A theory of robust experiments for choice under uncertainty[☆]

S. Grant[a,*], J. Kline[b], I. Meneghel[a], J. Quiggin[b], R. Tourky[a]

[a] *Australian National University*
[b] *The University of Queensland*

Abstract

Thought experiments are commonly used in the theory of behavior in the presence of risk and uncertainty to test the plausibility of proposed axiomatic postulates. The prototypical examples of the former are the Allais experiments and of the latter are the Ellsberg experiments. Although the lotteries from the former have objectively specified probabilities, the participants in both kinds of experiments may be susceptible to small deviations in their subjective beliefs. These may result from a variety of factors that are difficult to check in an experimental setting: including deviations in the understanding and trust regarding the experiment, its instructions and its method. Intuitively, an experiment is robust if it is tolerant to small deviations in subjective beliefs in models that are in an appropriate way close to the analyst's model. The contribution of this paper lies in the formalization of these ideas.

1. Introduction

The development of decision theory has been driven, in large measure, by thought experiments questioning the core postulates of the expected utility model, axiomatized for choice under risk by von Neumann and Morgenstern (1944) and for uncertainty by Savage (1954). Experimentalists have gone to great efforts to improve the design of experiments and elicitation of preferences from the participants aimed to test such theories (see Becker et al. (1964), Holt (1986) and Johnson et al. (2015) amongst others). In this paper we add to these

[☆] We thank George Mailath for discussions in the early stages of the development of this project. The comments of Jim Cox, Jürgen Eichberger and Glenn Harrison have greatly improved the exposition. We also thank Konrad Grabiszewski, Mamoru Kaneko, Luke McKenzie, Ludovic Renou, Stephen Roberts, Matthew Ryan and the audiences at Bielefeld, Georgia State, Paris-Dauphine, SAET2015, the Econometric Society World Congress 2015 and the Decision Day Workshop at Duke University. Finally, we thank two anonymous referees and an associate editor for very constructive and insightful comments and suggestions that have greatly improved the paper.

[*] Corresponding author

Email address: simon.grant@anu.edu.au (S. Grant)

efforts by providing a notion of *robustness* of an experiment that ensures the conclusions of the analyst would not be overturned by introducing a vanishingly small amount of doubt about what was in the minds of the participants. We give a simple way to check for robustness of an experiment and some methods to ensure an experiment is robust.[1]

The main idea is that a robust challenge to a decision-theoretic model arising from an experiment remains a challenge in any model that is close to the analyst's model. With this in mind, we demonstrate that many classic experiments such as the prototypical Allais and Ellsberg experiments and their derivatives (Machina, 2009), although well-conceived, do not pose convincing challenges to the particular decision theory the experiment was designed to test. We show, however, how they can be modified to overcome this problem.

The principal ideas, concepts and results can be readily introduced and illustrated in the context of Ellsberg's single-urn thought experiment. In that thought experiment, the reader is asked to "imagine an urn known to contain 30 red balls and 60 black and yellow balls, the latter in unknown proportion." (Ellsberg, 1961, p. 653). A ball is to be drawn from the urn. On the basis of the color of the ball drawn, first consider a choice between a bet that pays $100 if the ball drawn is red and nothing otherwise, denoted b_R, and a bet that pays $100 if the ball drawn is black and nothing otherwise, denoted b_B. Next consider a choice between a bet that pays $100 if the ball drawn is red or yellow and nothing if it is black, denoted b_{RY}, and a bet that pays $100 if the ball drawn is black or yellow and nothing if it is red, denoted b_{BY}. Ellsberg argues that anyone exhibiting the preference pattern $b_R \succ b_B$ and $b_{BY} \succ b_{RY}$ is "simply not acting 'as though' they assigned numerical or even qualitative probabilities to the events in question." (Ellsberg, 1961, p. 656) In particular, this means such a preference pattern is inconsistent with subjective expected utility theory.

Ellsberg's reasoning rests on the assumption that the subject in such an experiment takes the state space to be the *sample* space $\{s_R, s_B, s_Y\}$, where s_c is the sample-state in which a ball of color c is drawn from the urn *independent* of which bet has been chosen by the subject in either problem. By identifying each of these three states with the corresponding vector of bet-consequences we obtain the following 4×3 consequence matrix:

$$C = \begin{array}{c} \\ b_R \\ b_B \\ b_{RY} \\ b_{BY} \end{array} \begin{array}{c} \begin{array}{ccc} s_R & s_B & s_Y \end{array} \\ \begin{bmatrix} 100 & 0 & 0 \\ 0 & 100 & 0 \\ 100 & 0 & 100 \\ 0 & 100 & 100 \end{bmatrix} \end{array}.$$

The set of admissible preferences are ones that represent a subjective expected utility maximizing decision-maker characterized by a pair (u, p) where

[1]In fact, we find that some of the methods used by experimentalists have the effect of making the experiments robust in our sense (Halevy (2007) and Binmore et al. (2012)).

1. u is any (Bernoulli) utility function for which $u(0) < u(100)$.

2. p is any probability distribution over the sample space satisfying $p_R = \frac{1}{3}$, $p_B = q$ and $p_Y = \frac{2}{3} - q$, for some number q, $\frac{1}{90} \leqslant q \leqslant \frac{59}{90}$.[2]

We refer to the set $\mathcal{H}$ of all such pairs as the *admissible parameters* and the pair $(C, \mathcal{H})$ as an *experiment in belief form.* Any preference ordering $\succsim$ generated by an admissible parameter (u, p) is called an *admissible preference.*

Notice that in this experiment, every admissible preference ordering $\succsim$ satisfies $b_R \succsim b_B$ if and only if $b_{RY} \succsim b_{BY}$. This implies, in particular, that the preference ordering $b_{BY} \succ b_{RY} \succ b_R \succ b_B$ is not an admissible preference in the experiment.

As it happens, in this experiment, the sets of admissible and inadmissible preferences do not change if we allow for perturbations of any admissible beliefs in the direction of *any* belief over the sample space $\{s_R, s_B, s_Y\}$.[3] In this sense, we view this experiment as being *internally robust.*

Imagine now, however, that the experimenter suspects that some participants may have an alternative perception of the situation which corresponds to the experiment in belief form $(C', \mathcal{H}')$ given by the consequence matrix

$$C' = \begin{array}{c} \\ b_R \\ b_B \\ b_{RY} \\ b_{BY} \end{array} \begin{array}{c} \begin{array}{cccc} s_R & s_B & s_Y & s^* \end{array} \\ \begin{bmatrix} 100 & 0 & 0 & 0 \\ 0 & 100 & 0 & 0 \\ 100 & 0 & 100 & 0 \\ 0 & 100 & 100 & 100 \end{bmatrix} \end{array},$$

with $\mathcal{H}'$ being the set of all (u, p') for which there is $(u, p) \in \mathcal{H}$ satisfying $p'_s = p_s$ for all $s \in \{s_R, s_B, s_Y\}$, thus making $p'_{s^*} = 0$. That is, $(C', \mathcal{H}')$ is very similar to $(C, \mathcal{H})$ but with an additional state s^* that has zero probability in all the admissible parameters of the experiment. One possible interpretation for this state s^* is the participant conceives of, but places zero probability on, the possibility that the experimenter can "manipulate" the draw of a ball whose color is not red by substituting a yellow ball for a black one or a black ball for a yellow one, whenever such a substitution results in the bet paying out \$0 instead of \$100. Hence, the only bet that pays out \$100 in s^* is b_{BY}.

The version $(C', \mathcal{H}')$ admits exactly the same admissible and inadmissible preferences as $(C, \mathcal{H})$. However, if we now allow for (vanishingly) small perturbations of any admissible belief in the direction of *any* belief over the sample space, we find that the set of admissible preferences expands to what we refer to

---

[2]The restriction $p_R = \frac{1}{3}$ accords with the information that 30 out of the 90 balls are red. The restriction to positive probability for the other two states accords with the information that the urn contains both black and yellow balls, albeit in unknown proportion. Alternatively, as Ellsberg writes, "imagine a sample of two drawn from the 60 black and yellow balls has resulted in one black and one yellow." (Ellsberg, 1961, pp. 653-4)

[3]This includes inadmissible beliefs such as those for which $p_R \neq \frac{1}{3}$.

in the sequel as ε-admissible in $(C', \mathcal{H}')$. In particular, the inadmissible preference ordering $b_{BY} \succ b_{RY} \succ b_R \succ b_B$ is ε-admissible. This particular preference pattern is generated by a subjective expected utility maximizer characterized by a utility function satisfying $u(0) < u(100)$ and a belief

$$\left(\frac{1}{3}, \frac{1-\varepsilon}{3}, \frac{1-\varepsilon}{3}, \frac{2\varepsilon}{3}\right),$$

where ε can be *any* number in $(0, 1)$, no matter how small. That is, by allowing for small perturbations of any admissible belief (such as $(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}, 0)$) in the direction of any other probability distribution over the sample-space $\{s_R, s_B, s_Y, s^*\}$ (such as $(\frac{1}{3}, 0, 0, \frac{2}{3})$), we see this preference ordering is ε-admissible.[4]

Although $(C, \mathcal{H})$ may clearly be what Ellsberg had in mind, we contend that one cannot rule out, either from *a priori* reasoning or from any *ex post* experimental observation, that a participant did not have some alternative version in mind, for example, the version $(C', \mathcal{H}')$ described above. Notice that both of these versions have the same set of bets and, for each bet, the same set of consequences and, as we have already noted, the same sets of admissible preferences. However, the sample-spaces of the two versions differ, which we have seen leads to differences in the set of ε-admissible preferences. If we desire a notion of robustness that does not depend on which of these two versions a participant has in mind, then we need a stronger notion than internal robustness.

Even more troubling, the sample space of $(C', \mathcal{H}')$ was chosen somewhat arbitrarily. We introduced and motivated it by one story, but there are other possible stories one might tell, each with its own distinct probability model. There are many more models with widely varying sample spaces that are also consistent with the underlying uncertainty described in this Ellsberg scenario. Any of these potentially can serve as the version in the mind of a participant.

Building on this insight we define in Section 3 an equivalence class of experiments in belief form which includes all these potential versions. Our formal notion of robustness will be one that is invariant across such an equivalence class of experiments. The associated equivalence relation is based on the set of bets, the associated set of conceivable consequences for each bet and the set of admissible preferences over the bets, thus making no explicit reference to the sample space. Loosely speaking, each equivalence class of experiments in belief form can be interpreted as embodying a notion of *revealed preference* equivalence.

Given the equivalence class associated with an experiment in belief form $(C, \mathcal{H})$, we shall say a preference ordering over bets is *weakly-admissible* if it is ε-admissible in *some* version from this class. This leads naturally to our notion of robustness both for observed violations of the theory within an experiment and for the entire experiment itself.

[4]This is reminiscent of Kadane (1992) in which he proposes that the participants' 'healthy skepticism' of the experimenter and a suspicion that he might manipulate the design to their disadvantage could 'explain' both Allais and Ellsberg type phenomena without having to resort to a model of behavior that does not conform to expected utility.

- A preference ordering is *robustly inadmissible* if it is not weakly-admissible.
- An experiment is *robust* if every inadmissible preference ordering is robustly inadmissible.

Although this notion of robustness requires a property to hold on an entire equivalence class of experiments, we show it can be characterized solely in terms of the admissible preferences over the bets for any given version of the experiment. So in the case of the Ellsberg experiment, for example, we can characterize robustness purely in terms of $(C, \mathcal{H})$.

In particular, the main result of the paper is a characterization of robust inadmissibility and a corollary characterizing robust experiments. They are stated in terms of coarsenings and refinements of preference orderings and a notion of range dominance. A preference ordering over bets is *finer* than another preference ordering if any strict preference between a pair of bets in the latter implies the corresponding strict preference holds in the former. The latter preference ordering is also viewed as being *coarser* than the former. A preference ordering is range dominant with respect to a parameter if it ranks bet b over bet b' whenever the worst outcome from b is better than the best outcome from b'.

Characterization of Robust Inadmissibility (Theorem 1)

A preference ordering is robustly inadmissible if and only if there is no admissible parameter with respect to which it is range dominant that generates a coarser preference ordering.

Characterization of Robust Experiments (Corollary 1.1)

An experiment in belief form is robust if and only if every preference ordering that is range dominant with respect to some admissible parameter that generates a coarser preference ordering is also admissible.

Applying these results to the Ellsberg experiment, we see that the inadmissible preference ordering $b_{BY} \succ b_{RY} \succ b_R \succ b_B$ is not robustly inadmissible since it is range-dominant with respect to the admissible parameter (u, p), where $u(0) < u(100)$ and $p(s_B) = \frac{1}{3}$ $(= p(s_R))$, which generates the coarser preference relation $b_{BY} \sim' b_{RY} \succ' b_R \sim' b_B$. Correspondingly, we see that the Ellsberg experiment is not robust.

The simple explanation of the characterizations is that non-robust anomalous-inadmissible preferences are highly susceptible to (misspecification) error in cases where true preferences are at indifferences. Even vanishingly small perturbations can produce reported preferences that are inconsistent with the theory being tested.

Thus, the take-home message of Theorem 1 is that if a participant in an experiment reports a preference ordering that is inadmissible for that experiment, then the analyst should check whether there is any admissible preference ordering that is *coarser*. If that turns out to be the case, then the analyst cannot rule out the possibility that the participant is someone who conforms to the decision theory being tested, but who has an equivalent version of the experiment 'in mind' for which the observed 'anomalous' choices can be attributed to a (vanishingly) small perturbation of some admissible preference.

On the other hand, if every coarser preference ordering is also inadmissible, then this reported failure of admissibility is robust in the sense that it cannot be rationalized by a (vanishingly) small perturbation to *any* admissible preference within *any* of the equivalent versions. Thus, even if the experiment itself is not robust, this particular piece of data may be viewed as a *robust* rejection of the participant's behavior conforming to the theory being tested.

The rest of the paper is organized as follows. The formal framework is set up in Section 2. Robust experiments are introduced in Section 3 and a number of examples of non-robust experiments are studied in Section 4. We show in Section 5 how the analysis can be applied to experiments such as the Allais common consequence effect, that are presented in terms of (objective) lotteries and demonstrate how this classic experiment also fails our robustness criteria. Section 6 outlines different methods for robustifying an experiment. The most straightforward method involves eliciting certainty equivalents for each of the bets.[5]

This method can be illustrated in our Ellsberg example. Let e_b denote the *certainty equivalent* of bet b, and suppose the decision maker reveals the modal preference pattern $b_R \succ b_B$ and $b_{BY} \succ b_{RY}$. Including the four certainty equivalents, we have $b_R \sim e_R \succ e_B \sim b_B$ and $b_{BY} \sim e_{BY} \succ e_{RY} \sim b_{RY}$. Notice this means by range-dominance that $e_R > e_B$ and $e_{BY} > e_{RY}$. Furthermore, any coarsening that satisfies range-dominance must retain the two strict preferences $e_R \succ e_B$ and $e_{BY} \succ e_{RY}$.[6] Hence there is *no* admissible coarsening that would allow for either $b_R \sim b_B$ or $b_{BY} \sim b_{RY}$. Thus by Theorem 1 the preference pattern is robustly inadmissible.

2. Decision Theories and Experiments

An experiment is designed to test whether decision-makers in a certain population behave in a way that is consistent with a given decision theory.

2.1. Generalized Mean Utility Decision Theories

We are interested in theories of how a decision-maker evaluates alternatives under risk or uncertainty. The uncertainty is represented by a finite state space and the objects of choice are acts which are mappings from the state space to a consequence space, or equivalently, state-contingent consequence vectors. More formally, we associate with any non-empty set of consequences $\mathcal{C}$ and any

[5] We thank an anonymous referee for suggesting this method.

[6] We acknowledge that one should be careful about interpreting any difference in elicited certainty equivalents as reflecting a strict preference no matter how small it may be. At some level, such differences might more sensibly be explained by a mistake, misperception or lack of motivation on the part of the decision-maker.

finite non-empty state space S the set of acts $\mathcal{C}^S$, i.e., the set of vectors of state contingent consequences.

For any real-valued function $f : \mathcal{C} \to \mathbb{R}$, we denote by $f \circ a$ the state-contingent real-valued vector obtained from the act a by converting each consequence a_s into the real number $f(a_s)$. We denote the inner product of any two real-valued vectors x and y of the same dimension by $\langle x, y \rangle$. For a vector $x \in \mathbb{R}^{|S'|}$ and S a subset of S', we use x^S to denote the vector $(x_s)_{s \in S}$. We also use $\Delta^{|S|-1} \subset \mathbb{R}_+^{|S|}$ to denote the standard $(|S|-1)$-dimensional simplex.

A decision-maker's choice is assumed to be guided by a binary relation, the decision-maker's preferences, defined over a set of acts. A decision theory corresponds to a family of representations of preferences over a set of acts. Our focus in this paper is on preferences which admit what we refer to as a "generalized mean utility" representation.

Let Ω be the universal set of states and let $\mathcal{S}$ be the collection of all finite non-empty subsets of Ω. Let Γ denote the universal set of consequences and let $\mathcal{K}$ be the collection of all non-empty subsets of Γ.

Definition 1. Fix a set of consequences $\mathcal{C} \in \mathcal{K}$, a state space $S \in \mathcal{S}$, and a binary relation $\succsim$ over the set of acts $\mathcal{C}^S$. A *generalized mean utility* representation of $\succsim$ comprises a utility index $u : \mathcal{C} \to \mathbb{R}$ and a decision weight function $w : \mathbb{R}^{|S|} \to \Delta^{|S|-1}$ such that for any two acts a and a',

$$a \succsim a' \text{ if and only if } \langle u \circ a, w(u \circ a) \rangle \geqslant \langle u \circ a', w(u \circ a') \rangle \,.$$

Since we focus on generalized mean utility representations, the binary relation $\succsim$ guiding the decision-maker must be complete and transitive, that is, it must be a *preference ordering*. As the next two examples demonstrate, this formulation is rich enough to accommodate subjective expected utility maximizers as well as the generalization, Choquet expected utility maximizers of Schmeidler (1989), that allows for non-neutral attitudes toward perceived ambiguity.

Example 1 (Subjective Expected Utility)**.** Preferences in this family are characterized by a utility index u and a probability measure P defined over the set of subsets of S. Let p denote the probability vector in $\Delta^{|S|-1}$ for which $p_s = P(\{s\})$ for each $s \in S$. The corresponding generalized mean representation is given by the pair $(u\,, \overline{w}^p)$, where the decision weight function $\overline{w}^p$ is the constant function $\overline{w}^p(x) = p$ for all x. Notice that, for any act a, the inner product:

$$\langle u \circ a, \overline{w}^p(u \circ a) \rangle = \langle u \circ a, p \rangle = \sum_{s \in S} p_s u(a_s) \,,$$

which is indeed the expectation of the state-contingent utility vector $u \circ a$ with respect to the probability measure P.

In the sequel where there is no risk of confusion, we shall often use $(u\,, p)$ for the generalized mean parameter characterizing a subjective expected utility maximizer and refer to it as an SEU-parameter.

Example 2 (Choquet Expected Utility)**.** Preferences in this family are characterized by a utility index u and a capacity ν (or 'non-additive probability') defined over the set of subsets of S, that is normalized ($\nu(\emptyset) = 0$ and $\nu(S) = 1$) and set-monotonic ($E \subset F \subseteq S \Rightarrow \nu(E) \leqslant \nu(F)$).[7] A corresponding generalized mean utility representation is given by the pair $(u\,, w^{\nu})$, where w^{ν} is the decision weight function given by: for each $x \in \mathbb{R}^{|S|}$ and each state $s \in S$:

$$w_s^{\nu}(x) = \frac{\nu(\{t : x_t \geqslant x_s\}) - \nu(\{t : x_t > x_s\})}{|\{t : x_t = x_s\}|}$$

Notice that for any act a, the inner product:

$$\langle u \circ a, w^{\nu}(u \circ a)\rangle = \sum_{z \in \mathbb{R}} [\nu(\{s : u(a_s) \geqslant z\}) - \nu(\{s : u(a_s) > z\})] \times z\,.$$

which is indeed the Choquet integral of the state-contingent utility vector $u \circ a$ with respect to the capacity ν.

In the sequel where there is no risk of confusion, we shall often use $(u\,, \nu)$ for the generalized mean utility parameter characterizing a Choquet expected utility maximizer.

The above two examples can both be subsumed in the very general family of **MBA preferences** (for **M**onotonic, **B**ernoullian, **A**rchimedean) introduced and axiomatized by Cerreia-Vioglio et al. (2011). In fact in our setting any MBA preference relation admits a generalized mean utility representation.

Example 3 (MBA Preferences)**.** Preferences in this family are characterized by a utility index u, and a functional $I : \mathbb{R}^{|S|} \to \mathbb{R}$, that is monotonic, continuous and normalized in the sense that $I(\gamma\,, \gamma\,, \ldots \gamma) = \gamma$, for any $\gamma \in \mathbb{R}$. The preferences over acts that the pair $(u\,, I)$ represents, is the one generated by the functional $I\,(u \circ a)$.

If the preference relation $\succsim$ admits an MBA representation $(u\,, I)$, then we can construct a generalized mean utility representation using the same utility index u and defining $w(\cdot)$ as follows. For each $x \in \mathbb{R}^{|S|}$ and each state $s \in S$, set

$$w_s(x) = \begin{cases} \frac{\alpha(x)}{|\mathrm{argmin}_{t \in S}\, x_t|} & \text{if } s \in \mathrm{argmin}_{t \in S}\, x_t \\ \frac{1 - \alpha(x)}{|\mathrm{argmax}_{t \in S}\, x_t|} & \text{if } s \in \mathrm{argmax}_{t \in S}\, x_t \text{ and } \max_{t \in S} x_t > \min_{t \in S} x_t \\ 0 & \text{otherwise} \end{cases}$$

where

$$\alpha(x) = \max\left\{\alpha \in [0,1] : \alpha \min_{t \in S} x_t + (1 - \alpha) \max_{t \in S} x_t = I(x)\right\}.$$

[7] The capacity ν is a probability if, in addition to being normalized and set-monotonic, it is additive, that is, $\nu(E) + \nu(F) = \nu(E \cup F) + \nu(E \cap F)$.

Notice that for any constant vector $x = (\gamma\,, \gamma\,, \ldots\, \gamma) \in \mathbb{R}^{|S|}$, $w_s(x) = \frac{1}{|S|}$, for all $s \in S$. Furthermore, by construction we have that for any act $a \in \mathcal{C}^S$

$$\langle u \circ a, w(u \circ a)\rangle = \alpha(x) \min_{t \in S} x_t + (1 - \alpha(x)) \max_{t \in S} x_t = I(u \circ a)\,.$$

The flexibility of the family of generalized mean utility representations means that such representations are generally non-unique. For example, since MBA preferences include preferences that may be represented by SEU or CEU functional forms, we note that in addition to the generalized mean utility representations that we constructed in examples 1 and 2, respectively, we can also construct generalized mean utility representations in terms of a weighted sum of the utilities of the *extreme* outcomes, with intermediate outcomes 'only' contributing indirectly via their effect on those decision weights.

The non-uniqueness of generalized mean utility representations is not an issue for us. Any evidence that cannot be reconciled with a particular decision theory, such as, say, SEU, means that *no* generalized mean utility representation consistent with SEU can generate preferences that rationalize this evidence.

We now outline our notion of a theory that is to be tested in an experiment. A theory tells us for each consequence and state space pair the set of generalized mean utility representations that generate preferences over the corresponding set of acts that conform to the precepts of the particular decision theory. We assume that a theory is '(linearly) open' in the sense that if the decision-weight function of any generalized mean utility representation is slightly 'perturbed' in the direction of a constant decision weight function, we remain within the theory. That is, any decision-weight function may be 'flattened' slightly without leaving the theory. Furthermore, the theory is coherent in the sense that if we consider a larger state-space, then there is a natural embedding of the set of generalized mean utility representations involving the original state-space to the larger state-space.

More formally, we define a theory as follows.

Definition 2. A *theory* $\mathcal{T}$ assigns to each consequence and state space pair $(\mathcal{C}, S) \in \mathcal{K} \times \mathcal{S}$ a subset of generalized mean utility representations $\mathcal{T}(\mathcal{C}\,, S)$ in which for every $(u, w) \in \mathcal{T}(\mathcal{C}, \mathcal{S})$:

1. (*linearly open*) the set $W_u = \{w' : (u, w') \in \mathcal{T}(\mathcal{C}, S)\}$ is open in the sense that for any $w' \in W_u$ and any $p \in \Delta^{|S|-1}$, there is an $\varepsilon^p \in (0, 1)$ such that $(1 - \varepsilon)w' + \varepsilon\overline{w}^p \in W_u$ for all $\varepsilon \in (0, \varepsilon^p)$;

2. (*coherent*) and for any set $S' \in \mathcal{S}$ that is a superset of S, there exists a $(u, w') \in \mathcal{T}(\mathcal{C}, S')$ such that $w'_s(x) = w_s(x^S)$ for all $s \in S$, whenever $x \in \mathbb{R}^{|S'|}$ and $\{x_s : s \in S\} = \{x_s : s \in S'\}$.

As an example the MBA theory $\mathcal{T}^{\text{MBA}}$, can be defined as follows. For each $(\mathcal{C}, S) \in \mathcal{K} \times \mathcal{S}$, set

$$\mathcal{T}^{\text{MBA}}(\mathcal{C}, S) := \left\{ (u, w) : \begin{array}{l} \text{there exists some MBA representation } (u, I) \\ \text{s.t. } \langle u \circ a, w\,(u \circ a)\rangle = I\,(a)\,, \text{ for all } a \in \mathcal{C}^S \end{array} \right\}$$

It is straightforward to verify that $\mathcal{T}^{\text{MBA}}$ is linearly open and coherent. In an analogous manner we can define the SEU theory $\mathcal{T}^{\text{SEU}}$ and the CEU theory $\mathcal{T}^{\text{CEU}}$. Finally, the generalized mean utility theory $\mathcal{T}^{\text{GMU}}$ which consists of all generalized mean utility representations is also a theory. While each of the theories given here include SEU theory, that is, they include the constant decision weight functions, our definition of a theory does not require it.

2.2. Experiment in Belief Form

An experiment is designed to test a decision theory, or, more typically, some suitably specified restriction of that theory. It begins with a finite set of *bets* $B = \{b_1, b_2, \ldots, b_{|B|}\}$ (sometimes referred to as *actions*, *choices* or *prospects*) and a finite set of *sample states* $S = \{s_1, s_2, \ldots, s_{|S|}\}$ and a set of consequences $\mathcal{C}$. We use b to denote a generic bet in B and s to denote a generic sample state in S.

Definition 3. A *consequence matrix* is a function from $B \times S$ to $\mathcal{C}$ associating with each ordered pair $(b, s) \in B \times S$ a consequence $c_{bs} \in \mathcal{C}$.

It can be represented in matrix form as:

$$C = \begin{array}{c} \\ b_1 \\ b_2 \\ \vdots \\ b_{|B|} \end{array} \begin{array}{c} \begin{array}{cccc} s_1 & s_2 & \cdots & s_{|S|} \end{array} \\ \begin{bmatrix} c_{b_1s_1} & c_{b_1s_2} & \cdots & c_{b_1s_{|S|}} \\ c_{b_2s_1} & c_{b_2s_2} & \cdots & c_{b_2s_{|S|}} \\ \vdots & \vdots & \ddots & \vdots \\ c_{b_{|B|}s_1} & c_{b_{|B|}s_2} & \cdots & c_{b_{|B|}s_{|S|}} \end{bmatrix} \end{array}$$

Given a bet b and sample-state s, let

$$C_b = \begin{bmatrix} c_{bs_1} & c_{bs_2} & \cdots & c_{bs_{|S|}} \end{bmatrix} \quad \text{and} \quad C_s = \begin{bmatrix} c_{b_1s} \\ c_{b_2s} \\ \vdots \\ c_{b_{|B|}s} \end{bmatrix}$$

denote the bet b row of C and the sample-state s column of C, respectively. For each bet b let $\underline{C}_b$ be the set $\{c_{bs} \colon s \in S\}$, which is the set of *consequences of the bet* b. Likewise, let $\underline{C}$ be the set $\{c_{bs} \colon b \in B, s \in S\}$, of *consequences of the experiment.*

We characterize an experiment by a consequence matrix C together with a set of parameters which define the set of admissible preferences over bets.

Definition 4. An *experiment in belief form* testing a null hypothesis $\mathcal{H}$ within the theory $\mathcal{T}$ is a pair $(C, \mathcal{H})$ where C is a $|B| \times |S|$ consequence matrix and $\mathcal{H} \subseteq \mathcal{T}(\underline{C}, S)$ is the set of admissible parameters of the experiment encoding the null hypothesis being tested.

An experiment in belief form $(C, \mathcal{H})$ is always testing a null hypothesis of the form: every decision maker in the population has preferences that can be represented by a parameter in $\mathcal{H}$. The test is always within some theory $\mathcal{T}$. We can take the theory to be as large as $\mathcal{T}^{\text{GMU}}$, but the experimenter typically constructs the null hypothesis within some informed restricted theory like $\mathcal{T}^{\text{SEU}}$ or $\mathcal{T}^{\text{CEU}}$. When the null hypothesis $\mathcal{H} = \mathcal{T}(\underline{C}, S)$, the experiment is regarded as a test of the theory $\mathcal{T}$. Since the null hypothesis is always encoded in the experiment in belief form $(C, \mathcal{H})$, we will sometimes simply say the experiment in belief form $(C, \mathcal{H})$ within theory $\mathcal{T}$, and even drop the theory when it is clear from the context.

A parameter $(u, w) \in \mathcal{T}(\underline{C}, S)$ is said to *generate* the binary relation $\succsim$ on B whenever $b \succsim b'$ if and only if $\langle u \circ c_b, w(u \circ c_b) \rangle \geqslant \langle u \circ c_{b'}, w(u \circ c_{b'}) \rangle$. This leads to the notion of admissible preferences within a given experiment.

Definition 5. A binary relation $\succsim$ on B is *admissible* in $(C, \mathcal{H})$ if there is a parameter $(u, w) \in \mathcal{H}$ that generates $\succsim$. We refer to any binary relation $\succsim$ on B for which there is no parameter in $\mathcal{H}$ that generates it as *inadmissible* in $(C, \mathcal{H})$.

As an example of these concepts, the version $(C, \mathcal{H})$ of the Ellsberg experiment from Section 1 is indeed an experiment that tests subjective expected utility theory, as does the alternative version $(C', \mathcal{H}')$.

The analyst designs an experiment with the aim of testing a decision theory, such as subjective expected utility, possibly in combination with some other global restriction, such as constant absolute risk aversion. Our formulation allows the analyst to choose a consequence matrix C and an accompanying set of parameters $\mathcal{H}$ that is tailored to express her challenge to the theory. The subjects participate in the designed experiment and their preferences over the set of bets are elicited. If the observed preferences of a subject turn out to be inadmissible then the analyst concludes that the decision-maker in question has preferences that do not conform to the theory (as restricted by $\mathcal{H}$), posing a challenge to the theory.

3. Robust Experiments

3.1. Internally Robust Experiments

For the experimental results to constitute a robust challenge to the theory being tested, the set of *inadmissible* preference orderings in the experiment should be unaffected by (vanishingly) small perturbations in the admissible parameters in the experiment $(C, \mathcal{H})$. To capture robustness, we first introduce a weaker notion of admissibility that expands the set of admissible preference orderings and makes it more difficult to reject a theory.

Definition 6. A binary relation $\succsim$ is ε*-admissible* in $(C, \mathcal{H})$ within theory $\mathcal{T}$ if there is an admissible parameter $(u, w) \in \mathcal{H}$ and a parameter $(u, w') \in \mathcal{T}(\underline{C}, S)$ such that for all $\varepsilon \in (0, 1)$, the parameter $(u, (1-\varepsilon)w + \varepsilon w')$ is in $\mathcal{T}(\underline{C}, S)$ and generates $\succsim$.

Our notion of ε-admissibility is based on the idea that a subject may have a slightly different experiment in mind, albeit with the same state space, which can be captured by a slight perturbation in the decision weight outside the admissible set, but remaining within the theory $\mathcal{T}$. Our first notion of robustness is one for which this relaxation of admissibility does not expand the set of admissible preference orderings.

Definition 7. An experiment in belief form $(C, \mathcal{H})$ is *internally robust* within the theory $\mathcal{T}$ if every ε-admissible preference ordering is admissible.

In an internally robust experiment an inadmissible preference ordering challenges a theory more convincingly than in an experiment that is not internally robust, where inadmissible preferences may in fact be ε-admissible. Internal robustness ensures that for vanishingly small differences between the actual experiment, and the one in the subject's mind, the set of admissible preferences are the same. The restriction of the theory $\mathcal{T}$ ensures the preferences of the subject conform to the theory in question.

Returning to the illustrative Ellsberg example given in Section 1, we can now verify that the version $(C, \mathcal{H})$ is internally robust within $\mathcal{T}^{\text{SEU}}$ while the alternative version $(C', \mathcal{H}')$ is not. So, if the analyst is convinced that the participants perceive the experimental design according to $(C, \mathcal{H})$ and its associated state space S, then the analyst will view the observation of any inadmissible preference pattern in the experiment $(C, \mathcal{H})$ as constituting a violation of the theory. With the second experiment $(C', \mathcal{H}')$, the violation of subjective expected utility posed by the preference pattern $b_{BY} \succ b_{RY}$ and $b_R \succ b_B$ is not as robust. The apparent violation may arise from a vanishingly small perturbation of an admissible decision weight function (in this case, the subjective probability belief of a subjective expected utility maximizer).

3.2. Equivalent Versions and Fully Robust Experiments

In this section we first define an equivalence class of experiments which includes both versions of the Ellsberg experiment from the illustrative example in Section 1. We then extend ε-admissibility to define what we mean for an experiment to be fully robust.

Definition 8. Fix a theory $\mathcal{T}$. Two experiments $(C, \mathcal{H})$ and $(C', \mathcal{H}')$ in which $\mathcal{H} \subseteq \mathcal{T}(\underline{C}, S)$ and $\mathcal{H}' \subseteq \mathcal{T}(\underline{C}', S')$ are *(equivalent) versions* of a test within theory $\mathcal{T}$, if they have the same bets, the same consequences for each bet, and the same admissible preferences.

Two versions $(C, \mathcal{H})$ and $(C', \mathcal{H}')$ of a test within theory $\mathcal{T}$ are essentially testing the same null hypothesis. The sets of admissible parameters $\mathcal{H}$ and $\mathcal{H}'$ differ according to the state space, but each encodes the same relevant information about the test since each admits the same set of preferences.

For the experimental results to constitute a fully robust challenge to the theory being tested, the set of *inadmissible* preference orderings in the experiment should be unaffected by (vanishingly) small perturbations in the decision-weight function parameter of *any* version of the experiment $(C, \mathcal{H})$. So to capture (full) robustness, we extend our notion of ε-admissibility to the equivalence class of versions of $(C, \mathcal{H})$.

Definition 9. A preference ordering $\succsim$ is *weakly-admissible* in $(C, \mathcal{H})$ within theory $\mathcal{T}$ if it is ε-admissible in some version $(C', \mathcal{H}')$.

Definition 10. A binary relation $\succsim$ is *robustly-inadmissible* in $(C, \mathcal{H})$ within theory $\mathcal{T}$ if it is not weakly admissible in $(C, \mathcal{H})$.

Definition 11. An experiment in belief form $(C, \mathcal{H})$ within theory $\mathcal{T}$ is *robust* if every weakly admissible preference ordering is admissible.

Underpinning our robustness notion is the idea that the analyst who designs the experiment $(C, \mathcal{H})$ does not know which version of the experiment is in the mind of the participant. So, for a *particular* observed violation of subjective expected utility to be deemed robust, we require that no matter which version of the experiment the participant has in mind, the violation cannot be attributed to a (vanishingly) small perturbation of an admissible probability belief within that version. Correspondingly, the experiment itself is deemed robust, if a perturbation of a subjective belief in *any* version of the experiment does not affect its set of admissible preferences.

We state the following proposition which characterizes robust experiments both in terms of weak admissibility and internal robustness.

Proposition 1. The following statements are equivalent for an experiment in belief form $(C, \mathcal{H})$ within theory $\mathcal{T}$:

1. It is robust.

2. Every weakly-admissible preference ordering is admissible.

3. Every version of the experiment is internally robust.

Although robustness requires a property to hold on an entire equivalence class of experiments, we characterize it entirely in terms of the admissible preference orderings over the bets for any given version of the experiment. In particular, robustness can be characterized solely in terms of the analyst's version of the experiment $(C, \mathcal{H})$. For this, we introduce two notions about preference orderings. The first notion compares different preference orderings in terms of refinements or coarsenings of preference orderings. The second notion is one of dominance for a given preference ordering.

Definition 12. Let $\succsim$ and $\succsim'$ be preference orderings on B. We say that $\succsim'$ is *finer* than $\succsim$ on B if for all $b, \hat{b} \in B$ we have $b \succ \hat{b}$ implies $b \succ' \hat{b}$. Correspondingly, we say that $\succsim'$ on B is *coarser* than $\succsim$ on B, whenever $\succsim$ is finer than $\succsim'$.

Definition 13. Fix an experiment in belief form $(C, \mathcal{H})$ with bets B and state space S. A preference ordering $\succsim$ on B is *range dominant* with respect to $(u, w) \in \mathcal{T}(\underline{C}, S)$ if for all $b, \hat{b} \in B$:

1. $\min_{s\in S} u(c_{bs}) \geqslant \max_{s\in S} u(c_{\hat{b}s})$ implies $b \succsim \hat{b}$,

2. $\min_{s\in S} u(c_{bs}) > \max_{s\in S} u(c_{\hat{b}s})$ implies $b \succ \hat{b}$.

Now we can express the main results of the paper.

Theorem 1. A preference ordering is robustly inadmissible if and only if there is no admissible parameter with respect to which it is range dominant that generates a coarser preference ordering.

We give two comments about the theorem: one about the only-if part, and one about the if-part. The only-if part of Theorem 1 states that for an inadmissible ordering to be robustly inadmissible, it cannot be range-dominant with respect to some admissible parameter and finer than the preference ordering generated by that admissible parameter.

The if-part of Theorem 1 implies that *any* inadmissible preference ordering that is range dominant with respect to some admissible parameter and is finer than the preference ordering generated by that parameter is *also* weakly admissible. This comes from our strong notion of robustness which considers all versions and all perturbations of the decision weight function with the same utility index u. If, on the other hand, the analyst had some additional information about the possible versions or perturbations, then some of the refinements of admissible preferences could be excluded. We have explored one such case with the notion of internal robustness of a specific experiment.

Using Theorem 1, we can immediately characterize robust experiments.

Corollary 1.1. An experiment in belief form is robust if and only if every preference ordering that is range dominant with respect to some admissible parameter that generates a coarser preference ordering is also admissible.

4. Examples of Non-robust Experiments

We present four examples. The first is a non-robust test of constant absolute risk aversion (CARA). The second is a single-urn Ellsberg-style experiment based on an example that appears in Eichberger et al. (2007, p. 892). Although the experiment itself is not robust, we show that there is a robustly inadmissible preference pattern that accords with our intuition of how an ambiguity averse decision-maker might choose. The third and fourth are two-urn Ellsberg-style experiments which correspond to the "Reflection Example" and to the "50:51 Example", respectively, from Machina (2009).

Example 4 (A Non-robust Test of CARA)**.** The subjects are presumed to be expected utility maximizers that are either CARA or DARA (decreasing absolute risk aversion). For this experiment we seek to challenge the hypothesis that the subjects are all CARA. We offer the subjects four bets involving the toss of a fair coin. So the state space is $S = \{H, T\}$ and the payoffs of the bets are given by the following consequence matrix:

$$C = \begin{array}{c} \\ b_1 \\ b_2 \\ b_3 \\ b_4 \end{array} \begin{array}{c} \begin{array}{cc} H & T \end{array} \\ \begin{bmatrix} 5 & 5 \\ 2 & 10 \\ 55 & 55 \\ 52 & 60 \end{bmatrix} \end{array}.$$

Here the set of parameters $\mathcal{H}$ can be expressed by pairs (α, p^*), where $\alpha > 0$ is the coefficient of absolute risk aversion for the Bernoulli utility function of the form $u(x) = -e^{-\alpha x}$ and $p^*(H) = \frac{1}{2}$.

Notice that the preference ordering $b_4 \succ b_3 \succ b_1 \succ b_2$ is inadmissible for the CARA model. However, for the admissible pair (α^*, p^*) in which

$$-e^{-5\alpha^*} = -\frac{1}{2}e^{-2\alpha^*} - \frac{1}{2}e^{-10\alpha^*}$$

we have $b_4 \sim^* b_3 \succ^* b_1 \sim^* b_2$, thus making $b_4 \succ b_3 \succ b_1 \succ b_2$ weakly admissible and hence not robustly inadmissible. ■

Example 5 (A Robustly Inadmissible Preference)**.** Consider an urn that contains 200 balls numbered 1 to 200. The balls numbered 1 to 66 are red, the balls numbered 67 to $200 - 2n$ are black and the remainder (that is, those numbered from $[201 - 2n]$ to 200) are yellow. The only information a participant has about n is that it is an integer and that $1 \leqslant n \leqslant 66$. Let O (respectively, E) be the event that the ball drawn from the urn has an odd (respectively, even) number on it. Let R (respectively, B, Y) be the event that color of the ball drawn is red (respectively, black, yellow). Let OR be the event that the ball drawn from the urn has an odd number and its color is red, and so on. Notice that the number of balls that are black with an odd number on them or yellow with an even number on them is 67 no matter what value n takes. Similarly, the number of balls that are black with an even number on them or yellow with an odd number on them is also 67 no matter what value n takes. We take the sample space to be $S = \{OR, OB, OY, ER, EB, EY\}$ and the set of bets to be $B = \{b_1, b_2, b_3, b_4, b_5, b_6\}$. The payoffs are given in the following consequence matrix

$C^{\mathrm{RI}} =$

	OR	OB	OY	ER	EB	EY
b_1	\$100	\$0	\$0	\$100	\$0	\$0
b_2	\$0	\$100	\$0	\$0	\$0	\$100
b_3	\$100	\$0	\$0	\$0	\$0	\$0
b_4	\$0	\$100	\$0	\$0	\$0	\$0
b_5	\$0	\$0	\$0	\$100	\$0	\$0
b_6	\$0	\$0	\$0	\$0	\$0	\$100

The bet b_1 is a standard 'unambiguous' bet that the color of the ball drawn is red. The bet b_2 can be viewed as a way of implementing the suggestion by

Raiffa (1961) to avoid the ambiguity associated with a bet on black or a bet on yellow by randomly choosing which of these two colors to bet on. Here we are using the property of whether the number on the ball drawn is odd or is even 'to randomize' between the choices of black or yellow. A choice between bets b_3 and b_4 corresponds to a choice between betting on red versus betting on black *conditional* on the number of the ball drawn is odd. Similarly, a choice between bets b_5 and b_6 is a choice between betting on red versus betting on yellow, conditional on the number of the ball drawn is even.

The set of admissible parameters $\mathcal{H}^{\mathrm{RI}}$ can be characterized by pairs (u, p) where u is any Bernoulli utility function with $u(0) < u(100)$ and p is a probability vector in $\Delta^{|S|-1}$, satisfying

$$p_{OR} = p_{ER} = \frac{33}{200} \qquad p_{OB} = p_{EB} = q \qquad p_{OY} = p_{EY} = \frac{67}{200} - q\,,$$

where $\frac{1}{200} \leqslant q \leqslant \frac{66}{200}$.

We note that in the experiment $(C^{\mathrm{RI}}, \mathcal{H}^{\mathrm{RI}})$, the preference pattern $b_2 \succ b_1$, $b_3 \succ b_4$ and $b_5 \succ b_6$ is robustly inadmissible. To see this, notice that for any preference generated by an admissible parameter (u, p) we must have $b_2 \succ b_1$, and $b_3 \succsim b_4 \Rightarrow b_6 \succ b_5$. This follows since for any $(u, p) \in \mathcal{H}^{\mathrm{RI}}$,

$$\begin{gathered} p_{OB} + p_{EY} > p_{OR} + p_{ER} \\ \max\{p_{OB}, p_{EY}\} > p_{OR} = p_{ER}\,. \end{gathered}$$

This implies that every coarsening of $b_2 \succ b_1$ or $b_3 \succ b_4$ or $b_5 \succ b_6$ is inadmissible. By Theorem 1 the inadmissible preference pattern is seen to be robustly inadmissible. But this pattern accords with what we expect from someone who exhibits ambiguity aversion, since b_1, b_2, b_3, b_5 are all unambiguous (and b_2 first-order stochastically dominates b_1) while b_4 and b_6 are bets for which there is ambiguity about the probability of winning. ■

We now turn to the "Reflection Example" from Machina (2009). It was designed as an Ellsberg-style experiment to generate choice paradoxes that could not be explained by any member of the generalization of expected utility known as Choquet Expected Utility (CEU). Although the experiment is designed to elicit preference patterns that are inadmissible for this larger class of preferences, such inadmissible patterns are not robustly inadmissible even for the smaller class of subjective expected utility maximizers.

Example 6 (The Reflection Example from Machina (2009))**.** The subject is presented with two urns each containing 100 balls that are either Red or Black. Urn 1 is known to contain 50 balls of each color. The proportion of red balls in urn 2 is unknown. There are four bets b_1, b_2, b_3, and b_4.

We take the sample space to be $S = \{RR, RB, BR, BB\}$, where $\gamma_1\gamma_2$ is the state in which a ball of color γ_i is drawn from urn i. The consequence matrix is

$$C^{\mathrm{R}} = \begin{array}{c} \\ b_1 \\ b_2 \\ b_3 \\ b_4 \end{array} \begin{array}{c} \begin{array}{cccc} RR & RB & BR & BB \end{array} \\ \begin{bmatrix} \$4000 & \$8000 & \$4000 & \$0 \\ \$4000 & \$4000 & \$8000 & \$0 \\ \$0 & \$8000 & \$4000 & \$4000 \\ \$0 & \$4000 & \$8000 & \$4000 \end{bmatrix} \end{array}.$$

Exploiting the symmetry between bets b_1 and b_4, and similarly, between bets b_2 and b_3, Machina argues that we might expect a decision-maker to exhibit the preferences $b_1 \sim b_4$ and $b_2 \sim b_3$. When comparing b_1 to b_2 he observes that although the events in which they yield the best (respectively, the worst) outcome are 'similarly' ambiguous, the event in which the middle outcome occurs is unambiguous for b_2 but not for b_1. Hence, one might argue that an ambiguity-averse decision-maker would strictly prefer b_2 to b_1.

The class of preferences Machina has in mind is Choquet expected utility. Recall from Example 2, preferences in this class are characterized by a Bernoulli utility function u and a subjective belief which is a capacity ν (or 'non-additive probability') defined over the set of subsets of S. Fixing a capacity ν, its conjugate, denoted $\overline{\nu}$, is the capacity given by $\overline{\nu}(E) = 1 - \nu(E^c)$.

Given a pair (u, ν) the Choquet expected utility of bet $b \in B$ is given by:

$$\sum_{z \in \mathbb{R}} [\nu(\{s : u(c_{sb}) \geqslant z\}) - \nu(\{s : u(c_{sb}) > z\})] \times z\,.$$

Correspondingly, we say that a pair (u, ν) generates the preference ordering $\succsim$ over the set of bets B whenever $b \succsim b'$ if and only if the Choquet expected utility of b is greater than or equal to the Choquet expected utility of b'.

We take $\mathcal{H}^{\mathrm{R}}$ to be the set of pairs (u, ν), where u is any Bernoulli utility function with $u(0) < u(4000) < u(8000)$ and ν is a capacity that along with its conjugate $\overline{\nu}$ satisfy the following 'natural' symmetry conditions:

$$\begin{aligned} &\nu(RR) = \nu(RB) = \nu(BR) = \nu(BB) > 0 \\ &\overline{\nu}(RR) = \overline{\nu}(RB) = \overline{\nu}(BR) = \overline{\nu}(BB) > 0 \end{aligned} \qquad .$$

We say a preference ordering $\succsim$ is (CEU-)admissible in the experiment $(C^{\mathrm{R}}, \mathcal{H}^{\mathrm{R}})$ if it can be represented by some $(u, \nu) \in \mathcal{H}^{\mathrm{R}}$. However, notice that for any $(u, \nu) \in \mathcal{H}^{\mathrm{R}}$, the Choquet expected utility of bet b_1 is:

$$\begin{aligned} &(\nu(RB) - \nu(\emptyset))\, u(8000) + (\nu(\{RR, RB, BR\}) - \nu(RB))\, u(4000) \\ &\quad + \ (\nu(S) - \nu(\{RR, RB, BR\}))u(0) \\ &= \nu(RB)u(8000) + (1 - \nu(RB) - (1 - \nu(\{RR, RB, BR\}))u(4000) \\ &\quad + \ (1 - \nu(\{RR, RB, BR\}))u(0) \\ &= \nu(RB)u(8000) + (1 - \nu(RB) - \overline{\nu}(BB))u(4000) + \overline{\nu}(BB)u(0)\,. \end{aligned}$$

Similarly, the (Choquet) expected utilities of the other three bets, b_2, b_3, b_4 are given by, respectively,

$$\nu(BR)u(8000) + (1 - \nu(BR) - \overline{\nu}(BB))u(4000) + \overline{\nu}(BB)u(0)\,,$$
$$\nu(RB)u(8000) + (1 - \nu(RB) - \overline{\nu}(RR))u(4000) + \overline{\nu}(RR)u(0)\,,$$
$$\nu(BR)u(8000) + (1 - \nu(BR) - \overline{\nu}(RR))u(4000) + \overline{\nu}(RR))u(0)\,.$$

Given the above equality constraints on any admissible capacity, it follows that the Choquet expected utilities of all four bets must be equal. Hence the only preference ordering that is admissible is the trivial one in which $b \sim^{\mathrm{T}} b'$, for all b, b' in B. Thus the preference ordering in which b_2 is strictly preferred to b_1 as suggested by Machina, is inadmissible.

However, the trivial preference relation is also admissible for the uniform probability distribution on S. Hence it follows from Theorem 1 that any inadmissible preference relation, including the one suggested by Machina, is not robustly inadmissible even restricting preferences to the smaller class of subjective expected utility maximizers. ■

For our final example in this section we analyze the "50:51 Example", the other main thought experiment presented in Machina (2009). We show the preference pattern Machina argues as being intuitively plausible for an ambiguity averse individual is not only inadmissible for CEU maximizers but it is also robustly inadmissible for SEU maximizers. However, since a coarsening of these preferences is admissible for some CEU maximizer, it follows from Theorem 1 that this preference pattern is not robustly (CEU-)inadmissible.

Example 7. The 50:51 Example from Machina (2009) The subject is presented with a single urn containing 101 balls. Fifty balls are marked with either 1 or 2, the other fifty-one balls are marked with either 3 or 4. Each ball is equally likely to be drawn. There are four bets b_1, b_2, b_3, b_4.

We take the sample space to be $S = \{1, 2, 3, 4\}$, where s is the event in which a ball marked with an s is drawn from the urn. The consequence matrix is

$$C^{\mathrm{M}} = \begin{array}{c} \\ b_1 \\ b_2 \\ b_3 \\ b_4 \end{array} \begin{array}{c} \begin{array}{cccc} 1 & 2 & 3 & 4 \end{array} \\ \begin{bmatrix} \$8000 & \$8000 & \$4000 & \$4000 \\ \$8000 & \$4000 & \$8000 & \$4000 \\ \$12000 & \$8000 & \$4000 & \$0 \\ \$12000 & \$4000 & \$8000 & \$0 \end{bmatrix} \end{array}.$$

An individual who prefers bets with known odds might well express the preference pattern $b_1 \succ b_2$ and $b_4 \succ b_3$.

As was the case in Example 6, the class of preferences Machina has in mind is Choquet expected utility. We take $\mathcal{H}^{\mathrm{M}}$ to be the set of pairs (u, ν), where u is any Bernoulli utility function with $u(0) < u(4000) < u(8000) < u(12000)$ and ν is a capacity that along with its conjugate $\overline{\nu}$ satisfy the following 'natural' symmetry conditions along with inequalities reflecting the known aspects of the urn's composition:

$$0 < \nu(1) = \nu(2) \leqslant \nu(3) = \nu(4)\,, \quad \nu(\{1,2\}) = \frac{50}{101}\,, \quad \nu(\{3,4\}) = \frac{51}{101}\,,$$
$$0 < \overline{\nu}(1) = \overline{\nu}(2) \leqslant \overline{\nu}(3) = \overline{\nu}(4)\,.$$

We say a preference ordering $\succsim$ is (CEU-)admissible in the experiment $(C^{\mathrm{M}}, \mathcal{H}^{\mathrm{M}})$ if it can be represented by some $(u,\nu) \in \mathcal{H}^{\mathrm{M}}$. However, by calculations similar to the ones detailed in Example 6, it follows that for any $(u,\nu) \in \mathcal{H}^{\mathrm{M}}$, the Choquet expected utility of bet b_1 is greater than or equal to the that of bet b_2 if and only if $\nu(\{1,2\}) \geqslant \nu(\{1,3\})$. But this is true if and only if the Choquet expected utility of bet b_3 is greater than or equal to that of bet b_4. Hence the preference pattern $b_1 \succ b_2$ and $b_4 \succ b_3$ is (CEU-)inadmissible. Moreover, it is robustly (SEU-)inadmissible since the only probability vector that satisfies the inequality constraints above is $p = \left(\frac{25}{101}, \frac{25}{101}, \frac{51}{202}, \frac{51}{202}\right)$. Thus the only preference orderings that are (SEU-)admissible must have $b_2 \succ_{\mathrm{SEU}} b_1$ and $b_4 \succ_{\mathrm{SEU}} b_3$.

However, notice that for any $(u,\nu) \in \mathcal{H}^{\mathrm{M}}$ with $\nu(\{1,2\}) = \nu(\{1,3\})$, the induced preference ordering $\succsim'$ has $b_1 \sim' b_2$ and $b_3 \sim' b_4$. Hence by Theorem 1 it follows that for the larger family of Choquet expected utility preferences, the preference pattern $b_1 \succ b_2$ and $b_4 \succ b_3$ is not robustly inadmissible. ■

5. Lottery Based Experiments

In some experiments, like the Allais common consequence and common ratio experiments, the design is presented effectively in terms of lotteries with objective probabilities. The experimenter then elicits preferences over these lotteries and checks for instance whether they can be represented by some class of admissible preferences defined on lotteries. These can be treated as a subclass of experiments in belief form since we shall see in Proposition 2, they always admit at least one rendition in our formulation, with a state space and a single admissible probability that induces those lotteries.

Suppose that we are given a finite set of simple lotteries $\mathcal{L} = \{\ell_1, \ell_2, \ldots, \ell_{|B|}\}$. We denote by $\underline{\ell_b}$ the (finite) support of lottery ℓ_b. We write $K = \{1, 2, \ldots, |K|\}$ as an index set for the set of consequences given by the set $\cup_{b\in B}\underline{\ell_b}$. The *lottery matrix* L induced by the lotteries is the $|B| \times |K|$ matrix

$$L = \begin{array}{c} \\ \ell_1 \\ \ell_2 \\ \vdots \\ \ell_{|B|} \end{array} \begin{array}{c} \begin{array}{cccc} c_1 & c_2 & \cdots & c_{|K|} \end{array} \\ \begin{bmatrix} \ell_{1c_1} & \ell_{1c_2} & \cdots & \ell_{1c_{|K|}} \\ \ell_{2c_1} & \ell_{2c_2} & \cdots & \ell_{2c_{|K|}} \\ \vdots & \vdots & \ddots & \vdots \\ \ell_{|B|c_1} & \ell_{|B|c_2} & \cdots & \ell_{|B|c_{|K|}} \end{bmatrix} \end{array},$$

where ℓ_{bc} is the probability assigned by lottery ℓ_b to consequence c, with $\ell_{bc} > 0$ if and only if $c \in \underline{\ell_b}$. Notice that a lottery matrix has no column of all zeros.

Definition 14. A *lottery based experiment* is a triple $(L\,,\mathcal{U}\,,\mathcal{V})$, where:

1. L is a lottery matrix induced from the set of lotteries $\mathcal{L}$;
2. $\mathcal{U}$ is a set of (Bernoulli) utility indices over K; and,
3. $\mathcal{V}$ is a set of functions from the space of simple lotteries over utilities to the real line, each member of which is continuous, respects first-order stochastic dominance, and normalized in the sense that for each degenerate lottery δ_α that yields α with probability one we have $V(\delta_\alpha) = \alpha$.

For any $u \in \mathcal{U}$ and any $b \in B$, let ℓ_b^u denote the lottery over utilities induced by u from the lottery ℓ_b. That is, $\ell_b^u(x) = \sum_{k \in K\,:\, u(c_{bk})=x} \ell_{bc_k}$.

A pair $(u, V) \in \mathcal{U} \times \mathcal{V}$ is said to *generate* the binary relation $\succsim$ on the set of lotteries $\mathcal{L}$ whenever $\ell_b \succsim \ell_{b'}$ if and only if $V(\ell_b^u) \geqslant V(\ell_{b'}^u)$. As was the case with experiments in belief form, this leads naturally to the notion of admissible preferences for a lottery-based experiment.

Definition 15. A binary relation $\succsim$ on $\mathcal{L}$ is *admissible* in $(L\,,\mathcal{U}\,,\mathcal{V})$, if there is a pair $(u, V) \in \mathcal{U} \times \mathcal{V}$ that generates $\succsim$. We refer to any binary relation $\succsim$ on $\mathcal{L}$ for which there is no pair in $\mathcal{U} \times \mathcal{V}$ that generates it as *inadmissible* in $(L\,,\mathcal{U}\,,\mathcal{V})$.

Recall, in the approach we have taken so far, a decision maker models the uncertainty she faces by a state space and hence takes the objects of choice to be acts, that is, mappings from that state space to a consequence space. So in order to assess the 'robustness' of a lottery based experiment we need first to find a corresponding experiment in belief form that can be viewed as *inducing* the lottery based experiment.

Definition 16. The lottery based experiment $(L\,,\mathcal{U}\,,\mathcal{V})$ is *induced* by the experiment in belief form $(C, \mathcal{H})$ within theory $\mathcal{T}$, if

1. $\underline{\ell_b} = \underline{C_b}$ for each $b = 1\,, \dots\,, |B|\,;$
2. for every $(u, V) \in \mathcal{U} \times \mathcal{V}$ there exists $(u\,, w) \in \mathcal{H}$ such that
$$\langle u \circ C_b\,, w(u \circ C_b)\rangle = V(\ell_b^u) \text{ for all } b = 1\,, \dots\,, |B|\,;$$
3. for every $(u, w) \in \mathcal{H}$ there exists $(u, V) \in \mathcal{U} \times \mathcal{V}$ such that
$$\langle u \circ C_b\,, w(u \circ C_b)\rangle = V(\ell_b^u) \text{ for all } b = 1\,, \dots\,, |B|\,.$$

The next proposition states that for any lottery based experiment we can 'reverse engineer' an experiment in belief form that induces it.

Proposition 2. Every lottery based experiment $(L\,,\mathcal{U}\,,\mathcal{V})$ can be induced by some experiment in belief form $(C, \mathcal{H})$ within some theory $\mathcal{T}$. Moreover, any other experiment in belief form $(C', \mathcal{H}')$ within the same theory $\mathcal{T}$ that induces $(L\,,\mathcal{U}\,,\mathcal{V})$ is a version of $(C, \mathcal{H})$.

In light of Proposition 2, we shall assess the robustness of a lottery-based experiment in terms of the robustness of (the equivalence class of) an experiment in belief form that induces it. We illustrate Proposition 2 for the Allais common consequence experiment. Moreover, we demonstrate that a lottery based experiment can be induced by multiple experiments in belief form with distinct sample spaces.

Allais Common Consequence Experiment

In the common consequence experiment (Allais, 1953, p. 527) the reader is asked to consider the following four lotteries $\{\ell_1, \ell_2, \ell_3, \ell_4\}$ over monetary consequences specified (in millions) as:

$$\ell_1(c) = \begin{cases} 1 & \text{if } c = \$1 \\ 0 & \text{if } c \neq \$1 \end{cases} \qquad \ell_2(c) = \begin{cases} \frac{10}{100} & \text{if } c = \$5 \\ \frac{89}{100} & \text{if } c = \$1 \\ \frac{1}{100} & \text{if } c = \$0 \\ 0 & \text{if } c \notin \{\$0, \$1, \$5\} \end{cases}$$

$$\ell_3(c) = \begin{cases} \frac{11}{100} & \text{if } c = \$1 \\ \frac{89}{100} & \text{if } c = \$0 \\ 0 & \text{if } c \notin \{\$0, \$1\} \end{cases} \qquad \ell_4(c) = \begin{cases} \frac{10}{100} & \text{if } c = \$5 \\ \frac{90}{100} & \text{if } c = \$0 \\ 0 & \text{if } c \notin \{\$0, \$5\} \end{cases}$$

Notice that $\underline{\ell_1} \cup \underline{\ell_2} \cup \underline{\ell_3} \cup \underline{\ell_4} = \{0, 1, 5\}$, thus the lottery based experiment representing this situation is the triple $(L, \mathcal{U}, \{\mathbb{E}\})$, where L is the 4×3 lottery matrix

$$L = \begin{bmatrix} 0 & 1 & 0 \\ \frac{1}{100} & \frac{89}{100} & \frac{1}{10} \\ \frac{89}{100} & \frac{11}{100} & 0 \\ \frac{9}{10} & 0 & \frac{1}{10} \end{bmatrix},$$

$\mathcal{U}$ is the set of all utility functions satisfying $u(0) < u(1) < u(5)$, and $\mathbb{E}$ is the standard expectation operator, that is, it returns the expected value of a lottery.

To construct one experiment in belief form that generates $(L, \mathcal{U}, \{\mathbb{E}\})$, take a sample space comprising just three elements $S = \{s_1, s_2, s_3\}$ with a corresponding consequence matrix

$$C = \begin{bmatrix} 1 & 1 & 1 \\ 5 & 0 & 1 \\ 1 & 1 & 0 \\ 5 & 0 & 0 \end{bmatrix}.$$

Set the unique admissible belief to be

$$p = \begin{bmatrix} p_1 \\ p_2 \\ p_3 \end{bmatrix} = \begin{bmatrix} \frac{1}{10} \\ \frac{1}{100} \\ \frac{89}{100} \end{bmatrix}.$$

The set of admissible parameters $\mathcal{H}$ is the set of pairs (u, p), where u is a utility function satisfying $u(0) < u(1) < u(5)$.[8]

Although the experiment in belief form above induces a lottery based experiment that corresponds to the Allais common consequence experiment, we stress that it is not the only experiment in belief form that does this. Furthermore, the sample spaces of alternative experiments in belief form will generally involve different correlation structures for the lotteries. The version above is one in which the correlation is greatest, allowing us to specify a sample space of minimal cardinality of three.

Another possible version has the four lotteries distributed independently as formulated by Loomes and Sugden (1982). In this case, the sample space S' needs at least twelve elements. An example of a consequence matrix with twelve states is:

$$C' = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 5 & 5 & 5 & 5 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 5 & 5 & 0 & 0 & 5 & 5 & 0 & 0 & 5 & 5 & 0 & 0 \end{bmatrix} .$$

The single admissible belief p' is the 12-dimensional column vector, in which

$$p'_s = \ell_1(c_{1s})\ell_2(c_{2s})\ell_3(c_{3s})\ell_4(c_{4s}) \, ,$$

for all the twelve states $s \in S'$.

As we shall now see the Allais lottery based experiment is not robust. The preference ordering $\succsim$ satisfying

$$\ell_1 \succ \ell_2 \succ \ell_4 \succ \ell_3$$

is inadmissible in both $(C, \mathcal{H})$ and $(C', \mathcal{H}')$. Thus it is inadmissible in the associated lottery based experiment $(L, \mathcal{U}, \{\mathbb{E}\})$.

Consider, however, the belief p^* on S', in which $p^*(s^*) = 1$, where s^* is the state in S' that has associated with it the vector of bet consequences

$$C'_{s^*} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 5 \end{bmatrix} .$$

By straightforward calculation it follows that for a utility function u' in which $u'(1) = \frac{10}{11}u'(5) + \frac{1}{11}u'(0)$ and a belief $\hat{p} = \frac{1}{2}p' + \frac{1}{2}p^*$ on S', we have

$$\begin{aligned} &\left\langle u' \circ C'_{\ell_1}, (1-\varepsilon)\, p + \varepsilon \hat{p} \right\rangle > \left\langle u' \circ C'_{\ell_2}, (1-\varepsilon)\, p + \varepsilon \hat{p} \right\rangle \\ &> \left\langle u' \circ C'_{\ell_4}, (1-\varepsilon)\, p + \varepsilon \hat{p} \right\rangle > \left\langle u' \circ C'_{\ell_3}, (1-\varepsilon)\, p + \varepsilon \hat{p} \right\rangle , \end{aligned}$$

[8] This formulation is closest to the way the example is presented in Allais (1953). It represents any experiment in which the design is essentially equivalent to the following. The subject is asked to consider the draw of a ball from an urn containing 100 balls numbered from 1 to 100, with (i) state s_1 corresponding to the event of the draw of a ball with any number from 1 to 10, (ii) state s_2 corresponding to the draw of the ball numbered 11, and (iii) state s_3 corresponding to the event of the draw of a ball with any number from 12 to 100.

for any $\varepsilon \in (0,1)$. Hence, $\succsim$ is ε-admissible in $(C', \mathcal{H}')$ which means it is not robustly inadmissible in $(C, \mathcal{H})$ and hence not robustly inadmissible in $(L, \mathcal{U}, \{\mathbb{E}\})$ either.

Furthermore, the coarser preference ordering $\succsim'$ satisfying

$$\ell_1 \sim' \ell_2 \succ' \ell_4 \sim' \ell_3$$

is admissible in $(C, \mathcal{H})$ as it is generated by the admissible pair (u', p). So applying Corollary 1.1 we conclude that $(C, \mathcal{H})$ is not robust and thus neither is the Allais common consequence lottery based experiment $(L, \mathcal{U}, \{\mathbb{E}\})$.

Not only does this establish that the Allais lottery based experiment is not robust but since there is no 'natural' experiment in belief form that induces it, one might argue it is even 'less robust' than the Ellsberg experiment for which his version was shown to be internally robust.

6. Robustifying an Experiment

The idea of robustification begins with a given non-robust thought experiment $(C, \mathcal{H})$ testing a theory $\mathcal{T}$ and a specific preference ordering $\succsim$ that is inadmissible. The experimenter has a well-reasoned intuition that participants could display these preferences. The first task of robustification is to modify the experiment so that the hypothesized preference ordering is robustly inadmissible. A more ambitious objective is to do so in a way such that the modified experiment is robust.

We begin by showing how both tasks may be accomplished for a class of experiments, including the classic Ellsberg and Allais, involving two pairs of bets. This is achieved by a slight modification of one of the consequences of one of the bets.

Next, we consider participant based approaches in which experiments are modified for each participant to guarantee that, if a participant's elicited answers imply preferences that are inadmissible, then those preferences are robustly inadmissible in that personalized experiment. Such approaches have been employed by (amongst others) Halevy (2007) and Binmore et al. (2012).

6.1. Robustification of the Four Bet Experiment

Important characteristics of preferences may be expressed in terms of consistency requirements. Examples include the sure-thing principle in subjective expected utility, the independence axiom in expected utility, and CARA in risk theory. Experiments designed to test these consistency requirements directly will often give rise to inadmissible preferences that are not robustly so. This may be explained as follows.

Suppose that in some restriction of *a general MBA model* there are four bets for which the following consistency property holds,

$$b_1 \succsim b_2 \iff b_3 \succsim b_4 \text{, for all } (u, w) \in \mathcal{H} .$$

One approach to testing this consistency property is to design an experiment $(C, \mathcal{H})$ that includes choice pairs of this kind, for which the experimenter has an intuition that participants could display the inadmissible preference pattern $b_1 \succ b_2$ and $b_4 \succ b_3$ as in the Ellsberg one-urn example from Section 1, Example 4 from Section 4 and the Allais common consequence experiment from Section 5.

To the extent that the experimenter's intuitions about participants' responses are correct, experiments of this kind are likely to elicit preference orderings that violate the consistency requirement being tested and therefore inadmissible. However, since all such orderings admit a parameter $(u, w) \in \mathcal{H}$ that induces indifference between b_1 and b_2 (and hence, by the consistency property, between b_3 and b_4, as well), this pattern is not robustly inadmissible.

Since any u is an order-preserving transformation of the consequences, we can exploit the monotonicity of the induced preferences to robustify experiments of this kind. Consider a perturbed consequence matrix $\hat{C} \neq C$, in which $C_{b_1} \gg \hat{C}_{\hat{b}_1}$ and $\hat{C}_{\hat{b}_i} = C_{b_i}$ for $i = 2, 3, 4$.[9] Notice that for any $(u, w) \in \mathcal{H}$,

$$\langle u \circ C_{b_1}, w\left(u \circ C_{b_1}\right) \rangle > \left\langle u \circ \hat{C}_{\hat{b}_1}, w\left(u \circ \hat{C}_{\hat{b}_1}\right) \right\rangle .$$

Hence monotonicity and the consistency property together imply

$$\hat{b}_1 \succsim \hat{b}_2 \Longrightarrow b_1 \succ b_2 \iff b_3 \succ b_4 \iff \hat{b}_3 \succ \hat{b}_4 \text{, for all } (u, w) \in \mathcal{H} .$$

Thus for the inadmissible preference ordering $\hat{b}_1 \succ \hat{b}_2$ and $\hat{b}_4 \succ \hat{b}_3$, no coarsening is admissible. So by Theorem 1 the ordering is robustly inadmissible. One can also readily check using Corollary 1.1 that the modified experiment is robust.

Returning to Example 4, consider the perturbed consequence matrix

$$\hat{C} = \begin{array}{c} \\ \hat{b}_1 \\ \hat{b}_2 \\ \hat{b}_3 \\ \hat{b}_4 \end{array} \begin{array}{c} \begin{array}{cc} H & T \end{array} \\ \begin{bmatrix} 4 & 4 \\ 2 & 10 \\ 55 & 55 \\ 52 & 60 \end{bmatrix} \end{array},$$

obtained by reducing the consequence of bet b_1 in each state from 5 to 4. The choice pattern $\hat{b}_1 \succ \hat{b}_2$ and $\hat{b}_4 \succ \hat{b}_3$ is now robustly inadmissible for CARA preferences.

In this robustification as well as the ones that follow, we should be mindful that the perturbations in rewards should be significant enough to motivate the participants. If they are too small, then a violation might reflect a mistake or a misperception or a lack of motivation on the part of the decision-maker.[10]

Notice that there is a price to pay. In the original experiment, the pattern $b_2 \succ b_1$ and $b_3 \succ b_4$ was also inadmissible. In the new experiment the preference

[9]For any two vectors $x, y \in \mathbb{R}^n$ we write $x \gg y$ for the component-wise strict inequality, that is, $x \gg y \iff x_i > y_i$ for all i.

[10]This concern is raised and discussed in more detail by Smith (1982) and Harrison (1994).

ordering $\hat{b}_3 \succ \hat{b}_4 \succ \hat{b}_2 \succ \hat{b}_1$ is admissible. That is, the new experiment is a 'one-sided' instead of a 'two-sided' test. The acceptability of this trade-off depends on the strength of the intuition that the preference pattern $b_1 \succ b_2$ and $b_4 \succ b_3$ is more plausible than the reverse. In the case of Example 4, CARA is likely to be rejected in favor of DARA but not in favor of IARA (increasing absolute risk aversion).

6.2. Participant Based Robustness

Every subject in an experiment in belief form $(C, \mathcal{H})$ designed to test some theory $\mathcal{T}$ can be taken to have a single weighting function w. If w is an admissible belief, that is, $(u, w) \in \mathcal{A}$ for some utility function u, then that subject can be viewed as participating in a 'personalized' lottery based experiment. So for an experiment $(C, \mathcal{H})$ and each admissible belief w, let $\mathcal{H}^w = \{(u', w') \in \mathcal{H} \colon w' = w\}$. The pair $(C, \mathcal{H}^w)$ is an experiment in belief form that induces a lottery based experiment.

The next result states that an experiment in belief form for generalized mean-utility maximizers is robust if and only if each admissible personalized lottery based experiment is robust.

Proposition 3. An experiment $(C, \mathcal{H})$ testing theory $\mathcal{T}$ is robust if and only if the 'personalized' experiment $(C, \mathcal{H}^w)$ testing $\mathcal{T}$ is robust for every admissible decision weight function w.

So far we have considered an experiment $(C, \mathcal{H})$ for which the experimenter selects a sample of participants from the population that is being studied. The robustness of this experiment guarantees that any inadmissible elicited preference ordering from any participant is robustly inadmissible. Consider an individual participant i. If the experimenter elicits a preference ordering $\succsim^i$ over the bets, and it is inadmissible but weakly admissible in the experiment $(C, \mathcal{H})$, then the experimenter may not have enough information about the preferences of the particular participant to rule out the possibility that $\succsim^i$ is arising from a small perturbation in the expected utility maximizer's beliefs.

To see this, we consider by way of example, the case where a participant i is a subjective expected utility maximizer satisfying the restriction of subjective expected theory being tested. This participant i has a belief p^i over S and a utility function u^i, which represent her preferences but which are not known. From the perspective of the experimenter, i participates in the trivial experiment $(C, \mathcal{H}^i)$ where $\mathcal{H}^i = \{(u^i, p^i)\}$. Because she has just one belief, effectively this participant faces the lottery-based experiment $(L_{(C,\mathcal{H}^i)}, \{u^i\}, \{\mathbb{E}\})$.

The elicited preferences are inadmissible in this lottery based experiment as they are inadmissible in $(C, \mathcal{H})$. Importantly, although $\succsim^i$ may be weakly admissible in $(C, \mathcal{H})$, if the experimenter can establish that $(L_{(C,\mathcal{H}^i)}, \{u^i\}, \{\mathbb{E}\})$ is robust, then the experimenter is able to conclude that preferences $\succsim^i$ are not due to a small perturbation in the subjective beliefs of the particular participant i.

The experimenter, however, does not know the parameter (u^i, p^i) of the $(C, \mathcal{H}^i)$ and cannot determine these simply by the information given by the preferences $\succsim^i$. So the experimenter will want to gather additional information from the participant enabling the experimenter to make the personalized experiment robust.

We consider two different methods. The first method elicits certainty equivalents of each bet for the participant. The second, adjusts bets on the basis of the participant's initial choices and then elicits a second-round of choices. This process may be repeated to get more precise estimates of the participant's parameter.

6.2.1. Eliciting Certainty Equivalents

This was the procedure used in Section 1 to robustify the Ellsberg Single Urn example. Fix an experiment in belief form $(\mathcal{C}, \mathcal{A})$. The general procedure can be described as follows.

For each participant i:

(i) Elicit certainty equivalents for each bet using an appropriate incentivized procedure (such as the BDM mechanism (Becker et al., 1964)).

(ii) Elicit preferences over the original bets.

In the event that the elicited preferences over the bets and their certainty equivalents are inadmissible, either *(a)* the strict inequality between any pair of certainty equivalents corresponds to a strict preference between the corresponding pair of bets; or *(b)* not.

Case *(b)* constitutes a violation of transitivity. Furthermore, by range dominance there is no admissible parameter which will make these certainty equivalents indifferent. Thus by Theorem 1 this violation is robust.[11]

For case *(a)* it also follows from range dominance that any admissible coarsening must respect the strict inequalities between any pair of certainty equivalents. Thus again there is no admissible coarsening which means that by Theorem 1 the inadmissible preference relation is robustly inadmissible.

Since the inadmissible preference relation was arbitrary this means by Corollary 1.1 the experiment that has been "personalized" by adding these certainty equivalents is now robust.

6.2.2. Multi-round Preference Elicitation with Choice-based Adjustments to the Bets

In an Ellsberg experiment where there are only two consequences, one need only estimate the participant i's beliefs p^i setting the range of u^i as $\{0,1\}$, then checking for the robustness of $(C, \mathcal{H}^i)$.

[11] This situation essentially corresponds to the inadmissible preferences that are seen in the well-known *preference reversal* (Holt, 1986) experiments in which certainty equivalents are elicited for two bets, a so-called P-bet and a \$-bet, but the preference expressed between these two bets contradicts the one implied by the relative size of the two certainty equivalents. That is, the preference reversal experiments are robust in our sense.

For the special case of an Ellsberg experiment Binmore et al. (2012) employed the following procedure for each participant i:

1. Estimate the implied beliefs p^i of each participant by iteratively adding and subtracting balls in the experiment, eliciting the participant i's preferences, and stopping the iteration when i switches her strict preferences over two bets. Coupled with the contrapositive assumption of expected utility maximization, this switching (taken as representing indifferences) gives an estimate $\hat{p}^i$ of p^i.

2. Use the information given by $\hat{p}^i$ to robustify the personalized lottery-based experiment by means of a perturbation of the induced lotteries.

3. Elicit preferences in the robustified experiment.

7. Concluding Comments

Experiments like those proposed by Allais and Ellsberg have been influential in the development of alternatives to, and generalizations of, (subjective) expected utility theory. However, it has often been suggested that the apparently anomalous results of these experiments may result from sensitivity to small errors in decisions or deviations between the perceptions of the subject and those assumed by the experimenter.

The central task of this paper has been to formulate a rigorous definition of robustness for experiments and for observed choices that are inadmissible for a class of preferences under consideration. Most commonly, this is the class of expected utility preferences, but the method is equally applicable to tests of such hypotheses as constant absolute risk aversion as well as testing various generalizations of subjective expected utility.

The core result is that if inadmissible preferences can be made admissible by coarsening (that is, by one or more conversions of strict preference to indifference), then the inadmissibility is not robust.

Appendix

Let $B = \{b_1, b_2, \ldots, b_{|B|}\}$ be a finite set of bets, each associated with a finite consequence set $K_b \subseteq \mathcal{C}$. Let T be set the of all functions $t : B \to \bigcup_{b \in B} K_b$ satisfying $t(b) \in K_b$ for all $b \in B$.

Definition 17. A *canonical consequence matrix* based on the set of bets B and profile of consequences $(K_b)_{b \in B}$ is a $|B| \times |S|$ consequence matrix C satisfying:

1. for each $t \in T$, there is an $s \in S$ such that $t(b) = c_{bs}$ for all $b \in B$;

2. $\underline{C_b} = K_b$ for each $b \in B$.

Condition (1) of Definition 17 guarantees that every function in $t \in T$ is represented by some state $s \in S$. Condition (2) ensures that the consequence matrix respects the consequence set K_b for each bet $b \in B$.

Lemma 1. Let C be a canonical consequence matrix based on the set of bets B and profile of consequences $(K_b)_{b\in B}$. For each $b \in B$, let ℓ_b be any lottery over K_b. The probability distribution p on S defined by $p_s = \prod_{b\in B} \ell_b(c_{bs})$ for all $s \in S$ induces the lottery ℓ_b over K_b for each $b \in B$.

Proof. Notice that for each $b \in B$ and each $c \in \underline{C}_b = K_b$, the probability assigned to the lottery induced by p^L is equal to:

$$\begin{aligned}
&\sum_{s:\, c_{bs}=c} p_s \\
&= \ell_b(c) \left(\sum_{s:\, c_{bs}=c} \ell_1(c_{1s})\ell_2(c_{2s}) \ldots \ell_{b-1}(c_{(b-1)s})\, \ell_{b+1}(c_{(b+1)s}) \ldots \ell_B(c_{Bs}) \right) \\
&= \ell_b(c) \prod_{b'\neq b} \left(\sum_{k\in K} \ell_{b'}(k) \right) \\
&= \ell_b(c)\,,
\end{aligned}$$

as required. ■

Lemma 2. Let $(C, \mathcal{H})$ be an experiment in belief form where C is a canonical consequence matrix. If $\succsim$ is range dominant with respect to a parameter $(u, w) \in \mathcal{T}(\underline{C}\,, S)$, then $\succsim$ is generated by an SEU-parameter (u, p) with the same u.

Proof. Suppose $\succsim$ is range dominant with respect to (u, w). We use Lemma 1 to construct a probability distribution p on S such that the SEU-parameter (u, p) represents $\succsim$.

For each bet $b \in B$ we define the best outcome $\overline{u}_b = \max_{s\in S} u(c_{bs})$ and the worst outcome $\underline{u}_b = \min_{s\in S} u(c_{bs})$.

We partition B into equivalence classes $B_1, ..., B_k$ of bets from worst to best, that is, if $i < j$, then for each $b \in B_i$ and $b' \in B_j$ we have $b' \succ b$. For each equivalence class B_i, we define the worst of the best outcomes over bets in that class by $\overline{v}_i = \min_{b\in B_i} \overline{u}_b$, and the best of the worst outcomes $\underline{v}_i = \max_{b\in B_i} \underline{u}_b$.

Since $\succsim$ is range dominant with respect to (u, w), we obtain the following two results:

1. $\underline{v}_i < \overline{v}_j$ whenever $i < j$;
2. $\underline{v}_i \leqslant \overline{v}_i$ for all $i = 1, ..., k$.

Results (1) and (2) follow from parts (1) and (2) respectively of Definition 13.

We inductively define a value v_i to each equivalence class as follows:

$$v_1 = \underline{v}_1;$$

and for $i > 1$,

$$v_i = v_{i-1} + \frac{\underline{v}_{i-1} + \overline{v}_i}{2}.$$

Observe that v_i is strictly increasing in i and that for each class i, and each bet $b \in B_i$,

$$\underline{u}_b \leqslant \underline{v}_i \leqslant v_i \leqslant \overline{v}_i \leqslant \overline{u}_b.$$

Hence, for each bet $b \in B_i$, there is a lottery on the best and worst outcome for that bet that yields value v_i in expectation.

By Lemma 1, we can find a probability distribution p on S that induces these same lotteries. Hence, $\succsim$ is generated by the SEU-parameter (u, p). ■

Proof of Proposition 1. The proposition simply translates definitions and the proof is an immediate consequence of these. ■

Proof of Theorem 1. Let $(C, \mathcal{H})$ be an experiment in belief form testing a theory $\mathcal{T}$ where B and S denote the set of bets and the sample state space.

(If-part): We prove the contra-positive of the if-part. Suppose that $\succsim$ is weakly admissible in $(C, \mathcal{H})$. Then, $\succsim$ is ε-admissible in some version $(C', \mathcal{H}')$ of $(C, \mathcal{H})$. We let S' denote the sample state space of $(C', \mathcal{H}')$. Since $\succsim$ is ε-admissible in $(C', \mathcal{H}')$, there is an admissible parameter $(u, w) \in \mathcal{H}'$ and a parameter $(u', w') \in \mathcal{T}(\underline{C}', S')$ such that for all $\varepsilon \in (0, 1)$, the parameter $(u, (1 - \varepsilon)w + \varepsilon w')$ is in $\mathcal{T}(\underline{C}', S')$ and generates $\succsim$.

Since $\mathcal{T}$ is a generalized mean decision theory, $\succsim$ must be range dominant with respect to the parameter $(u, (1-\varepsilon)w + \varepsilon w')$. Since $(C', \mathcal{H}')$ and $(C, \mathcal{H})$ are versions, $\succsim$ will also be range dominant with respect to the admissible parameter $(u, w) \in \mathcal{H}'$.

Let $\succsim'$ be the preference ordering generated by the admissible parameter (u, w) in $\mathcal{H}'$. We show that $\succsim$ is finer than $\succsim'$. It suffices to show that $b \succ' b'$ implies $b \succ b'$. Since $b \succ' b'$, we know that $\langle u \circ c_b, w(u \circ c_b) \rangle > \langle u \circ c_{b'}, w(u \circ c_{b'}) \rangle$. So for all ε sufficiently close to zero,

$$\begin{aligned}&\langle ((1-\varepsilon)u \circ c_b + \varepsilon u' \circ c_b), ((1-\varepsilon)w(u \circ c_b) + \varepsilon w(u' \circ c_b)) \rangle > \\ &\qquad \langle ((1-\varepsilon)u \circ c_{b'} + \varepsilon u' \circ c_{b'}), ((1-\varepsilon)w(u \circ c_{b'}) + \varepsilon w(u' \circ c_{b'})) \rangle \ .\end{aligned}$$

Thus, $b \succ b'$.

(Only-if-part): We prove the contra-positive of the only-if-part. Suppose that $\succsim$ is range dominant with respect to some admissible parameter $(u, w) \in \mathcal{H}$ and it is finer than the preference ordering $\succsim'$ generated by (u, w). We will show that $\succsim$ is ε-admissible in a canonical version $(C', \mathcal{H}')$ of $(C, \mathcal{H})$.

To this end, let C' be any $|B| \times |S'|$ canonical consequence matrix based on B and $(\underline{C}_b)_{b \in B}$ that satisfies $S \subseteq S'$. Since the theory is coherent, for every parameter $(u, w) \in \mathcal{H} \subseteq \mathcal{T}(\underline{C}, S)$, there is a parameter $(u, w') \in \mathcal{T}(\underline{C}', S')$ that represents the same preferences as (u, w). Let $\mathcal{H}'$ be a selection of such parameters in $\mathcal{T}(\underline{C}', S')$, one for each $(u, w) \in \mathcal{H}$. By construction $(C', \mathcal{H}')$ is a version of $(C, \mathcal{H})$.

Since, by assumption, $\succsim$ is range dominant with respect to the admissible parameter $(u, w) \in \mathcal{H}$ it is also range dominant with respect to any admissible parameter $(u, w') \in \mathcal{H}'$, one of which must exist by our construction of $\mathcal{H}'$. It follows, by Lemma 2 that $\succsim$ is generated by an SEU-parameter (u, p) where $p \in \Delta^{|S'|-1}$.

Let (u, w') be the parameter in $\mathcal{H}'$ that corresponds to the parameter $(u, w) \in \mathcal{H}$ that generates $\succsim'$. Since the theory is linearly open and $\succsim$ is finer than $\succsim'$, we can find a $\delta \in (0, 1)$ such that $\succsim$ is generated by the parameter $(u, w'') \in \mathcal{T}(\underline{C}', S')$ where $w'' = (1-\delta)w' + \delta\overline{w}^p$. It follows that for any for any $\varepsilon \in (0, 1)$, the parameter $(u, (1-\varepsilon)w' + \varepsilon w'')$ is in the theory $\mathcal{T}(\underline{C}', S')$ and also generates $\succsim$. ■

Proof of Corollary 1.1. This proof follows from Theorem 1 and the definition of robust experiment. ■

Proof of Proposition 2. First we need to construct an experiment in belief form $(C, \mathcal{H})$ testing a theory $\mathcal{T}$ that induces the lottery based experiment $(L, \mathcal{U}, \mathcal{V})$.

First construct a canonical consequence matrix as in Lemma 1 which ensures that (1) of Definition 16 holds. Fix a pair (u, V) in $\mathcal{U} \times \mathcal{V}$. For each lottery ℓ_b, recall that ℓ_b^u is the associated induced lottery over utilities. Similarly, for each bet b, $u \circ C_b$ is the induced state contingent utility vector associated with that bet.

For any $u \circ C_b$ that is a constant vector set $w_s(u \circ C_b) = \frac{1}{|S|}$, for all $s \in S$. For any non-constant state contingent utility vector and any state $s \in S$ set:

$$w_s(u \circ C_b) = \begin{cases} \frac{\alpha}{\left|\operatorname{argmin}_{t \in S} u(c_{bt})\right|} & \text{if } s \in \operatorname{argmin}_{t \in S} u(c_{bt}) \\ \frac{1-\alpha}{\left|\operatorname{argmax}_{t \in S} u(c_{bt})\right|} & \text{if } s \in \operatorname{argmax}_{t \in S} u(c_{bt}) \end{cases},$$

where $\alpha \in (0, 1)$ is the unique solution to:

$$V(\ell_b^u) = \alpha \min_{t \in S} u(c_{bt}) + (1-\alpha) \max_{t \in S} u(c_{bt}).$$

The fact that α is well defined follows from the three properties V is assumed to exhibit: respecting first order stochastic dominance, continuity and normalization. Observe that for for any b we have $\langle u \circ C_b, w(u \circ C_b) \rangle = V(\ell_b^u)$.

Let $\mathcal{H}$ be the set of (u, w) defined as above and consider the grand theory $\mathcal{T}^{\text{GM}}$. By construction, the lottery based experiment $(L, \mathcal{U}, \mathcal{V})$ is induced by the experiment in belief form $(C, \mathcal{H})$ testing the null hypothesis $\mathcal{H}$ in the theory $\mathcal{T}^{\text{GM}}$.

Moreover, let $(C', \mathcal{H}')$ be an experiment in belief form within theory $\mathcal{T}$ that induces $(L, \mathcal{U}, \mathcal{V})$. Clearly the bets are the same, and by (1) of Definition 16, the set of consequences for each bet is the same, and by (2) and (3) of Definition 16 the sets of admissible preferences are the same. ■

Proof of Proposition 3. This follows from the definitions of robust experiment and personalized experiment. ■

References

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

Becker, G., Groot, M.D., Marshak, J., 1964. Measuring utility by a single-response sequential method. Behavioral Science , 226–232.

Binmore, K., Stewart, L., Voorhoeve, A., 2012. How much ambiguity aversion? Finding indifferences between Ellsberg's risky and ambiguous bets. Journal of Risk and Uncertainty 45, 215–238.

Cerreia-Vioglio, S., Ghirardato, P., Maccheroni, F., Marinnacci, M., Siniscalchi, M., 2011. Rational preferences under uncertainty. Economic Theory 48, 341–375.

Eichberger, J., Grant, S., Kelsey, D., 2007. Updating Choquet beliefs. Journal of Mathematical Economics 43, 888–899.

Ellsberg, D., 1961. Risk, ambiguity, and the Savage axioms. The Quarterly Journal of Economics 75, 643–669.

Halevy, Y., 2007. Ellsberg Revisited: An Experimental Study. Econometrica , 503–536.

Harrison, G.W., 1994. Expected utility theory and the experimentalists. Empirical Economics 19, 223–253.

Holt, C., 1986. Preference reversals and the independence axiom. American Economic Review , 508–513.

Johnson, C., Baillon, A., Bleichrodt, H., Li, Z., van Dolder, D., Wakker, P., 2015. PRINCE: An improved method for measuring incentivized preferences. Mimeo.

Kadane, J., 1992. Healthy scepticism as an expected-utility explanation of the phenomena of Allais and Ellsberg. Theory and Decision 32, 57–64.

Loomes, G., Sugden, R., 1982. Regret theory: an alternative theory of rational choice under uncertainty. The Economic Journal 92, 805–824.

Machina, M., 2009. Risk, ambiguity, and the rank-dependence axioms. American Economic Review 99, 385–392.

von Neumann, J., Morgenstern, O., 1944. Theory of Games and Economic Behavior. Princeton University Press.

Raiffa, H., 1961. Risk, ambiguity, and the Savage axioms: comment. Quarterly Journal of Economics 75, 690–694.

Savage, L.J., 1954. The Foundations of Statistics. John Wiley & Sons.

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

Smith, V.L., 1982. Microeconomic systems as an experimental science. American Economic Review 72, 923–955.