 2)$ - dimensional hyperplanes that are parallel (i.e., the intersections of the hyperplanes $I(p)$, $p \in \Delta(\mathcal{C})$ with the faces of the simplex are parallel lines). (See depiction in Figure 2).

Now, $P_q = \{\alpha q + (1-\alpha)\delta_{\hat{x}} \mid \alpha \in [0,1]\}$ is a linear path in the simplex. Consider the triangle T in $\Delta(\mathcal{C})$ defined by vertices $\delta_{\bar{c}}, \delta_{\hat{x}}$, and $\eta(t^{\succcurlyeq}(q))$. By definition, T is a two-dimensional object. Since the intersections $I(p) \cap T$ and $I(p') \cap T$ are parallel lines and $z_q = 1$, it holds that

$$\frac{t^{\succcurlyeq}(q) - t^{\succcurlyeq}(p)}{1 - z_p} = \frac{t^{\succcurlyeq}(q) - t^{\succcurlyeq}(p')}{1 - z_{p'}}.$$

Figure 2: Illustration for the proof of Proposition 2.



Thus, by the definition in equation (10), $\pi^{\succcurlyeq}(P_q; p) = \pi^{\succcurlyeq}(P_q; p')$ for all $p, p' \in P_q \setminus \{q\}$. ■

Given the invariance result in Proposition 2, it is meaningful to define the probability premium at P_q, as in Definition 4 below.

Definition 4 *For all $q \in \Delta_1(C)$, $\pi^{\succcurlyeq}(P_q)$ defined in* (10) *is the probability premium at P_q corresponding to $\succcurlyeq$.*

The expression $\pi^{\succcurlyeq}(P_q)$ has the interpretation of probability premium per "unit of unawareness". The probability premium measures how much probability the decision maker would be willing to shift from the best to the worst known consequence, per unit probability of the unknown consequence, in order to avoid an unknown consequence. It represents the cost, expressed in probability terms, an unawareness averse decision maker stands ready to bear in order to reduce her exposure to unawareness.

Let $z_p - z_{p'} > 0$ then, by monotonicity with respect to first order stochastic dominance, we have the following implications:

a. $p \succ p'$ and $\pi^{\succcurlyeq}(P_q) > 0$ if and only if the preference relation displays unawareness aversion at P_q.

b. $p' \succ p$ and $\pi^{\succcurlyeq}(P_q) < 0$ if and only if the preference relation displays unawareness proclivity at P_q.

c. $p \sim p'$ and $\pi^{\succcurlyeq}(P_q) = 0$, if and only if the preference relation displays unawareness neutrality at P_q.

Figure 2 depicts a situation where the preference relation displays unawareness proclivity at P_q. This is the case when the ray corresponding to the equivalence class is flatter than the indifference curves. At equivalence classes where the ray is steeper than the indifference curves, the preference relation displays unawareness aversion. It displays unawareness neutrality at the ray that is parallel to the indifference curves.

3.3 Interpersonal comparisons of attitudes toward unawareness

To compare decision makers' attitudes toward unawareness we assume that the preference relations being compared are ordinally equivalent. Formally, they agree on the ranking of the consequences in C.[6] Moreover, to compare decision makers' local attitudes toward unawareness we restrict the comparisons to their preferences among probability distributions belonging to the same equivalence class P_q. We say that one decision maker displays greater unawareness aversion than another if, for any pair of distributions p and p' belonging to the same equivalence class $P_q \subset \Delta(\mathcal{C})$, such that $z_p \geq z_{p'}$, if the latter individual prefers p over p' so does the former. In words, if the latter individual prefers a lower degree of unawareness (captured by the higher probability z_p of known consequences), so does the former individual.[7] Formally,

Definition 5

a. $\succcurlyeq^u$ *displays greater unawareness aversion at* P_q *than* $\succcurlyeq^v$ *if, for all* $p, p' \in P_q$ *such that* $z_p > z_{p'}$, *it holds that* $p \succcurlyeq^v p'$ *implies* $p \succ^u p'$.

b. $\succcurlyeq^u$ *displays greater unawareness aversion than* $\succcurlyeq^v$ *if it displays greater unawareness aversion at* P_q *for all* $q \in \Delta_1(C)$.

If the preference relations have generalized expected utility representation as in (3), then when $\succcurlyeq^u$ displays greater unawareness aversion than $\succcurlyeq^v$, it implies that a mean-utility-preserving increase in unawareness from the viewpoint of v implies a mean-utility-reducing increase in unawareness from the viewpoint of u. That is, a move along an indifference curve for $\succcurlyeq^v$ in the direction that increases the probability of the unknown would reduce the utility for $\succcurlyeq^u$.

The following theorem characterizes the interpersonal attitudes toward unawareness. Recall that $\phi_q^u := u(\hat{x})/\bar{U}_q$ and, similarly, $\phi_q^v := v(\hat{x})/\bar{V}_q$.

Theorem 1 *Let* $\succcurlyeq^u$ *and* $\succcurlyeq^v$ *be preference relations on* $\Delta(\mathcal{C})$ *that admit the generalized expected utility representations in (3). If* $\delta_{\bar{c}} \succ^j \delta_{\hat{x}} \succ^j \delta_{\underline{c}}$, $j \in \{u, v\}$ *then the following conditions are equivalent:*

(i) $\succcurlyeq^u$ *displays greater unawareness aversion than* $\succcurlyeq^v$.

(ii) $\phi_q^u < \phi_q^v$, *for all* $q \in \Delta_1(C)$.

(iii) $\pi^{\succcurlyeq^u}(P_q) > \pi^{\succcurlyeq^v}(P_q)$, *for all* $q \in \Delta_1(C)$.

[6] This assumption is analogous to the requirement, in the theory of multivariate risk aversion, that the underlying ordinal preferences of the relations being compared are the same (see Kihlstrom and Mirman [1974]). Similarly, the restriction, in the theory of risk aversion with state-dependent preference, that the preferences being compared have the same reference set (see Karni [1985]).

[7] This is analogous to saying that one individual displays greater risk aversion than another if any risk that is acceptable to the former is acceptable to the latter.

Proof. $(ii) \Rightarrow (i)$. Fix P_q and suppose that (ii) holds. Let $p, p' \in P_q$ such that $z_p > z_{p'}$ and $p \succcurlyeq^v p'$. By the representation (3), $p \succcurlyeq^v p'$ if and only if

$$z_p \bar{V}(p) + (1 - z_p) v(\hat{x}) \geq z_{p'} \bar{V}(p') + (1 - z_{p'}) v(\hat{x}), \tag{11}$$

where $\bar{V}(p) := \sum_{c \in C} v(c) p(c)/z_p$, for all $p \in \Delta(\mathcal{C})$. But $\bar{V}(p) = \bar{V}(p') = \bar{V}_q$. Hence, the expressions in (11) are equivalent to,

$$z_p \left(1 - \frac{v(\hat{x})}{\bar{V}_q}\right) \geq z_{p'} \left(1 - \frac{v(\hat{x})}{\bar{V}_q}\right) \tag{12}$$

Taking the difference we get

$$(z_p - z_{p'}) \left(1 - \frac{v(\hat{x})}{\bar{V}_q}\right) \geq 0. \tag{13}$$

But $z_p - z_{p'} > 0$. Thus $p \succcurlyeq^v p'$ if and only if

$$1 - \frac{v(\hat{x})}{\bar{V}_q} \geq 0. \tag{14}$$

By (ii), $u(\hat{x}) / \bar{U}_q < v(\hat{x}) / \bar{V}_q$. Thus

$$1 - \frac{u(\hat{x})}{\bar{U}_q} > 0. \tag{15}$$

By the same argument as above, since $\bar{U}(p) = \bar{U}(p') = U_q$. Thus,

$$p \succ^u p' \Leftrightarrow (z_p - z_{p'}) \left(1 - \frac{u(\hat{x})}{\bar{U}_q}\right) > 0. \tag{16}$$

Hence, (13) implies (16). Consequently, $p \succcurlyeq^v p'$ implies that $p \succ^u p'$. This completes the proof that (ii) implies (i).

$(i) \Rightarrow (ii)$. To prove this part we show the contraposition "not (ii)" implies "not (i)". Let $\hat{q} \in \Delta_1(C)$ and $p, p' \in P_{\hat{q}}$ be such that $z_p > z_{p'}$, and $p \sim^u p'$ and suppose that $v(\hat{x}) / \bar{V}_{\hat{q}} \leq u(\hat{x}) / \bar{U}_{\hat{q}}$. That such $\hat{q}$ exists is implied by the assumption that $\delta_{\bar{c}} \succ^u \delta_{\hat{x}} \succ^u \delta_{\underline{c}}$.

By (16) $p \sim^u p'$ if and only if $\left(1 - u(\hat{x}) / \bar{U}_{\hat{q}}\right) = 0$, by the supposition, $\left(1 - v(\hat{x}) / \bar{V}_{\hat{q}}\right) \geq 0$. By (13), $\left(1 - v(\hat{x}) / \bar{V}_{\hat{q}}\right) \geq 0$ if and only if $p \succcurlyeq^v p'$. Hence, for $p, p' \in P_{\hat{q}}$, $p \succcurlyeq^v p'$ does not imply $p \succ^u p'$. A contradiction. Thus, "not (i)".

$(i) \Rightarrow (iii)$. We show the contraposition "not (iii)" implies "not (i)." Suppose that $p, p' \in P_{\hat{q}}$ such that $z_p > z_{p'}$, $p \sim^u p'$ and $\pi^{\succcurlyeq^v}(P_{\hat{q}}) \geq \pi^{\succcurlyeq^u}(P_{\hat{q}})$. That such $\hat{q}$ exists is implied by the assumption that $\delta_{\bar{c}} \succ \delta_{\hat{x}} \succ \delta_{\underline{c}}$. But $p \sim^u p'$ implies that $t^{\succcurlyeq^u}(p) - t^{\succcurlyeq^u}(p') = 0$. Since $z_p - z_{p'} > 0$, by definition, $\pi^{\succcurlyeq^u}(P_{\hat{q}}) = 0$. By the supposition, $\pi^{\succcurlyeq^v}(P_{\hat{q}}) \geq 0$. Hence, $t^{\succcurlyeq^v}(p) - t^{\succcurlyeq^v}(p) = \left(z_p - z'_p\right) \pi^{\succcurlyeq^v}(P_{\hat{q}}) \geq 0$. Thus, $t^{\succcurlyeq^v}(p) \geq t^{\succcurlyeq^v}(p')$, which holds if and only if $p \succcurlyeq^v p'$. Consequently, $p \succcurlyeq^v p'$ does not imply $p \succ^u p'$, a contradiction.

$(iii) \Rightarrow (i)$. Let $p, p' \in P_q$ such that $z_p > z_{p'}$. By definition, $p \succcurlyeq^v p'$ if and only if $t^{\succcurlyeq^v}(p) - t^{\succcurlyeq^v}(p') = (z_p - z_{p'}) \pi^{\succcurlyeq^v}(P_q) \geq 0$. But $(z_p - z_{p'}) > 0$ implies that $\pi^{\succcurlyeq^v}(P_q) \geq 0$. By (iii), $\pi^{\succcurlyeq^u}(P_q) > 0$. Hence, by definition, $t^{\succcurlyeq^u}(p) - t^{\succcurlyeq^u}(p') = (z_p - z_{p'}) \pi^{\succcurlyeq^u}(P_q) > 0$. But, $t^{\succcurlyeq^u}(p) - t^{\succcurlyeq^u}(p') > 0$ if and only if $p \succ^u p'$. Thus, $\succcurlyeq^u$ displays greater unawareness aversion than $\succcurlyeq^v$. ■

Remark: It is noteworthy that if $\delta_{\underline{c}} \succ^v \delta_{\hat{x}}$ then condition (ii) is sufficient, but not necessary for $\succcurlyeq^u$ displaying greater unawareness aversion than $\succcurlyeq^v$. To grasp this, it suffices to note that there is not $q \in \Delta_1(C)$ such that $q \sim^v \delta_{\hat{x}}$. Hence, for all $p, p' \in P_q$ if $p \succcurlyeq^v p'$ and $z_p > z_{p'}$ it must be that $p \succ^v p'$. Thus, by the same argument as in the proof that $(ii) \Rightarrow (i)$ it holds that

$$1 - \frac{v(\hat{x})}{\bar{V}_q} > 0$$

But (i) implies that $p \succ^u p'$. Hence,

$$1 - \frac{u(\hat{x})}{\bar{U}_q} > 0.$$

These inequalities do not imply that $u(\hat{x}) / \bar{U}_q < v(\hat{x}) / \bar{V}_q$. Hence, (ii) is not necessary.

A similar observation applies to the case in which $\delta_{\hat{x}} \succ^v \delta_{\bar{c}}$.

4 Behavioral implications

The following example illustrates the choice behavior implications of attitudes toward unawareness. It also suggests an experimental design.

Consider the following decision problem. Let there be two urns B_1 and B_2. Each urn contains red and blue balls and one black ball. The proportion of red and blue is the same in B_1 and B_2, but B_2 contains twice as many red and blue balls as B_1. A ball will be drawn from each of the urns. A blue ball pays \$5, a red ball pays \$10. The payoff to a black ball is unspecified. The decision maker's problem is to choose which urn to bet on.

It is obvious that the level of unawareness (i.e., the probability of drawing the black ball whose payoff the decision maker is unaware of) is smaller for B_2 than it is for B_1. Moreover, the likelihood ratio of the red and blue balls is the same in the two urns, so they belong to the same equivalence class, P_q. A decision maker displaying unawareness aversion would choose B_2 over B_1 and, in general, display monotonically decreasing preferences with respect to the probability of drawing the black ball.

If the decision maker is not globally averse or inclined toward unawareness, then $u(\$10) > u(\hat{x}) > u(\$5)$. Then there exists $q = \alpha\delta_{\$10} + (1-\alpha)\delta_{\$5} \in \Delta(\{\$10, \$5\})$ such that the decision maker is locally unawareness neutral at P_q. For $q' = \alpha'\delta_{\$10} + (1-\alpha')\delta_{\$5}$ she will display unawareness aversion at $P_{q'}$, for every $\alpha' < \alpha$, and unawareness proclivity at $P_{q'}$ for all $\alpha' > \alpha$.

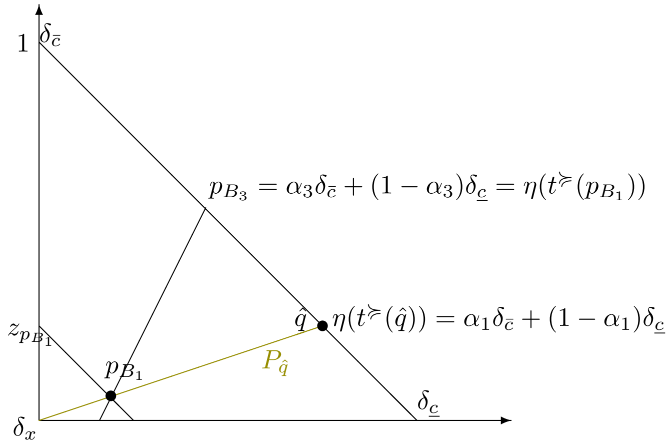
To find the q at which a decision maker is unawareness neutral, an experimenter could expose the decision maker to a series of choices between urns B_1 and B_2, where the proportion of red vs. blue balls is gradually changed in the two urns in a way that always keeps the proportion of red vs. blue fixed between the urns. When the decision maker is indifferent between the two urns, that proportion gives us his point of unawareness neutrality.

According to our definitions, if one decision maker displays greater unawareness aversion than another, then if the latter prefers to bet on B_2 over betting on B_1, so does the former.

What follows is an outline of a procedure designed to elicit a decision maker's local probability premium he is willing to pay to avoid being exposed to unanticipated outcomes. Let z_{B_1} denote the probability of drawing a ball of known color from urn B_1, and let α_1 denote the conditional (on the set of known colors) probability of red balls in B_1. Hence, urn B_1 corresponds to the lottery $p_{B_1} = (z_{B_1}\alpha_1, z_{B_1}(1-\alpha_1), 1 - z_{B_1}) \in P_{\hat{q}}$, where $\hat{q} = \alpha_1\delta_{\$10} + (1-\alpha_1)\delta_{\$5}$. Note that $t^{\succcurlyeq}(\hat{q}) = \alpha_1$.

Consider an urn B_3, which contains only red and blue balls. Then the probability of drawing a known ball from B_3 is $z_{B_3} = 1 > z_{B_1}$. Let the probability of red balls in B_3 be denoted α_3. Thus, urn B_3 corresponds to the lottery $p_{B_3} = (\alpha_3, 1-\alpha_3, 0) \in \Delta_1(\mathcal{C})$.

Figure 3: Illustration of urn experiment



Suppose that α_3 is such that the decision maker is indifferent between betting on B_3 and B_1. That is, $p_{B_1} \sim (\alpha_3, 1-\alpha_3, 0)$. Hence, $t^{\succcurlyeq}(p_{B_1}) = \alpha_3$. Then the local measure of the decision maker's attitude toward unawareness for urn B_1 is given by the probability premium $\pi(P_{\hat{q}}) = (\alpha_1 - \alpha_3) / (1 - z_{B_1})$. This is the probability premium the decision maker is willing to sacrifice to avoid the unawareness associated with B_1. Figure 3 provides an illustration of the situation.

5 Concluding Remarks

This paper introduces measures of decision makers' attitudes toward uncertain prospects, some of whose consequences are unknown unknowns. We have shown that these measures capture decision makers' dislike of facing unanticipated consequences and permits interpersonal comparisons of their attitudes toward unawareness.

An alternative idea of an unawareness probability premium is as follows: Fix $q \in \Delta_1(C)$ and consider $p, p' \in P_q$, such that $z_p > z_{p'}$. Then, by definition, the degree of unawareness associated with p' exceeds that of p. Arrange the elements of C in descending order, $\delta_{\bar{c}} = \delta_{c^1} \succcurlyeq \delta_{c^2} \succcurlyeq, ..., \succcurlyeq \delta_{c^n} = \delta_{\underline{c}}$. Then we can write the distributions p and p' as $p = (p_1, p_2, ..., p_n, 1 - z_p)$ and $p' = (p'_1, p'_2, ..., p'_n, 1 - z_{p'})$, where the last coordinate of each vector is the probability of the unknown consequence.

If $p \succ p'$, transfer probability mass ε from the top to the bottom, from δ_{c^1} to δ_{c^n} (i.e., $p_1 - \varepsilon$ to $p_n + \varepsilon$). If the probability mass p_1 is exhausted, and $\widetilde{p}_1 = (0, p_2, ..., p_n + p_1, 1 - z_p) \succ p'$ then shift probability mass ε from p_2 to p_n (i.e., $p_2 - \varepsilon$ to $p_n + p_1 + \varepsilon$). Continue the process until the point at which the resulting probability distribution, say $\widetilde{p}_k = (0, 0, ..., 0, p_{k+1} - \varepsilon, p_n + p_1 + p_2 + ... + p_k + \varepsilon, 1 - z_p)$, satisfies $\widetilde{p}_k \sim p'$. This process defines a path $\mathcal{H}$ in the space $\Delta(\mathcal{C})$ between p and p'. Let $\hat{p}_n = \Sigma_{i=1}^{k} p_i + \varepsilon$. The difference $\hat{p}_n - p_n$ measures the "sacrifice" the decision maker is willing to make to reduce the unawareness by the measure $z_p - z_{p'}$.[8]

Define $\rho^{\succcurlyeq}(P_q)$ by the equation

$$(z_p - z_{p'})\, \rho^{\succcurlyeq}(P_q) = \hat{p}_n - p_n. \tag{17}$$

Then, $\rho^{\succcurlyeq}(P_q)$ is a probability premium and, invoking the arguments in the proof of Theorem 1, it is easy to verify that a preference relation $\succcurlyeq^u$ displays greater unawareness aversion at P_q than $\succcurlyeq^v$ if and only if $\rho^{\succcurlyeq^u}(P_q) \geq \rho^{\succcurlyeq^v}(P_q)$.

The decision model is general in the sense that the set of consequences is unstructured. Therefore, our unawareness premium, that is, the "price" an individual is willing to pay for the opportunity to choose prospects that entail lower levels of unawareness, is measured in terms of "probability sacrifice."

Our measures of unawareness attitude provide a contribution towards operationalizing the analysis of a variety of questions that have to do with the behavioral implications of awareness of unawareness in a manner analogous to the use of measures of risk aversion. Our characterization of greater unawareness aversion is preference-based. Section 4 suggests a procedure for how our measure can be elicited experimentally.

An important aspects of the models of decision making under uncertainty of this paper are the presumption that decision makers are aware of the possible existence of unanticipated consequences, and that their attitudes toward this possibility are captured by their utility functions. The utility function assigns a real number to the "unknown unknown" outcome $\hat{x}$. It is noteworthy that, given the utility representation and the underlying axioms, $\hat{x}$ is treated analogously to every known consequences. Hence, our measures of unawareness attitudes could, in principle, be applied to any known consequence, say $c \in C$. To grasp this, let $q \in \Delta C - \{c\}$ and denote by $\bar{U}_q$ the expected utility of q. Then, by the same logic we employed before, the utility characterization of "c aversion" will be given by $1 - u(c)/\bar{U}_q$. In other words, the attitude toward c depends on its utility relative to the expected utility of q and is captured by measures analogous to the measures of unawareness attitudes developed here.

[8]This sacrifice is the same for all $p'', p''' \in P_q$ such that $z_{p''} - z_{p'''} = z_p - z_{p'}$.

References

[1] Becker, C. K., Melkonyan, T., Proto, E., Soanos, A., Trautman, S. T. (2020) "Reverse Bayesianism: Revising Beliefs in Light of Unforeseen Events," University of Heidelberg.

[2] Chakravarty, S., Kelsey, D., Teitelbaum, J. C. (2022). "Reverse Bayesianism and Act Independence," *Journal of Economic Theory,* 203, 105495.

[3] De Morgan, A. (1838) *An essay on probabilities, and their application to life contingencies and insurances offices,* London: Longman, Orne, Brown, Green, and Longmans.

[4] Dominiak, A., Tserenjigmid, G. (2018). "Belief Consistency and Invariant Risk Preferences," *Journal of Mathematical Economics* 79, 157-162.

[5] Dominiak, Adam and Gerelt Tserenjigmid (2022): "Ambiguity under growing awareness," *Journal of Economic Theory* 199, 105256.

[6] Eichberger, J. and Guerdjikova, A. (2024) "Data, Cases and States," unpublished manuscript.

[7] Ewens, W. J. (1972) "The Sampling Theory of Selectively Neutral Alleles," *Theoretical Population Biology* 3, 87–112.

[8] Grant, S., Quiggin, J. (2013) "Inductive Reasoning about Unawareness," *Economic Theory* 54, 717-755.

[9] Grant, Simon and John Quiggin (2013b): "Bounded awareness, heuristics and the precautionary principle," *Journal of Economic Behavior and Organization* 93, 17–31.

[10] Grant, S., Meneghel, I., Tourky, R. (2022) "Learning Under Unawareness," *Economic Theory*

[11] Halpern, Joseph Y. and Leandro C. Rego (2009): "Reasoning about knowledge of unawareness," *Games and Economic Behavior* 67, 503–525.

[12] Halpern, Joseph Y. and Leandro C. Rego (2013): "Reasoning about knowledge of unawareness revisited," *Mathematical Social Sciences* 65, 73–84.

[13] Heifetz, A., Meier, M., Schipper, B. C. (2006) "Interactive Unawareness," *Journal of Economic Theory* 130, 78-94.

[14] Heifetz, A., Meier, M., Schipper, B. C. (2008) "A Canonical Model for Interactive Unawareness," *Games and Economic Behavior* 62, 305-324.

[15] Heifetz, A., Meier, M., Schipper, B. C. (2013) "Unawareness, Beliefs, and Speculative Trade," *Games and Economic Behavior* 77, 100-121.

[16] Karni, E. (1985) *Decision Making under Uncertainty: The Case of State-Dependent Preferences.* Harvard University Press. Boston MA.

[17] Karni, E. (2022) "A Theory-Based Decision Model," *Journal of Economic Theory* 201,

[18] Karni, E. (2024) "Decisions and Discovery: A non-Bayesian Perspective," Unpublished manuscript.

[19] Karni, E., Vierø, M.-L. (2013). "'Reverse Bayesianism': A choice-based theory of growing awareness," *American Economic Review* 103, 2790-2810.

[20] Karni, Edi and Marie-Louise Vierø (2015): "Probabilistic sophistication and reverse Bayesianism," *Journal of Risk and Uncertainty* 50, 189-208.

[21] Karni, E., Vierø, M.-L. (2017). Awareness of Unawareness: A theory of Decision Making in the Face of Ignorance, *Journal of Economic Theory* 168, 301-328.

[22] Karni, E., Valenzuela-Stookey, Q., Vierø, M.-L. (2021)."'Reverse Bayesianism': A Generalization," *B. E. Journal of Theoretical Economics* 21, 557-569.

[23] Kihlstrom, R. E. and Mirman, L. J. (1974)"Risk Aversion with Many Commodities, "*Journal of Economic Theory* 8, 361-388.

[24] Schipper, B. C. (2022) "Predicting the Unpredictable under Subjective Expected Utility" unpublished manuscript.

[25] Vierø, M.-L. (2021) "An Intertemporal Model of Growing Awareness," *Journal of Economic-Theory* 197, 105351.

[26] Vierø, M.-L. (2021) "Lost in Objective Translation: Awareness of Unawareness when Unknowns are not simply Unknowns,"Working Paper Aarhus University.