Random Choice and Private Information*

Jay Lu†

August 2014

Abstract

We consider an agent who chooses from a set of options after receiving some private information. This information however is unobserved by an analyst, so from the latter's perspective, choice is probabilistic or *random*. We provide a theory in which information can be fully identified from random choice. In addition, the analyst can perform the following inferences even when information is unobservable: (1) directly compute ex-ante valuations of option sets from random choice and vice-versa, (2) assess which agent has better information by using choice dispersion as a measure of informativeness, (3) determine if the agent's beliefs about information are dynamically consistent, and (4) test to see if these beliefs are well-calibrated or rational.

* I am deeply indebted to both Faruk Gul and Wolfgang Pesendorfer for their continuous advice, guidance and encouragement. I am very grateful to Stephen Morris for some highly valuable and insightful discussions. I would also like to thank Roland Bénabou, Simone Cerreia-Vioglio, Sylvain Chassang, Mark Dean, Leandro Gorno, Doron Ravid and seminar audiences at Yale University, University of California Los Angeles, Northwestern University, Cornell University, Massachusetts Institute of Technology, University of British Columbia, Boston University and the University of Pennsylvania for helpful suggestions and constructive feedback.

† Department of Economics, Cornell University; `jay.lu@cornell.edu`

1 Introduction

In many economic situations, an agent's private information is not observable. We provide a theory for identifying this information from choice data. To be concrete, consider an agent choosing from a set of health insurance plans at the beginning of every year. Before choosing, she receives some private information (i.e. a signal) that influences her beliefs about her health for the rest of the year. For example, she may obtain a health exam that informs her about her likelihood of falling sick. As a result, her choice of an insurance plan depends on the realization of her signal that year. An analyst (i.e. an outside observer) knows the agent's set of health plans but does not observe her signal. Hence, from the analyst's perspective, the agent's choice of health insurance every year is probabilistic or *random*. This is captured by a frequency distribution of choices over time. Call this the *individual interpretation* of random choice.

Alternatively, consider a group of agents choosing from the same set of health insurance plans in a given year. Before choosing, each agent in the group has some private information about her health. For example, she may have some personal knowledge about her lifestyle that affects her choice of insurance. This information however, is beyond what is captured by all the characteristics observable by the analyst. To the analyst, agents are observationally identical.[1] Hence, from the analyst's perspective, the choice of health insurance within the group is probabilistic or *random*. This is captured by a frequency distribution of choices over agents in the group. Call this the *group interpretation* of random choice.

In both the individual and group interpretations, the information affecting an agent's choice is *strictly private* i.e. it is completely unobservable to the analyst. Many problems in information economics and decision theory fit in this framework. These include consumers choosing a retail banking service, private investors deciding on an investment, buyers bidding at a Treasury auction and users clicking on an online ad. In all these examples, it is likely that an analyst is unable to directly observe an agent's private information or has difficulty discerning how that information will be interpreted.

In this paper, we provide a theory where private information can be completely identified from random choice.[2] We also provide tools for performing some of the standard exercises of

[1] We can think of this group as the end result after applying all possible econometric (both parametric and non-parametric) analysis available to differentiate the observable data.

[2] An alternative approach is to directly elicit private information from survey data (for example, see Finkelstein and McGarry [20] and Hendren [27]). However, respondents may not accurately report their true beliefs or the data may be subject to other complications (such as excess concentrations at focal points). In

inference. First, the analyst can use random choice to directly compute ex-ante valuations of option sets (i.e. ex-ante utility in the individual interpretation or welfare in the group interpretation) and vice-versa. Call this *evaluating option sets*. Second, the analyst can assess which agent (or group of agents) has better information by using choice dispersion as a measure of informativeness. Call this *assessing informativeness*. Third, if valuations of option sets are known, then the analyst can compare them with random choice to detect when beliefs about information are dynamically inconsistent. Call this *detecting biases*. Finally, from the joint distribution of choices and payoff-relevant outcomes, the analyst can determine if beliefs are well-calibrated or rational. Call this *calibrating beliefs*.

When information is observable, the above inferences are important and well-understood exercises in information theory and information economics. We demonstrate how to carry out the same analysis even when information is *not* directly observable and can only be inferred from choice behavior. Our theorems reveal that when all relevant choice data is available, all these inferences can be performed just as effectively as in the case with observable information. The more practical question of drawing inferences when choice data is only partially available is left for future research.

Formally, each choice option corresponds to a state-contingent *act*. For example, a high-deductible health plan corresponds to the act that yields a high (low) payoff if the agent is healthy (sick). Call a set of acts a *decision-problem*. The timing is as follows. At time 0, the agent faces a decision-problem. At time 1, she receives some private information that informs her about the likelihood of the payoff-relevant state. At time 2, she evaluates each act based on its subjective expected utility and chooses the best act from the decision-problem. Since the agent's private information is unobservable to the analyst, the only observable choice data is a *random choice rule (RCR)* that specifies a choice distribution over acts for each decision-problem.

We characterize a random utility maximization (RUM) model where the utilities are subjective expected utilities.[3] In the individual interpretation, each realization of the random utility corresponds to a realization of the agent's private signal. In the group interpretation, each realization of the random utility corresponds to a random draw of an agent in the group. In both interpretations, the probability that an act is chosen is equal to the probability

contrast, our approach follows the original spirit of Savage [43] by inferring beliefs from choice behavior.

[3] For more about random utility maximization, see Block and Marschak [7], Falmagne [18], McFadden and Richter [35] and Gul, Natenzon and Pesendorfer [25].

that it attains the highest subjective expected utility in the decision-problem. Call this an *information representation* of the RCR.

In general, RUM models have difficulty dealing with indifferences in the random utility. We address this issue by drawing an analogy with deterministic choice. If two acts are indifferent (i.e. they have the same utility) under deterministic choice, then the model is silent about which act will be chosen. Similarly, under random choice, if two acts are indifferent (i.e. they have the same random utility), then the model is silent about what the choice probabilities are. This approach has two advantages. First, it allows the analyst to be agnostic about choice data that is beyond the scope of the model and provides some additional freedom to interpret data. Second, it allows for just enough flexibility so that we can include deterministic choice as a special case of random choice. In particular, the subjective expected utility model of Anscombe-Aumann [2] obtains as a degenerate case.

We first provide an axiomatic characterization of information representations. The first four axioms (*monotonicity*, *linearity*, *extremeness* and *continuity*) are direct translations of the random expected utility axioms from Gul and Pesendorfer [26]. Three new axioms are introduced. *Non-degeneracy* ensures there is no universal indifference. *C-determinism* ensures deterministic choice over constant acts, that is, acts that yield the same payoff in every state. As information only affects beliefs (not tastes), constant acts yield the same payoff regardless of beliefs so choice must be deterministic over them. Finally, *S-monotonicity* ensures that acts that dominate in every state are chosen for sure. It is the random choice version of the state-by-state monotonicity condition under deterministic choice. Theorem 1 asserts that a RCR has an information representation if and only if it satisfies these seven axioms. Theorem 2 asserts that analyzing binary decision-problems is sufficient for identifying the agent's private information. We thus provide a choice-based theory for information.

We then introduce a key technical tool that will feature prominently in our subsequent analysis. Given a decision-problem, consider the addition of an enticing *test act* (for example, a fixed payoff). As the value of the test act decreases, the probability that some act in the original decision-problem will be chosen over the test act will increase. Call this the *test function* for the decision-problem. Test functions are cumulative distribution functions that characterize the utility distributions of decision-problems. They serve as sufficient statistics for identifying information.

Following, we address our main questions of inference. First, we *evaluate option sets*. In the individual interpretation, the valuation of an option set is the ex-ante utility of the set before any information is received. In the group interpretation, the valuation of an option set is the welfare or total utility of the set for all agents in the group.[4] Given random choice, is there a method to directly compute valuations? Theorem 3 shows that computing integrals of test functions recovers valuations. Conversely, Theorem 4 shows that computing the marginal valuations of decision-problems with respect to test acts recovers random choice. These operations are mathematical inverses of each other; even when information is not observable, the analyst can directly compute valuations from random choice and vice-versa. This provides a methodology for elicitation that is similar to classical results from consumer and producer theory (Theorem 4 for example is the random choice analog of Hotelling's Lemma).

Next, we *assess informativeness*. In the classical approach of Blackwell [5, 6], better information is characterized by higher ex-ante valuations. Theorem 5 shows that under random choice, better information is characterized by second-order stochastic dominance of test functions. Given two agents (or two groups of agents), one is better informed than the other if and only if test functions under the latter second-order stochastic dominate those of the former. This equates an unobservable multi-dimensional ordering of information with observable single-dimensional stochastic dominance rankings. Intuitively, a more informative signal structure (or more private information in a group of observationally identical agents) is characterized by greater dispersion or randomness in choice. For example, in the special case where information corresponds to events that partition the state space, better information is exactly characterized by less deterministic choice.

We then apply these results to *detect biases*. Suppose ex-ante valuations of option sets are observable via a preference relation. Can the analyst detect when this preference relation is inconsistent with random choice? In the individual interpretation, this describes a form of dynamic inconsistency where the time-0 preference relation suggests a more (or less) informative signal than that implied by time-2 random choice. Call this *prospective overconfidence* (or *underconfidence*). An example of the former is the *diversification bias* where agents initially prefer large option sets but subsequently always choose the same option. An example of the latter is the *confirmation bias* where beliefs becomes more extreme (i.e. dispersed) after agents receive their signals. Both exhibit *subjective misconfidence*. These biases also apply in the group interpretation. For example, consider a firm that chooses sets

[4] McFadden [34] calls this the "social surplus".

of health plans based on total employee welfare (i.e valuation). Inconsistency now suggests that the firm has an incorrect assessment of the distribution of beliefs among its employees. By studying both valuations and random choice, the analyst can detect these biases even when information is not directly observable.

Finally, we *calibrate beliefs*. As we have adopted a subjective treatment of beliefs, we have been silent about whether beliefs are well-calibrated. By well-calibrated, we mean that the beliefs implied by random choice are consistent with both choice data *and* actual state realizations. In the individual interpretation, this implies that the agent has rational expectations about her signals. In the group interpretation, this implies that agents have beliefs that are predictive of actual state realizations suggesting that there is genuine private information in the group. Given joint data on choices and state realizations, can the analyst tell if beliefs are well-calibrated? Define a *conditional test function* where the payoffs of a *conditional test act* are varied only in a given state. Theorem 6 shows that beliefs are well-calibrated if and only if conditional and unconditional test functions share the same mean. This provides a test for rational beliefs and can be combined with previous results on subjective misconfidence to obtain measures of objective misconfidence.

2 An Informational Model of Random Choice

2.1 Random Choice Rules

Let S be a finite objective state space and X be a finite set of prizes. For example, S could be the binary states for sick and healthy. Let ΔS and ΔX be their respective probability simplexes. Interpret ΔS as the set of beliefs about S and ΔX as the set of lotteries over prizes. Each choice option corresponds to a state-contingent payoff called an *act*. Following the setup of Anscombe and Aumann [2], an act is formally a mapping $f : S \to \Delta X$. Let H be the set of all acts. Call a finite set of acts a *decision-problem*. Let $\mathcal{K}$ be the set of all decision-problems, which we endow with the Hausdorff metric.[5] For notational convenience,

[5] For two sets F and G, the Hausdorff metric is given by

$$d_h(F,G) := \max\left(\sup_{f\in F}\inf_{g\in G}|f-g|, \sup_{g\in F}\inf_{f\in G}|f-g|\right)$$

let f denote the singleton set $\{f\}$ whenever there is no risk of confusion.

The primitive (i.e. the choice data) is a *random choice rule (RCR)* that specifies choice probabilities for acts in every decision-problem. In the *individual interpretation* of random choice, the RCR specifies the frequency distribution of choices by an agent if she chooses from the same decision-problem repeatedly. In the *group interpretation* of random choice, the RCR specifies the frequency distribution of choices in the group if every agent in the group chooses from the same decision-problem.

In the classic model of rational choice, if two acts are indifferent (i.e. they have the same utility), then the model is silent about which act will be chosen. We introduce an analogous innovation to address indifferences under random choice and random utility. If two acts are indifferent (i.e. they have the same random utility), then we declare that the random choice rule is unable to specify choice probabilities for each act in the decision-problem. For instance, it could be that one act is chosen over another with probability a half, but any other probability would also be perfectly consistent with the model. Similar to how the classic model is silent about which act will be chosen in the case of indifference, the random choice model is silent about what the choice probabilities are. In both cases, indifference is interpreted as choice behavior that is beyond the scope of the model. This provides the analyst with additional freedom to interpret data.

Formally, indifference is modelled as non-measurability with respect to some σ-algebra $\mathcal{H}$ on H. For example, if $\mathcal{H}$ is the Borel algebra, then this corresponds to the benchmark case where every act is measurable and there are no indifferences. In general, $\mathcal{H}$ can be coarser than the Borel algebra. Given any decision-problem, the decision-problem itself must be measurable. This is because we know that some act will be chosen for sure from the decision-problem. For $F \in \mathcal{K}$, let $\mathcal{H}_F$ be the σ-algebra generated by $\mathcal{H} \cup \{F\}$.[6] Let Π be the set of all probability measures on any measurable space of H.

Definition. A *random choice rule (RCR)* is a $(\rho, \mathcal{H})$ where $\rho : \mathcal{K} \to \Pi$ and $\rho(F)$ is a measure on $(H, \mathcal{H}_F)$ with support $F \in \mathcal{K}$.

Let ρ_F denote the measure $\rho(F)$ for every $F \in \mathcal{K}$. A RCR thus assigns a probability measure on $(H, \mathcal{H}_F)$ for each decision-problem $F \in \mathcal{K}$ such that $\rho_F(F) = 1$.[7] Interpret $\rho_F(G)$ as the probability that some act in G will be chosen in the decision-problem $F \in \mathcal{K}$. For ease

[6] This definition imposes a form of common measurability across all decision-problems. It can be relaxed if we strengthen the monotonicity axiom.

[7] The definition of $\mathcal{H}_F$ ensures that $\rho_F(F)$ is well-defined.

of exposition, we denote RCRs by ρ with the implicit understanding that it is associated with some $\mathcal{H}$. To address the fact that $G \subset F$ may not be $\mathcal{H}_F$-measurable, we use the outer measure[8]

$$\rho_F^* (G) := \inf_{G \subset G' \in \mathcal{H}_F} \rho_F (G')$$

As both ρ and ρ^* coincide on measurable sets, let ρ denote ρ^* without loss of generality.

A RCR is *deterministic* iff all choice probabilities are either zero or one. What follows is an example of a deterministic RCR; its purpose is to highlight (1) the use of non-measurability to model indifferences and (2) the modeling of classic deterministic choice as a special case of random choice.

Example 1. Let $S = \{s_1, s_2\}$ and $X = \{x, y\}$. Without loss of generality, let $f = (a, b) \in [0,1]^2$ denote the act $f \in H$ where

$$\begin{aligned} f(s_1) &= a\delta_x + (1-a)\,\delta_y \\ f(s_2) &= b\delta_x + (1-b)\,\delta_y \end{aligned}$$

Let $\mathcal{H}$ be the σ-algebra generated by sets of the form $B \times [0,1]$ where B is a Borel set on $[0,1]$. Consider the RCR $(\rho, \mathcal{H})$ where $\rho_F(f) = 1$ if $f_1 \geq g_1$ for all $g \in F$. Acts are ranked based on how likely they will yield prize x if state s_1 occurs. This could describe an agent who prefer x to y and believes that s_1 will realize for sure. Let $F = \{f, g\}$ where $f_1 = g_1$ and note that neither f nor g is $\mathcal{H}_F$-measurable; the RCR is unable to specify choice probabilities for f or g. This is because both acts yield x with the same probability in state s_1. The two acts are thus "indifferent". Observe that ρ corresponds exactly to classic deterministic choice where f is preferred to g iff $f_1 \geq g_1$.

2.2 Information Representation

We now describe the role of private information. Recall the timing of the model. At time 1, the agent receives some private information about the underlying state. In the case of health insurance for example, this could be a signal about her likelihood of falling sick. At time 2, she chooses the best act in the decision-problem given her updated belief. Since her private

[8] Lemma A1 in the Appendix ensures that this is well-defined.

information is unobservable to the analyst, choice is probabilistic and can be modeled by the analyst as a RCR. Call this an *information representation* of the RCR.

Since each signal realization corresponds to a posterior belief $q \in \Delta S$, we model private information as an *signal distribution* μ over the canonical signal space ΔS. This approach allows us to circumvent issues due to updating and work directly with posterior beliefs. Note that in the group interpretation, μ is simply the distribution of beliefs in the group. Consider the degenerate distribution $\mu = \delta_q$ for some $q \in \Delta S$. In the individual interpretation, this corresponds to the case where the agent receives no information and retains her initial prior. In the group interpretation, this corresponds to the case where all agents in the group share the same belief. In either interpretation, choice is deterministic in this example.

Let $u : \Delta X \to \mathbb{R}$ be an affine utility function. The subjective expected utility of an act $f \in H$ given the belief $q \in \Delta S$ is $q \cdot (u \circ f)$.[9] In order to study the role of information in random choice, we hold u fixed (we relax this in Section 2.4). In the individual interpretation, this implies that signals only affect beliefs but not tastes. In the group interpretation, this implies that agents have heterogeneous beliefs but homogeneous tastes (i.e. risk aversion). Choice is stochastic only as a result of varying beliefs.

Given a utility function, a signal distribution is *regular* iff the expected utilities of two acts are either always or never equal. This relaxes the standard restriction in traditional RUM where utilities are never equal.

Definition. μ is *regular* iff $q \cdot (u \circ f) = q \cdot (u \circ g)$ with μ-measure zero or one.

Let (μ, u) consist of a regular μ and a non-constant u. Define an information representation as follows.

Definition (Information Representation)**.** ρ is *represented by* (μ, u) iff for $f \in F \in \mathcal{K}$,

$$\rho_F(f) = \mu \{q \in \Delta S \mid q \cdot (u \circ f) \geq q \cdot (u \circ g) \ \forall g \in F\}$$

This is a RUM model where the random utilities are subjective expected utilities that depend on unobservable private information. If a RCR is represented by (μ, u), then the probability of choosing $f \in F$ is precisely the probability that f attains the highest subjective expected utility in F. Since the signal distribution μ is subjective, any inference about the agent's private information can only be gleaned by studying the RCR.

[9] For any act $f \in H$, let $u \circ f \in \mathbb{R}^S$ denote its utility vector where $(u \circ f)(s) = u(f(s))$ for all $s \in S$.

One of the classic critiques of subjective expected utility (especially in the context of health insurance) is the state independence of the (taste) utility. In an information representation, utilities are independent of both the unobservable subjective states affecting information and the objective state space S. We address the former below in Section 2.4 where we characterize a general model that allows for unobservable utility shocks. The latter can by addressed by any random choice generalization of the classic solutions to state-dependent utility (see Karni, Schmeidler and Vind [30] and Karni [29]).[10]

We follow with two examples of information representations.

Example 2. Let $S = \{s_1, s_2\}$, $X = \{x, y\}$ and $u(a\delta_x + (1-a)\delta_y) = a \in [0,1]$. Let $\mu = \delta_q$ for $q \in \Delta S$ such that $q_{s_1} = 1$. Since the agent believes that s_1 will occur for sure, she only cares about payoffs in that state. Let (μ, u) represent ρ, and let $F = \{f, g\}$. If $u(f(s_1)) \geq u(g(s_1))$, then

$$\rho_F(f) = \mu\{q \in \Delta S \mid q \cdot (u \circ f) \geq q \cdot (u \circ g)\} = 1$$

If $u(f(s_1)) = u(g(s_1))$, then $q \cdot (u \circ f) = q \cdot (u \circ g)$ μ-a.s. so

$$\rho_F(f) = \rho_F(g) = 1$$

and neither f nor g is $\mathcal{H}_F$-measurable. This is exactly the RCR described in Example 1 above.

Example 3. Let $S = \{s_1, s_2, s_3\}$, $X = \{x, y\}$ and $u(a\delta_x + (1-a)\delta_y) = a \in [0,1]$. Let μ be the uniform measure on ΔS and (μ, u) represent ρ. Given two acts f and g such that $u \circ f = v \in [0,1]^3$ and $u \circ g = w \in [0,1]^3$, we have

$$\rho_{f \cup g}(f) = \mu\{q \in \Delta S \mid q \cdot v \geq q \cdot w\}$$

Thus, the probability that f is chosen over g is the area of ΔS intersected with the halfspace $q \cdot (v - w) \geq 0$.

Example 2 above is exactly the standard subjective utility model where agents believe that s_1 will realize for sure. It serves to demonstrate how our random choice model includes standard subjective expected utility as a special case.

[10] In practice however, the empirical literature on health insurance has largely assumed state independence due to a dearth of empirical evidence (see Finkelstein, Luttmer and Notowidigdo [19]).

We end this section with a technical remark about regularity. As mentioned above, indifferences in traditional RUM must occur with probability zero. Since all choice probabilities are specified, these models run into difficulty when there are indifferences in the random utility.[11] Our definition of regularity circumvents this by allowing for just enough flexibility so that we can model indifferences using non-measurability.[12] In Example 2, if $q \cdot (u \circ f) = q \cdot (u \circ g)$ μ-a.s., then neither f nor g is $\mathcal{H}_{f \cup g}$-measurable. Acts that have the same utility μ-a.s. correspond exactly to non-measurable singletons. Note that our definition still imposes certain restrictions on μ. For example, multiple mass points are not allowed if μ is regular.[13]

2.3 Axiomatic Characterization

We now provide an axiomatic characterization of information representations. Given two decision-problems F and G, let $aF + (1-a)\,G$ denote the Minkowski mixture of the two sets for some $a \in [0,1]$.[14] Let $\text{ext}F$ denote the set of extreme acts of $F \in \mathcal{K}$.[15] We assume $f \in F \in \mathcal{K}$ throughout. The first three axioms below are standard restrictions on RCRs.

Axiom 1 (Monotonicity)**.** *$G \subset F$ implies $\rho_G(f) \geq \rho_F(f)$.*

Axiom 2 (Linearity)**.** *$\rho_F(f) = \rho_{aF+(1-a)g}(af + (1-a)\,g)$ for $a \in (0,1)$.*

Axiom 3 (Extremeness)**.** *$\rho_F(extF) = 1$.*

Monotonicity is standard for any RUM. To see this, note that when we enlarge the decision-problem, we introduce new acts that could dominate acts in the original decision-problem. Thus, the probability that the original acts are chosen can only decrease.

[11] Note that if we assumed that acts are mappings $f : S \to [0,1]$, then we could obtain a consistent model by assuming that indifferences never occur. Nevertheless, this would not allow us to include deterministic choice as a special case. Moreover, while extending a model with these mappings to the Anscombe-Aumann space is standard under deterministic choice, the extension under random choice is more intricate and warrants our approach.

[12] More precisely, our definition of regularity permits strictly positive measures on sets in ΔS that have less than full dimension. Regularity in Gul and Pesendorfer [26] on the other hand, requires μ to be full-dimensional (see their Lemma 2). See Block and Marschak [7] for the case of finite alternatives.

[13] See Example 7 below.

[14] The Minkowski mixture for $\{F, G\} \subset \mathcal{K}$ and $a \in [0,1]$ is defined as

$$aF + (1-a)\,G := \{\, af + (1-a)\,g \mid \ (f,g) \in F \times G \}$$

[15] Formally, $f \in \text{ext}F \in \mathcal{K}$ iff $f \in F$ and $f \neq ag + (1-a)\,h$ for some $\{g,h\} \subset F$ and $a \in (0,1)$.

Linearity and extremeness follow from the fact that the random utilities in our model are linear (i.e. agents are subjective expected utility maximizers). Linearity is exactly the random choice analog of the standard independence axiom. In fact, it is *the* version of the independence axiom tested in many experimental settings (see Kahneman and Tversky [28] for example). Extremeness implies that only extreme acts of the decision-problem will be chosen. This follows from the fact that when linear utilities are used for evaluation, mixtures of acts in the decision-problem are never chosen (except for cases of indifference). Note that both rule out behaviors associated with random non-linear utilities (such as ambiguity aversion for example).

We now introduce the continuity axiom. Given a RCR, let $\mathcal{K}_0 \subset \mathcal{K}$ be the set of decision-problems where every act in the decision-problem is measurable with respect to the RCR. In other words, $F \in \mathcal{K}_0$ iff $f \in \mathcal{H}_F$ for all $f \in F$. Let Π_0 be the set of all Borel measures on H endowed with the topology of weak convergence. Since all acts in $F \in \mathcal{K}_0$ are $\mathcal{H}_F$-measurable, $\rho_F \in \Pi_0$ for all $F \in \mathcal{K}_0$ without loss of generality.[16] Call ρ *continuous* iff it is continuous on the restricted domain $\mathcal{K}_0$.

Axiom 4 (Continuity). *$\rho : \mathcal{K}_0 \to \Pi_0$ is continuous.*

If $\mathcal{H}$ is the Borel algebra, then $\mathcal{K}_0 = \mathcal{K}$. In this case, our continuity axiom condenses to standard continuity. In general though, the RCR is not continuous over all decision-problems. In fact, the RCR is discontinuous precisely at decision-problems that contain indifferences. In other words, choice data that is beyond the scope of the model exhibits discontinuities with respect to the RCR. However, every decision-problem is arbitrarily (Hausdorff) close to some decision-problem in $\mathcal{K}_0$, so continuity is preserved over almost all decision-problems.[17]

The first four axioms are necessary and sufficient for random expected utility (see Gul and Pesendorfer [26]). We introduce three new axioms. An act $f \in H$ is *constant* iff $f(s)$ is the same for all $s \in S$. A decision-problem is *constant* iff it contains only constant acts. Given $f \in H$ and $s \in S$, define $f_s \in H$ as the constant act that yields the payoff $f(s)$ in every state. For $F \in \mathcal{K}$, let $F_s := \bigcup_{f \in F} f_s$ be the constant decision-problem consisting of f_s for all $f \in F$.

Axiom 5 (Non-degeneracy). *$\rho_F(f) < 1$ for some F and $f \in F$.*

Axiom 6 (C-determinism). *$\rho_F(f) \in \{0, 1\}$ for constant F.*

[16] We can easily complete ρ_F so that it is Borel measurable.
[17] See Lemma A15 in the Appendix.

Axiom 7 (S-monotonicity). *$\rho_{F_s}(f_s) = 1$ for all $s \in S$ implies $\rho_F(f) = 1$.*

Non-degeneracy rules out the trivial case of universal indifference. C-determinism states that the RCR is deterministic over constant decision-problems. This is because choice is stochastic only as a result of varying beliefs and in a constant decision-problem, every act yields the same payoff regardless of the state. Note that if ρ is represented by (μ, u), then ρ induces a preference relation over constant acts that is exactly represented by u. S-monotonicity states that if an act is the best regardless of which state occurs, then it must be chosen for sure. It is the random choice analog of the standard state-by-state monotonicity condition from deterministic choice. Taken together, Axioms 1-7 are necessary and sufficient for an information representation.

Theorem 1. *ρ has an information representation iff it satisfies Axioms 1-7.*

Proof. See Appendix. □

Theorem 2 states that studying binary choices is enough to completely identify private information. In other words, given two agents (or two groups of agents), comparing binary choices is sufficient to completely differentiate between the two information structures.[18]

Theorem 2 (Uniqueness). *Suppose ρ and τ are represented by (μ, u) and (ν, v) respectively. Then the following are equivalent:*

(1) *$\rho_{f \cup g}(f) = \tau_{f \cup g}(f)$ for all f and g*

(2) *$\rho = \tau$*

(3) *$(\mu, u) = (\nu, \alpha v + \beta)$ for $\alpha > 0$*

Proof. See Appendix. □

Note that if we allow the utility u to be constant, then Axiom 7 can be dropped in Theorem 1 without loss of generality. However, the uniqueness of μ in the representation would obviously fail in Theorem 2.

[18] Chambers and Lambert [10] also study the elicitation of unobservable information. While we consider an infinite collection of binary decision-problems to obtain uniqueness, they consider a single decision-problem but with an infinite set of choice options.

2.4 General Case: Random Subjective Expected Utility

We now consider a general case where choice is driven by both belief and taste (i.e. risk preference) shocks. Let $\mathbb{R}^X$ be the space of affine utility functions $u : \Delta X \to \mathbb{R}$ and π be a measure on $\Delta S \times \mathbb{R}^X$. Interpret π as the joint distribution over beliefs and tastes. Assume that u is non-constant π-a.s.. Note that the marginal distribution of π on ΔS corresponds exactly to the signal distribution μ. The corresponding regularity condition on π follows.

Definition. π is *regular* iff $q \cdot (u \circ f) = q \cdot (u \circ g)$ with π-measure zero or one.

Define a random subjective expected utility (RSEU) representation as follows.

Definition (RSEU Representation)**.** ρ is *represented by* a regular π iff for $f \in F \in \mathcal{K}$,

$$\rho_F(f) = \pi \left\{ (q, u) \in \Delta S \times \mathbb{R}^X \mid q \cdot (u \circ f) \geq q \cdot (u \circ g) \;\; \forall g \in F \right\}$$

This is a RUM model where the random subjective expected utilities depend not only on beliefs but tastes as well. In the individual interpretation, this describes an agent who receives unobservable shocks to both beliefs and tastes. In the group interpretation, this describes a group with heterogeneity in both beliefs and risk aversion. Note that in the special case where $\pi(\Delta S \times \{u\}) = 1$ for some non-constant $u \in \mathbb{R}^X$, this reduces to an information representation.

To characterize a RSEU representation, C-determinism must be relaxed. In particular, we replace S-monotonicity with a state-by-state independence axiom. For $f \in H$ and $\{s_1, s_2\} \subset S$, define $f_{s_1,s_2} \in H$ as the act obtained from f by replacing the payoff in s_2 with the payoff in s_1. In other words, $f_{s_1,s_2}(s_2) = f(s_1)$ and $f_{s_1,s_2}(s) = f(s)$ for all $s \neq s_2$.

Axiom 8 (S-independence)**.** $\rho_F(f_{s_1,s_2} \cup f_{s_2,s_1}) = 1$ *for* $F = \{f, f_{s_1,s_2}, f_{s_2,s_1}\}$

S-independence ensures that two acts that are constant over two states will be chosen for sure over an act that is non-constant over those states. This follows from the fact that non-constant acts are only chosen over constant acts when an agent has state-dependent utility. This is the random choice version of the state-by-state independence axiom.[19]

The proposition below shows that by replacing C-determinism and S-monotonicity with S-independence, we obtain a RSEU representation.

[19] Under deterministic choice, S-independence reduces to the condition that $f_{s_1,s_2} \succeq f$ or $f_{s_2,s_1} \succeq f$. Theorem 7 implies that this is equivalent to state-by-state independence axiom in the presence of the other standard axioms. Note that the definition of null states becomes unnecessary in this characterization.

Proposition 1. *ρ has a RSEU representation iff it satisfies Axioms 1-5 and 8.*

Proof. See Appendix. □

3 Test Functions

We now introduce a key technical tool that will play an important role in our subsequent analysis. To motivate the discussion, imagine enticing the agent with a *test act* that yields a fixed payoff in every state. Given a decision-problem, what is the probability that the agent will choose some act in the original decision-problem over the test act? If the test act is very valuable (i.e. the fixed payoff is high), then this probability will be low. As we lower the value of the test act, this probability will rise. Call this the *test function* for the original decision-problem.

To be concrete, suppose the original decision-problem is a set of health plans and the test act corresponds to a no-deductible (full-insurance) health plan introduced by the insurance company. The test function of the original set of plans is the probability of choosing some original plan over the test plan as a function of the test plan's premium. As we increase its premium, the test plan becomes less valuable and the demand for an original plan increases.

Call an act the *best* (*worst*) act under ρ iff in any binary choice comparison, the act (other act) is chosen with certainty. In other words, $\rho_{f \cup \overline{f}}\left(\overline{f}\right) = \rho_{f \cup \underline{f}}\left(f\right) = 1$ for all $f \in H$. If ρ is represented by (μ, u), then there exists a best and a worst act.[20] *Test acts* are mixtures between the best and worst acts.

Definition. A *test act* is $f^a := a\underline{f} + (1-a)\overline{f}$ for some $a \in [0,1]$.

Test acts are constant under information representations. Define *test functions* as follows.

Definition. Given ρ, the *test function* of $F \in \mathcal{K}$ is $F_\rho : [0,1] \to [0,1]$ where

$$F_\rho\left(a\right) := \rho_{F \cup f^a}\left(F\right)$$

[20] To see this, recall that C-determinism implies that ρ induces a preference relation over constant acts that is represented by u. Since u is affine, we can always find a best and worst act in the set of all constant acts. S-monotonicity ensures that these are also the best and worst acts over all acts.

Let F_ρ denote the test function of decision-problem $F \in \mathcal{K}$ given ρ. If $F = f$ is a singleton act, then denote $f_\rho = F_\rho$. As a increases, the test act f^a progresses from the best to worst act and becomes less attractive. Thus, the probability of choosing something in F increases so test functions are increasing. They are in fact cumulative distribution functions under information representations.

Lemma 1. *If ρ has an information representation, then F_ρ is a cumulative for all $F \in \mathcal{K}$.*

Proof. See Appendix. □

Test functions are the random choice generalizations of best-worst mixtures that yield indifference under deterministic choice. They completely characterize utility distributions. An immediate corollary is that test functions for singletons are sufficient for identifying information.

Corollary 1. *Let ρ and τ have information representations. Then $\rho = \tau$ iff $f_\rho = f_\tau$ for all $f \in H$.*

Proof. Follows from Theorem 2. □

Corollary 1 implies that we can treat test functions as sufficient statistics for identifying private information. We end this section with an example.

Example 4. Recall Example 3 where $S = \{s_1, s_2, s_3\}$, $X = \{x, y\}$, $u\left(a\delta_x + (1-a)\,\delta_y\right) = a \in [0,1]$ and μ is the uniform measure on ΔS. Let (μ, u) represent ρ. Thus, $u \circ \overline{f} = (1,1,1)$ and $u \circ \underline{f} = (0,0,0)$. For $a \in [0,1]$, the test act f^a satisfies

$$u \circ f^a = a\underline{f} + (1-a)\,\overline{f} = (1-a,\ 1-a,\ 1-a)$$

Consider two act f and g where $u \circ f = (1,0,0)$ and $u \circ g = (b,b,b)$ for some $b \in [0,1]$. Their test functions are

$$\begin{aligned} f_\rho(a) &= \rho_{f \cup f^a}(f) = \mu\left\{q \in \Delta S \mid q_{s_1} \geq 1-a\right\} = a \\ g_\rho(a) &= \rho_{g \cup f^a}(g) = \mu\left\{q \in \Delta S \mid b \geq 1-a\right\} = \mathbf{1}_{[1-b,1]}(a) \end{aligned}$$

Since g is constant, its utility is fixed regardless of beliefs. Its test function increases abruptly at the critical value $1-b$. On the other hand, the utility of f depends on beliefs, so its test function increases more gradually as a increases.

4 Evaluating Option Sets

We now address the first exercise of inference. We show that there is an intimate relationship between random choice and ex-ante *valuations* of option sets (i.e. decision-problems). Consider a *valuation preference relation* $\succeq$ over decision-problems.

Definition (Subjective Learning). $\succeq$ is *represented by* (μ, u) iff it is represented by

$$V(F) = \int_{\Delta S} \sup_{f \in F} q \cdot (u \circ f) \; \mu(dq)$$

In the individual interpretation, V gives the agent's ex-ante valuation of decision-problems prior to receiving her signal. For example, if the agent expects to receive a very informative signal about her health, then she may exhibit a strict preference for flexibility. This is the subjective learning representation axiomatized by Dillenberger, Lleras, Sadowski and Takeoka [16] (henceforth DLST).

In the group interpretation, V gives the total utility or "social surplus" (see McFadden[34]) of decision-problems for all agents in the group. Consider a firm that chooses health insurance based on total employee welfare. In this case, the firm prefers more flexible (i.e. larger) sets of health plans if it thinks that employee beliefs about their health are disperse.

In this section, assume that ρ has a best and worst act and F_ρ is a well-defined cumulative for all $F \in \mathcal{K}$.[21] Call ρ *standard* iff it is monotone, linear and continuous.

Definition. ρ is *standard* iff it is monotone, linear and continuous.

Any ρ that has an information representation is standard. On the other hand, the condition is relatively mild; it is insufficient to ensure that a random utility representation even exists.

Given random choice data, can the analyst directly compute valuations? Note that if a decision-problem is very valuable, then its acts will be chosen frequently and its the test function will take on high values. Consider evaluating decision-problems as follows.

Definition. Given ρ, let $\succeq_\rho$ be represented by $V_\rho : \mathcal{K} \to [0,1]$ where

$$V_\rho(F) := \int_{[0,1]} F_\rho(a) \, da$$

Theorem 3 confirms that $\succeq_\rho$ is the valuation preference relation corresponding to the RCR ρ. The analyst can simply use V_ρ to compute valuations.

[21] The best and worst acts are respectively defined as constant acts $\bar{f}$ and $\underline{f}$ where $\rho_{f \cup \bar{f}}(\bar{f}) = \rho_{f \cup \underline{f}}(f) = 1$ for all $f \in H$.

Theorem 3. *The following are equivalent:*

(1) ρ *is represented by* (μ, u)

(2) ρ *is standard and* $\succeq_\rho$ *is represented by* (μ, u)

Proof. See Appendix. □

Thus, if ρ has an information representation, then the integral of the test function F_ρ is exactly the valuation of F. An immediate consequence is that if $F_\rho(a) \geq G_\rho(a)$ for all $a \in [0,1]$, then $V_\rho(F) \geq V_\rho(G)$. Thus, first-order stochastic dominance of test functions implies higher valuations.

Theorem 3 also demonstrates that if a standard RCR induces a preference relation that has a subjective learning representation, then that RCR must have an information representation. In fact, both the RCR and the preference relation are represented by the same (μ, u). This serves as an alternate characterization of information representations using properties of its induced preference relation.

The discussion above suggests a converse: given valuations, can the analyst directly compute random choice? First, $\succeq$ is *dominant* iff it satisfies the following.

Definition. $\succeq$ is *dominant* iff $f_s \succeq g_s$ for all $s \in S$ implies $F \sim F \cup g$ for $f \in F$.

Dominance is one of the axioms of a subjective learning representation in DLST. It captures the intuition that adding acts that are dominated in every state does not affect ex-ante valuations. Define a RCR induced by a preference relation as follows.

Definition. Given $\succeq$, let $\rho_\succeq$ denote any standard ρ such that a.e.

$$\rho_{F \cup f_a}(f_a) = \frac{dV(F \cup f_a)}{da}$$

where $V : \mathcal{K} \to [0,1]$ represents $\succeq$ and $f_a := a\overline{f} + (1-a)\underline{f}$.

Given any preference relation $\succeq$, the RCR $\rho_\succeq$ may not even exist. On the other hand, there could be a multiplicity of RCRs that satisfy this definition. Theorem 4 shows that for our purposes, these issues do not matter. If $\succeq$ has a subjective learning representation, then $\rho_\succeq$ exists and is the unique RCR corresponding to $\succeq$.

Theorem 4. *The following are equivalent:*

(1) $\succeq$ *is represented by* (μ, u)

(2) $\succeq$ *is dominant and* $\rho_{\succeq}$ *is represented by* (μ, u)

Proof. See Appendix. □

Thus, the analyst can use $\rho_{\succeq}$ to directly compute random choice from valuations. The probability that the act f_a is chosen is exactly its marginal contribution to the ex-ante valuation of the decision-problem.[22] For example, consider a set of health plans that includes a no-deductible (full-insurance) test plan. If increasing the premium of the test plan does not affect the valuation of the set, then the test plan is never chosen from the set. Any violation of this would indicate some form of inconsistency (which we explore in Section 6).

Given $\succeq$ represented by (μ, u), the corresponding RCR $\rho = \rho_{\succeq}$ can be constructed as follows. First, define ρ so that it coincides with u over all constant acts. Then use the definition of $\rho_{\succeq}$ to specify $\rho_{F \cup f_a}(f_a)$ for all $a \in [0, 1]$ and $F \in \mathcal{K}$. Finally, linearity extends ρ to all decision-problems. By Theorem 4, the ρ so constructed is represented by (μ, u).

The other implication is that if a dominant preference relation induces a RCR that has an information representation, then that preference relation has a subjective learning representation. As in Theorem 3, this is an alternate characterization of subjective learning representations using properties of its induced RCR.

Theorem 4 is the random choice version of Hotelling's Lemma from classical producer theory. The analogy follows if we interpret choice probabilities as "outputs", conditional utilities as "prices" and valuations as "profits".[23] Similar to how Hotelling's Lemma is used to compute firm outputs from the profit function, Theorem 4 can be used to compute random choice from valuations.

Similar to classical results from consumer and producer theory (such as Hotelling's Lemma) that provide a methodology for relating data, Theorems 3 and 4 allow an analyst to compute valuations directly from random choice and vice-versa. Integrating test functions give valuations, while differentiating valuations yields random choice. This com-

[22] In the econometrics literature, this is related to the Williams-Daly-Zachary Theorem that also follows from an envelope argument (see McFadden [34]). The presence of constant acts in the Anscombe-Aumann setup however means Theorem 3 has no counterpart.

[23] Formally, let y be a probability on F, and for each y, let $\mathcal{Q}_y = \{Q_f\}_{f \in F}$ denote some partition of ΔS such that $\mu(Q_f) = y(f)$. For $f \in F$, let $p_f := \int_{Q_f} \frac{q \cdot (u \circ f)}{\mu(Q_f)} \mu(dq)$ denote the conditional utility of f. Interpret y as "output" and p as "price" so $V(F) = \sup_{y, \mathcal{Q}_y} p \cdot y$ is the maximizing "profit". Note that $a = p_{f_a}$ is exactly the price of f_a. The caveat is that prices are fixed in Hotelling's Lemma while in our case, p_f depends on $\mathcal{Q}_y$.

putation is direct since identification of the signal distribution and utility is unnecessary. In the individual interpretation, observing choice data in one time period allows the analyst to directly compute choice data in the other. We summarize these insights below.

Corollary 2. *Let* $\succeq$ *and* ρ *be represented by* (μ, u)*. Then* $\succeq_\rho = \succeq$ *and* $\rho_\succeq = \rho$*.*

Proof. Follows immediately from Theorems 3 and 4. $\square$

The following demonstrates how these operations work.

Example 5. Let $S = \{s_1, s_2\}$, $X = \{x, y\}$ and $u(a\delta_x + (1-a)\delta_y) = a \in [0,1]$. Associate each $q \in \Delta S$ with $t \in [0,1]$ such that $t = q_{s_1}$. Let μ have density $6t(1-t)$. Let $\succeq$ and ρ be represented by (μ, u) and $V : \mathcal{K} \to [0,1]$ represents $\succeq$. Two health plans are offered: a no-deductible (full-insurance) plan f and a high-deductible plan g. Let $u \circ f = \left(\frac{2}{5}, \frac{2}{5}\right)$, $u \circ g = \left(\frac{1}{4}, \frac{3}{4}\right)$ and $F = \{f, g\}$.

Valuations from random choice: The test function of F is given by

$$F_\rho(a) = \mu\left\{t \in [0,1] \;\middle|\; \max\left\{\frac{2}{5},\ t\frac{1}{4} + (1-t)\frac{3}{4}\right\} \geq 1 - a\right\}$$

It is straightforward to check that $F_\rho(a) = 0$ for $a \leq \frac{1}{4}$, $F_\rho(a) = 1$ for $a \geq \frac{3}{5}$ and

$$F_\rho(a) = (4a-1)^2(1-a)$$

for $a \in \left(\frac{1}{4}, \frac{3}{5}\right)$. Integrating the test function yields the valuation of the set

$$V(F) = \int_{[0,1]} F_\rho(a)\, da = \int_{\left[\frac{1}{4}, \frac{3}{5}\right]} (4a-1)^2(1-a)\, da + \frac{2}{5} \approx 0.511$$

Random choice from valuations: Let $f_a := a\overline{f} + (1-a)\underline{f}$ where $a \in [0,1]$ and note that $f_{\frac{2}{5}} = f$. It is straightforward to check that for $a \in \left(\frac{1}{4}, \frac{3}{4}\right)$

$$\begin{aligned} V(g \cup f_a) &= \int_{[0,1]} \max\left\{t\frac{1}{4} + (1-t)\frac{3}{4},\ a\right\}\mu(dt) \\ &= -4a^4 + 8a^3 - \frac{9}{2}a^2 + a + \frac{27}{64} \end{aligned}$$

Differentiating $V(g \cup f_a)$ at $a = \frac{2}{5}$ yields the probability of choosing plan f

$$\rho_F(f) = \rho_{g \cup f_a}(f_a) = \left.\frac{dV(g \cup f_a)}{da}\right|_{a=\frac{2}{5}} = \left.(4a-1)^2(1-a)\right|_{a=\frac{2}{5}} = \frac{27}{125}$$

5 Assessing Informativeness

5.1 Random Choice Characterization of Better Information

Given two agents (or two groups of agents), can the analyst use random choice to infer who gets better information even when information is not directly observable? First, consider the classic methodology when information *is* observable. A transition kernel[24] on ΔS is *mean-preserving* iff it preserves average beliefs about S.

Definition. The transition kernel $K : \Delta S \times \mathcal{B}(\Delta S) \to [0,1]$ is *mean-preserving* iff for all $q \in \Delta S$,

$$\int_{\Delta S} p \; K(q, dp) = q$$

Let μ and ν be two signal distributions. We say μ is *more informative* than ν, iff the distribution of beliefs under μ is a mean-preserving spread of the distribution of beliefs under ν.

Definition. μ is *more informative* than ν iff there is a mean-preserving transition kernel K such that for all $Q \in \mathcal{B}(\Delta S)$

$$\mu(Q) = \int_{\Delta S} K(p, Q) \;\; \nu(dp)$$

If μ is more informative than ν, then the information structure of ν can be generated by adding noise or "garbling" μ. This is Blackwell's [5, 6] ranking of informativeness based on signal sufficiency. In other words, μ is a sufficient signal for ν. If K is the identity kernel, then no information is lost and $\nu = \mu$.

In the classical approach, Blackwell [5, 6] showed that better information is characterized by higher ex-ante valuations. What is the random choice characterization of better information? First, consider a degenerate signal distribution corresponding to an uninformative signal (or a group of agents all with the same belief). Choice is deterministic in this case, so the test function of a singleton act corresponds a unit mass distribution. Another agent (or group of agents) has a dispersed signal distribution. Depending on the posterior realization,

[24] $K : \Delta S \times \mathcal{B}(\Delta S) \to [0,1]$ is a transition kernel iff $q \to K(q, Q)$ is measurable for all $Q \in \mathcal{B}(\Delta S)$ and $Q \to K(q, Q)$ is a measure on ΔS for all $q \in \Delta S$.

an act can either be valuable or not. Its test function corresponds to a more dispersed distribution. If both agents (or groups) have the same average beliefs, then the test function under the better informed will be a mean-preserving spread of that of the other. For general test functions, this is captured by second-order stochastic dominance.

Definition. $F \geq_{SOSD} G$ iff $\int_{\mathbb{R}} \phi dF \geq \int_{\mathbb{R}} \phi dG$ for all increasing concave $\phi : \mathbb{R} \to \mathbb{R}$.

Theorem 5. *Let ρ and τ be represented by (μ, u) and (ν, u) respectively. Then μ is more informative than ν iff $F_\tau \geq_{SOSD} F_\rho$ for all $F \in \mathcal{K}$.*

Proof. See Appendix. □

Theorem 5 equates an unobservable multi-dimensional information ranking with an observable single-dimensional stochastic dominance relation. An analyst can assess informativeness simply by comparing test functions via second-order stochastic dominance. It is the random choice characterization of better information. The intuition is that better information corresponds to more dispersed (i.e. random) choice while worse information corresponds to more concentrated (i.e. deterministic) choice. In the individual interpretation, the agent who gets a more informative signal about her health will exhibit greater dispersion in her annual choice of health insurance. In the group interpretation, the group with more private information will exhibit greater variation in the distribution of health insurance choices.

In DLST, better information is characterized by a greater preference for flexibility in the valuation preference relation. This is the preference relation version of Blackwell's [5, 6] result. A preference relation exhibits *more preference for flexibility* than another iff whenever the other prefers a set to a singleton, the first must do so as well.

Definition. $\succeq_1$ has *more preference for flexibility* than $\succeq_2$ iff $F \succeq_2 f$ implies $F \succeq_1 f$.

Corollary 3 relates our random choice characterization of better information with more preference for flexibility.

Corollary 3. *Let ρ and τ be represented by (μ, u) and (ν, u) respectively. Then the following are equivalent:*

(1) *$F_\tau \geq_{SOSD} F_\rho$ for all $F \in \mathcal{K}$*

(2) *$\succeq_\rho$ has more preference for flexibility than $\succeq_\tau$*

(3) *μ is more informative than ν*

Proof. By Theorem 5, (1) and (3) are equivalent. By Corollary 2, $\succeq_\rho$ and $\succeq_\tau$ are represented by (μ, u) and (ν, u) respectively. Hence, by Theorem 2 of DLST, (2) is equivalent to (3). □

Greater preference for flexibility and greater choice dispersion are the behavioral manifestations of better information. In the individual interpretation, a more informative signal corresponds to greater preference for flexibility (ex-ante) and more randomness in choice (ex-post). In the group interpretation, more private information corresponds to a greater group preference for flexibility and more heterogeneity in insurance choice. Note that by Corollary 2, Corollary 3 could have been formulated entirely in terms of preference relations. Also note the prominent role of test functions: computing their integrals evaluates options sets while comparing them via second-order stochastic dominance assesses informativeness.

If μ is more informative than ν, then it follows that the two distributions must have the same average.

Definition. μ and ν *share average beliefs* iff

$$\int_{\Delta S} q \; \mu(dq) = \int_{\Delta S} q \; \nu(dq)$$

In the individual interpretation, two agents share average beliefs iff they share the same prior about S (note the distinction between this prior and the more general "prior" over the universal space $\Delta S \times S$)[25]. In the group interpretation, two groups share average beliefs iff the average belief about S in both groups are the same. The analyst can determine if two agents (or two groups of agents) share average beliefs by comparing means of singleton test functions.

Lemma 2. *Let ρ and τ be represented by (μ, u) and (ν, u) respectively. Then μ and ν share average beliefs iff f_ρ and f_τ share the same mean for all $f \in H$.*

Proof. See Appendix. □

Combined with Theorem 5, Lemma 2 implies that a necessary condition for μ being more informative than ν is that every f_ρ is a mean-preserving spread of f_τ. This condition however is insufficient for assessing informativeness. It corresponds to a strictly weaker stochastic dominance relation known as the linear concave order.[26] Note that if f_ρ and f_τ have the same mean, then f has the same ex-ante valuation. Thus, from the analyst's perspective,

[25] Agreeing on the latter prior necessitates that both agents must have identical information structures.

[26] See Section 3.5 of Muller and Stoyan [36] for more about the linear concave order.

the random choice characterization of better information may sometimes be richer than the valuation characterization.

We end this section with an illustrative example.

Example 6. Let $S = \{s_1, s_2\}$, $X = \{x, y\}$ and $u(a\delta_x + (1-a)\delta_y) = a \in [0,1]$. Associate each $q \in \Delta S$ with $t \in [0,1]$ such that $t = q_{s_1}$. Let μ have density $6t(1-t)$ and ν be the uniform distribution. Note that ν is more informative than μ. Let ρ and τ be represented by (μ, u) and (ν, u) respectively. As in Example 5, consider the set of plans $F = \{f, g\}$ where $u \circ f = \left(\frac{2}{5}, \frac{2}{5}\right)$ and $u \circ g = \left(\frac{1}{4}, \frac{3}{4}\right)$. Recall that $F_\rho(a) = 0$ for $a \leq \frac{1}{4}$, $F_\rho(a) = 1$ for $a \geq \frac{3}{5}$ and

$$F_\rho(a) = (1-4a)^2(1-a)$$

for $a \in \left(\frac{1}{4}, \frac{3}{5}\right)$. The test function of F under τ satisfies $F_\tau(a) = 0$ for $a \leq \frac{1}{4}$, $F_\tau(a) = 1$ for $a \geq \frac{3}{5}$ and

$$F_\tau(a) = 2a - \frac{1}{2}$$

for $a \in \left(\frac{1}{4}, \frac{3}{5}\right)$. Hence, $F_\rho \geq_{SOSD} F_\tau$. Note that the test functions of g under ρ and τ respectively satisfy $g_\rho(a) = g_\tau(a) = 0$ for $a \leq \frac{1}{4}$, $g_\rho(a) = g_\tau(a) = 1$ for $a \geq \frac{3}{5}$ and

$$g_\rho(a) = (4a-1)^2(1-a)$$
$$g_\tau(a) = \frac{1}{2}(4a-1)$$

for $a \in \left(\frac{1}{4}, \frac{3}{4}\right)$. Hence, g_τ is a mean-preserving spread of g_ρ and $g_\rho \geq_{SOSD} g_\tau$ as well.

5.2 Special Case: Partitional Information

In this section, we study the special case where information corresponds to events that partition the state space. Fix a probability over S and consider a collection of events that form a partition of S. For instance, the events "healthy" and "sick" form a binary partition in the health insurance example. At time 1, an agent receives private information that reveals which event the true state is in. Given this information, she then updates her belief according to Bayes' rule. At time 2, she chooses an act from the decision-problem using this updated belief. Call this a *partitional information representation* of a RCR.

Let $\left(S, 2^S, r\right)$ be a probability space for some $r \in \Delta S$. Assume that r has full support without loss of generality.[27] Given an algebra $\mathcal{F} \subset 2^S$, let $Q_{\mathcal{F}}$ be the conditional probability given $\mathcal{F}$, that is, for $s \in S$ and the event $E \subset S$,

$$Q_{\mathcal{F}}(s, E) = \mathbb{E}_{\mathcal{F}}[\mathbf{1}_E]$$

where $\mathbb{E}_{\mathcal{F}}$ is the conditional expectation operator given $\mathcal{F}$. Note that we can interpret the conditional probability as a mapping $Q_{\mathcal{F}} : S \to \Delta S$ from states to beliefs. Thus, $\mathcal{F}$ induces a signal distribution $\mu_{\mathcal{F}} := r \circ Q_{\mathcal{F}}^{-1}$. Information corresponds to the event consisting of all states $s \in S$ where the belief is $q = Q_{\mathcal{F}}(s)$. These events form a natural partition of the state space S. Let $u : \Delta X \to \mathbb{R}$ be an affine utility so for any $f \in H$ and $s \in S$ such that $Q_{\mathcal{F}}(s) = q$

$$\mathbb{E}_{\mathcal{F}}[u \circ f] = q \cdot (u \circ f)$$

Let $(\mathcal{F}, u)$ denote an algebra $\mathcal{F}$ and a non-constant u.

We would like to consider the RCR generated by the signal distribution $\mu_{\mathcal{F}}$. However, excepting the case where $\mathcal{F}$ is trivial, $\mu_{\mathcal{F}}$ is in general not regular. The following example demonstrates how violations of regularity can create issues with our method of modeling indifferences.

Example 7. Let $S = \{s_1, s_2, s_3\}$, $X = \{x, y\}$ and $u(a\delta_x + (1-a)\delta_y) = a \in [0,1]$. Let $r = \left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)$ and $\mathcal{F}$ be generated by the partition $\{s_1, s_2 \cup s_3\}$. Let $q_1 := \delta_{s_1}$ and $q_2 := \frac{1}{2}\delta_{s_2} + \frac{1}{2}\delta_{s_3}$ so

$$\mu_{\mathcal{F}} = \frac{1}{3}q_1 + \frac{2}{3}q_2$$

Consider acts f, g and h where $u \circ f = (1,0,0)$, $u \circ g = (1,1,0)$ and $u \circ h = (0,0,1)$. Now

$$\begin{aligned} q_1 \cdot (u \circ h) &= 0 < q_1 \cdot (u \circ g) = 1 = q_1 \cdot (u \circ f) \\ q_2 \cdot (u \circ f) &= 0 < q_2 \cdot (u \circ g) = \frac{1}{2} = q_2 \cdot (u \circ h) \end{aligned}$$

so $\mu_{\mathcal{F}}\{q \in \Delta S \mid q \cdot (u \circ f) = q \cdot (u \circ g)\} = \frac{1}{3}$. Hence, $\mu_{\mathcal{F}}$ is not regular.

Let $F := \{f, g, h\}$ and note that f and g are indifferent one third of the time. By similar reasoning, g and h are indifferent two-thirds of the time. Using non-measurability to model indifferences implies that no singleton act in F is measurable. This fails to capture all the choice data implied by the model (it omits the fact that h will definitely not be chosen one

[27] That is $r_s > 0$ for all $s \in S$.

third of the time for example).

Example 7 illustrates that since $\mu_{\mathcal{F}}$ violates regularity, complications arise whenever a decision-problem contains acts that are neither always nor never equal (i.e. they have the same random utility on some $\mu_{\mathcal{F}}$-measure that is strictly between zero and one). We circumvent this issue by only considering decision-problems that do not contain such acts. Call these *generic*.

Definition. $F \in \mathcal{K}$ is *generic* under $\mathcal{F}$ iff for all $\{f, g\} \subset F$, $q \cdot (u \circ f) = q \cdot (u \circ g)$ with $\mu_{\mathcal{F}}$-measure zero or one.

Regularity is equivalent to requiring that all decision-problems are generic. Note that generic decision-problems are dense in the set of all decision-problems. Moreover, any $\mu_{\mathcal{F}}$ can always be approximated as the limit of a sequence of regular μ's. We are now ready for the formal definition.

Definition (Partitional Information)**.** ρ is *represented by* $(\mathcal{F}, u)$ iff for $f \in F \in \mathcal{K}$ where F is generic,

$$\rho_F(f) = r\{s \in S \mid \mathbb{E}_{\mathcal{F}}[u \circ f] \geq \mathbb{E}_{\mathcal{F}}[u \circ g] \;\; \forall g \in F\}$$

A partitional information representation is thus an information representation over generic decision-problems with signal distribution $\mu_{\mathcal{F}}$. A decision-problem $F \in \mathcal{K}$ is *deterministic* under ρ iff $\rho_F(f) \in \{0, 1\}$ for all $f \in F$. Let $\mathcal{D}_\rho$ denote the set of all generic decision-problems that are deterministic under ρ. Proposition 2 shows that in a partitional information model, the analyst can assess informativeness simply by comparing deterministic decision-problems.

Proposition 2. *Let ρ and τ be represented by $(\mathcal{F}, u)$ and $(\mathcal{G}, u)$ respectively. Then $\mathcal{F} \subset \mathcal{G}$ iff $\mathcal{D}_\tau \subset \mathcal{D}_\rho$.*

Proof. See Appendix. □

Thus, in the special case where information correspond to events that partition the state space, better information is equivalent to less deterministic (i.e. more random) choice. This captures the intuition shared by Theorem 5. Note that Theorem 5 still holds in this setting; the only complication is dealing with test functions for non-generic decision-problems.[28]

[28] One way to resolve this issue is to define the test function of $F \in \mathcal{K}$ at $a \in [0, 1]$ as $\lim_{b \downarrow a} F_\rho(b)$. Since generic decision-problems are dense, this is a well-defined cumulative. Theorem 5 then follows naturally.

6 Detecting Biases

In this section, we study situations when valuations and random choice are inconsistent. By inconsistent, we mean that private information inferred from $\succeq$ is misaligned with that inferred from ρ. In the individual interpretation, this misalignment describes an agent whose *prospective* (ex-ante) beliefs about her signal are misaligned with her *retrospective* (ex-post) beliefs. This is an informational version of the naive Strotz [45] model involving dynamic inconsistency with regards to information. In the group interpretation, this misalignment describes a situation where valuations of option sets indicate a more (or less) dispersed distribution of beliefs in the group than that implied by random choice. For example, a firm that evaluates health plans based on total employee welfare may overestimate (underestimate) the dispersion of employee beliefs and choose a more (less) flexible set of health plans than necessary. Since both interpretations are similar, for ease of exposition, we focus on the individual interpretation in this section.

Consider an agent who expects to receive a very informative signal. Ex-ante she prefers large option sets and may be willing to pay a cost in order to postpone choice and "keep her options open". Ex-post however, she consistently chooses the same option. For example, in the diversification bias, although an agent initially prefers a large option set containing a variety of foods, in the end, she always chooses the same food from the set.[29] If her choice is driven by informational reasons, then we can infer from her behavior that she initially anticipated a more informative signal than what her later choice suggests. This could be due to a misplaced "false hope" of better information. Call this *prospective overconfidence.*

On the flip side, there may be situations where ex-post choice reflects greater confidence than that implied by ex-ante preferences. To elaborate, consider an agent who expects to receive a very uninformative signal. Hence, ex-ante, large option sets are not very valuable. However, after receiving her signal, the agent becomes increasingly convinced of its informativeness. Both good and bad signals are interpreted more extremely, and she updates her beliefs by more than what she anticipated initially. This could be the result of some confirmatory bias where consecutive good and consecutive bad signals generate posterior beliefs that are more dispersed.[30] Call this *prospective underconfidence.*

[29] See Read and Loewenstein [41]. Note that in our case, the uncertainty is over future beliefs and not tastes. Nevertheless, there could be informational reasons for why one would prefer one food over another (a food recall scandal for a certain candy for example).

[30] See Rabin and Schrag [39] for a model and literature review of the confirmatory bias.

Since beliefs in our model are subjective, we are silent as to which period's choice behavior is more "correct". Both prospective overconfidence and underconfidence are relative comparisons involving *subjective misconfidence*. This is a form of belief misalignment that is independent of the true information structure and in some sense more fundamental. We show in Section 7 that given a richer data set (such as the joint data over choices and state realizations), the analyst can discern which period's choice behavior is correct.

Let the pair $(\succeq, \rho)$ denote both the valuation preference relation $\succeq$ and the RCR ρ. Motivated by Theorem 5, define *prospective overconfidence and underconfidence* as follows.

Definition. $(\succeq, \rho)$ exhibits:

(1) *prospective overconfidence* iff $F_\rho \geq_{SOSD} F_{\rho_\succeq}$ for all $F \in \mathcal{K}$

(2) *prospective underconfidence* iff $F_{\rho_\succeq} \geq_{SOSD} F_\rho$ for all $F \in \mathcal{K}$

Corollary 4. *Let $\succeq$ and ρ be represented by (μ, u) and (ν, u) respectively. Then the following are equivalent:*

(1) *$(\succeq, \rho)$ exhibits prospective overconfidence (underconfidence)*

(2) *$\succeq$ has more (less) preference for flexibility than $\succeq_\rho$*

(3) *μ is more (less) informative than ν*

Proof. By Corollary 2, $\rho_\succeq$ and $\succeq_\rho$ are represented by (μ, u) and (ν, u) respectively. The rest follows from Corollary 3. □

Corollary 4 provides a choice-theoretic foundation for subjective misconfidence. Both (1) and (2) are restrictions on observable behavior while (3) is an unobservable condition on the underlying information structures. Note that by Corollary 3, we could have equivalently defined prospective overconfidence (underconfidence) via more (less) preference for flexibility.

Corollary 4 also allows the analyst to order levels of prospective overconfidence and underconfidence via Blackwell's partial ordering of information structures. In other words, by studying the choice behaviors of two agents, an analyst can distinguish when one is more prospectively overconfident (or underconfident) than the other. This provides a unifying methodology to measure the severity of various behavioral biases, including the diversification and confirmatory biases.

7 Calibrating Beliefs

Following the footsteps of Savage [43] and Anscombe and Aumann [2], we have adopted a purely subjective treatment of beliefs. Our theory identifies when observed choice behavior is consistent with some distribution of beliefs but is unable to recognize when these beliefs may be incorrect. For example, our notions of misconfidence in the previous section are descriptions of *subjective* belief misalignment and not measures of *objective* misconfidence.

In this section, we incorporate additional data to achieve this distinction. By studying the joint distribution over choices and state realizations, an analyst can test whether agents' beliefs are objectively well-calibrated. In the individual interpretation, this implies that the agent has rational expectations about her signals. In the group interpretation, this implies that agents have beliefs that are predictive of actual state realizations and suggests that there is genuine private information in the group.

If information is observable, then calibrating beliefs is a well-understood statistical exercise.[31] We show that the analyst can calibrate beliefs even when information is not observable. For example, in the case of health insurance, an analyst may observe a correlation between choosing health insurance and ultimately falling sick. Even though information is not observable, data on both choices (whether an agent chooses health insurance or not) and state realizations (whether an agent gets sick or not) can be analyzed to infer if beliefs are well-calibrated.

Let $r \in \Delta S$ be some observed distribution over states. Assume that r has full support without loss of generality. In this section, the primitive consists of r and a *conditional random choice rule (cRCR)* that specifies choice frequencies conditional on the realization of each state. Recall that Π is the set of all probability measures on any measurable space of H. Define a cRCR as follows.

Definition. A *Conditional Random Choice Rule (cRCR)* is a $(\rho, \mathcal{H})$ where $\rho : S \times \mathcal{K} \rightarrow \Pi$ and $(\rho_s, \mathcal{H})$ is a RCR for all $s \in S$.

Unless otherwise stated, ρ in this section refers to a cRCR. For $s \in S$ and $f \in F \in \mathcal{K}$, $\rho_{s,F}(f)$ is the probability of choosing f conditional on state s realizing. For example, let f be a health plan and s the state of falling sick. In the individual interpretation, $\rho_{s,F}(f)$ is the frequency that the agent chooses plan f in all years in which she falls sick. In the group

[31] For example, see Dawid [14].

interpretation, $\rho_{s,F}(f)$ is the frequency of agents who chooses plan f among the subgroup of agents who fall sick. The probability that $f \in F$ is chosen *and* $s \in S$ occurs is given by $r_s \rho_{s,F}(f)$. Since each $\rho_{s,F}$ is a measure on $(H, \mathcal{H}_F)$, the measurable sets of $\rho_{s,F}$ and $\rho_{s',F}$ coincide for all s and s'. Define the unconditional RCR as

$$\bar{\rho} := \sum_{s \in S} r_s \rho_s$$

noting that $\bar{\rho}_F(f)$ is the unconditional probability of choosing $f \in F$.

The probability r in conjunction with the cRCR ρ completely specify the joint distribution over choices and state realizations. The marginal distributions of this joint distribution on choices and state realizations are $\bar{\rho}$ and r respectively. In both the individual and group interpretations, this form of state-dependent choice data is easily obtainable.[32]

Information now corresponds to a unique joint distribution over beliefs about S and actual state realizations. Define μ_s as the signal distribution conditional on $s \in S$ realizing. For example, let $s \in S$ be the state of falling sick. In the individual interpretation, μ_s is the agent's signal distribution in the years in which she falls sick. In the group interpretation, μ_s is the distribution of beliefs in the subgroup of agents who fall sick.

Let $\mu := (\mu_s)_{s \in S}$ be the collection of *conditional signal distributions*. Unless otherwise stated, μ in this section refers to this collection. Recall that $u : \Delta X \to \mathbb{R}$ is an affine utility function. Let (μ, u) denote some μ and a non-constant u. A cRCR ρ has an information representation iff each RCR ρ_s has an information representation for all $s \in S$.

Definition. ρ is represented by (μ, u) iff ρ_s is represented by (μ_s, u) for all $s \in S$.

By Theorem 1, a cRCR ρ has an information representation iff for every $s \in S$, the RCR ρ_s satisfies Axioms 1 to 7. The existence of an information representation does not imply that beliefs are well-calibrated. Well-calibrated beliefs require μ to be consistent with the observed frequency of states r. First, define the unconditional distribution of beliefs as

$$\bar{\mu} := \sum_{s \in S} r_s \mu_s$$

[32] In the individual interpretation, this data can be easily obtained in experimental work (for example, see Caplin and Dean [8]). In the group interpretation, this data is also readily available (for example, see Chiappori and Salanié [11]).

Definition. μ is *well-calibrated* iff for all $s \in S$ and $Q \in \mathcal{B}(\Delta S)$,

$$\mu_s(Q) = \int_Q \frac{q_s}{r_s} \bar{\mu}(dq)$$

Well-calibration implies that μ satisfies Bayes' rule. For each $s \in S$, μ_s is exactly the conditional signal distribution as implied by μ. In other words, choice behavior implies beliefs that agree with the observed joint data on choices and state realizations. In order to see this, let Q be the set of beliefs that rank $f \in F$ higher than all other acts in F. Note that $\rho_{s,F}(f) = \mu_s(Q)$ for all $s \in S$. By the definitions of $\bar{\rho}$ and $\bar{\mu}$, we also have $\bar{\rho}_F(f) = \bar{\mu}(Q)$. Hence

$$\frac{r_s \rho_{s,F}(f)}{\bar{\rho}_F(f)} = \frac{r_s \mu_s(Q)}{\bar{\mu}(Q)} = \frac{\int_Q q_s \bar{\mu}(dq)}{\bar{\mu}(Q)}$$

so the conditional probability that s occurs given that f is chosen agrees exactly with μ. In the individual interpretation, this implies that the agent has rational (i.e. correct) expectations about her signals. In the group interpretation, this implies that all agents in the group have rational (i.e. correct) beliefs about their future health and so there is genuine private information in the group. The following example illustrates.

Example 8. Let $S = \{s_1, s_2\}$ and again, associate each $q \in \Delta S$ with $t \in [0,1]$ such that $t = q_{s_1}$. Let $r = \left(\frac{1}{2}, \frac{1}{2}\right)$ and $\bar{\mu}$ have density $6t(1-t)$. Let μ_{s_1} and μ_{s_2} have densities $12t^2(1-t)$ and $12t(1-t)^2$ respectively. For $b \in [0,1]$,

$$\int_{[0,b]} \frac{q_{s_1}}{r_{s_1}} \bar{\mu}(dq) = \int_{[0,b]} 12t^2(1-t)\,dt = \mu_{s_1}[0,b]$$
$$\int_{[0,b]} \frac{q_{s_2}}{r_{s_2}} \bar{\mu}(dq) = \int_{[0,b]} 12t(1-t)^2\,dt = \mu_{s_2}[0,b]$$

Thus, μ_{s_1} and μ_{s_2} correspond exactly to the conditional distributions consistent with $\bar{\mu}$ and r. If we let $\mu := (\mu_{s_1}, \mu_{s_2})$, then μ is well-calibrated.

Can the analyst directly test for well-calibrated beliefs? Let ρ be represented by (μ, u). Since u is fixed under ρ_s for all $s \in S$, both best and worst acts are well-defined for ρ. Given a state $s \in S$, define a *conditional worst act* that yields the worst act if s occurs and the best act otherwise.

Definition (Conditional worst act)**.** For $s \in S$, let $\underline{f}^s$ be such that $\underline{f}^s(s') = \underline{f}$ if $s' = s$ and $\underline{f}^s(s') = \overline{f}$ otherwise.

Define a *conditional test act* $f_s^a := a\underline{f}^s + (1-a)\overline{f}$ as the mixture between the best act and conditional worst act. Define *conditional test functions* as follows.

Definition. Given ρ, the *conditional test function* of $F \in \mathcal{K}$ is $F_\rho^s : [0, r_s] \to [0,1]$ where

$$F_\rho^s(r_s a) := \rho_{s, F \cup f_s^a}(F)$$

Conditional test functions specify conditional choice probabilities as we vary the conditional test act from the best act to the conditional worst act. As in unconditional test functions, as a increases, the conditional test act becomes less attractive so F_ρ^s increases. The domain of the conditional test function is scaled by a factor r_s. Call F_ρ^s *well-defined* iff $F_\rho^s(r_s) = 1$, so well-defined conditional test functions are cumulatives on the interval $[0, r_s]$. Let $\mathcal{K}_s$ denote all decision-problems with well-defined conditional test functions.

Theorem 6. *Let ρ be represented by (μ, u). Then μ is well-calibrated iff F_ρ^s and $F_{\bar{\rho}}$ share the same mean for all $F \in \mathcal{K}_s$ and $s \in S$.*

Proof. See Appendix. □

Theorem 6 equates well-calibrated beliefs with the requirement that both conditional and unconditional test functions have the same mean. It is a random choice characterization of rational beliefs. The following example illustrates.

Example 9. Let $S = \{s_1, s_2\}$ and again, associate each $q \in \Delta S$ with $t \in [0,1]$ such that $t = q_{s_1}$. Following Example 8, let $r = \left(\frac{1}{2}, \frac{1}{2}\right)$, $\bar{\mu}$ have density $6t(1-t)$ and μ_{s_1} have density $12t^2(1-t)$. Let $F := \{f, g\}$ where $u \circ f = \left(\frac{1}{4}, \frac{3}{4}\right)$ and $u \circ g = (0,1)$. Conditional on s_1, the probability of choosing something in F over a conditional test act $f_{s_1}^a$ is

$$\mu_{s_1}\left\{t \in [0,1] \;\middle|\; \max\left\{1-t,\ t\frac{1}{4} + (1-t)\frac{3}{4}\right\} \geq 1 - at\right\}$$

If we let $F_\rho^{s_1}$ be this conditional probability scaled by $r_{s_1} = \frac{1}{2}$, then $F_\rho^{s_1}(a) = 0$ for $a \leq \frac{3}{8}$, $F_\rho^{s_1}(a) = 1$ for $a \geq \frac{1}{2}$ and

$$F_\rho^{s_1}(a) = 1 + \frac{1}{2(1-4a)^3} + \frac{3}{16(1-4a)^4}$$

for $a \in \left(\frac{3}{8}, \frac{1}{2}\right)$. The unconditional test function $F_{\bar{\rho}}$ satisfies $F_{\bar{\rho}}(a) = (3-2a)a^2$ for $a \leq \frac{1}{2}$,

$F_{\bar{\rho}}(a) = (1-a)(1-4a)^2$ for $a \in \left(\frac{1}{2}, \frac{3}{4}\right)$ and $F_{\bar{\rho}}(a) = 1$ for $a \geq \frac{3}{4}$. Note that

$$\int_{[0,1]} a \; dF_{\rho}^{s_1} = \frac{29}{64} = \int_{[0,1]} a \; dF_{\bar{\rho}}$$

so both test functions have the same mean. This follows from the fact that μ_1 is well-calibrated.

Suppose that in addition to the cRCR ρ, the analyst also observes the valuation preference relation $\succeq$ over all decision-problems. In this case, if beliefs are well-calibrated, then any misalignment between $\succeq$ and ρ is no longer solely subjective. For example, in the individual interpretation, any prospective overconfidence (underconfidence) can now be interpreted as objective overconfidence (underconfidence) with respect to the true information structure. Hence, by enriching choice behavior with data on state realizations, the analyst can make objective claims about belief misalignment.

8 Related Literature

This paper is related to a long literature on stochastic choice. Information representation is a RUM model.[33] Testable implications of RUM were first studied by Block and Marschak [7], and the model was later characterized by McFadden and Richter [35], Falmagne [18] and Cohen [13]. Gul and Pesendorfer [26] obtain a more intuitive characterization by enriching the choice space with lotteries. More recently, Gul, Natenzon and Pesendorfer [25] characterize a special class of RUM models called attribute rules that can approximate any RUM model.

In relation to this literature, we show that a characterization of random expected utility can be comfortably extended to the realm of Anscombe-Aumann acts. The axioms of subjective expected utility yield intuitive analogs in random choice. Moreover, by allowing our RCR to be silent on acts that are indifferent, we are able to include deterministic choice as a special case of random choice, overcoming an issue that most RUM models have difficulty with. Finally, one could interpret Theorem 3 as presenting an alternative characterization of RUM via properties of its induced valuation preference relation.

[33] RUM is used extensively in discrete choice estimation. Most models in this literature assume specific parametrizations such as the logit, the probit, the nested logit, etc. (see Train [46]).

Some recent papers have also investigated the relationship between stochastic choice and information. Natenzon [37] studies a model where an agent gradually learns her own tastes. Caplin and Dean [8] and Matejka and McKay [33] study cRCRs where the agent exhibits rational inattention. Ellis [17] studies a similar model with partitional information so the resulting cRCR is deterministic. In contrast, the information structure in our model is fixed, which is closer to the standard model of information processing and choice. Note that since the information structure in these other models is allowed to vary with the decision-problem, the resulting random choice model is not necessarily a RUM model. Caplin and Martin [9] do characterize and test a model where the information structure is fixed. We can recast their model in our richer Anscombe-Aumann setup, in which case our conditions for a well-calibrated cRCR imply their conditions. Note that by working with a richer setup, our representation can be uniquely identified from choice behavior.

This paper is also related to the large literature on choice over menus (i.e. option sets). This line of research commenced with Kreps' [31] seminal paper on preference for flexibility and was extended to the lottery space by Dekel, Lipman and Rustichini [15] (henceforth DLR) and more recently to the Anscombe-Aumann space by DLST [16]. Our main contribution to this literature is showing that there is an intimate link between ex-ante choice *over* option sets (i.e. our valuation preference relation) and ex-post random choice *from* option sets. Theorem 4 can be interpreted as characterizing the ex-ante valuation preference relation via properties of its ex-post random choice. Ahn and Sarver [1] also study this relationship although in the lottery space. Their work connecting DLR preferences with Gul and Pesendorfer [26] random expected utility is analogous to our results connecting DLST preferences with our random choice model (their Axiom 1 is an ordinal version of Theorem 4). As our choice options reside in the richer Anscombe-Aumann space, we are able to achieve a much tighter connection between the two choice behaviors (we elaborate on this further in Appendix E). Fudenberg and Strzalecki [21] also analyze the relationship between preference for flexibility and random choice but in a dynamic setting with recursive random utilities. In contrast, in both Ahn and Sarver [1] and our model, the ex-ante choice over option sets is static. Saito [42] also establishes a relationship between greater preference for flexibility and more randomness, although the agent in his model deliberately randomizes due to ambiguity aversion.

Grant, Kajii and Polak [23, 24] also study decision-theoretic models involving information. However, they consider generalizations of the Kreps and Porteus [32] model where the agent has an intrinsic preference for information even when she is unable to or unwilling to act on that information. In contrast, in our model, the agent prefers information only as a result of its instrumental value as in the classical sense of Blackwell.

Another strand of related literature studies the various biases in regards to information processing. This includes the confirmatory bias (Rabin and Schrag [39]), the hot-hand fallacy (Gilovich, Vallone and Tversky [22]) and the gambler's fallacy (Rabin [38]). Our model can be applied to study all these behaviors. To see this, assume that ex-ante, the agent is immune to these biases but after receiving her signal, she becomes afflicted and exhibits ex-post random choice that reflects these biases. In this setup, the confirmatory bias and the hot-hand fallacy correspond to prospective underconfidence while the gambler's fallacy corresponds to prospective overconfidence. Corollary 4 also allows us to rank the severity of these biases via the Blackwell ordering of information structures. Finally, although not necessarily about biased information processing, the diversification bias (Read and Loewenstein [41]) can also be studied in this setup.

In the strategic setting, Bergemann and Morris [3] study information structures in Bayes' correlated equilibria. In the special case where there is a single bidder, our results translate directly to their setup for a single-person game. Thus, we could interpret our model as describing the actions of a bidder assuming that the bids of everyone else are held fixed. Kamenica and Gentzkow [**?**] and Rayo and Segal [40] characterize optimal information structures where a sender can control the information that a receiver gets. In these models, the sender's ex-ante utility is a function of the receiver's random choice rule. Our results relating random choice with valuations thus provide a technique for expressing the sender's utility in terms of the receiver's utility and vice-versa.

Finally, this paper is related to the recent literature on testing for private information in insurance markets. Hendren [27] uses elicited subjective beliefs from survey data to test whether there is more private information in one group of agents (insurance rejectees) than another group (non-rejectees). Under the group interpretation, Theorem 5 allows us to perform this same test by inferring beliefs directly from choice data. Also, we can interpret Theorem 6 as providing a sufficient condition for the presence of private information that is similar to tests for private information in the empirical literature (e.g. Chiappori and Salanié [11], Finkelstein and McGarry [20] and also Hendren [27]).

References

[1] D. Ahn and T. Sarver. Preference for flexibility and random choice. *Econometrica*, 81(1):341–361, 2013.

[2] F. Anscombe and R. Aumann. A definition of subjective probability. *The Annals of Mathematical Statistics*, 34(1):199–205, 1963.

[3] D. Bergemann and S. Morris. Bayes correlated equilibrium and the comparison of information structures. Mimeo, 2013.

[4] P. Billingsley. *Probability and Measure*. John Wiley and Sons, Inc., second edition, 1986.

[5] D. Blackwell. Comparison of experiments. In *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, pages 93–102, 1951.

[6] D. Blackwell. Equivalent comparisons of experiments. *The Annals of Mathematical Statistics*, 24(2):265–272, 1953.

[7] H. Block and J. Marschak. Random orderings and stochastic theories of response. In I. Olkin, editor, *Contributions to Probability and Statistics*, pages 97–132. Stanford University Press, 1960.

[8] A. Caplin and M. Dean. Rational inattention and state dependent stochastic choice. Mimeo, 2013.

[9] A. Caplin and D. Martin. A testable theory of imperfect perception. Mimeo, 2013.

[10] C. Chambers and N. Lambert. Dynamically eliciting unobservable information. Mimeo, 2014.

[11] P. Chiappori and B. Salanié. Testing for asymmetric information in insurance markets. *The Journal of Political Economy*, 108(1):56–78, 2000.

[12] E. Çinlar. *Probability and Stochastics*. Springer, 2011.

[13] M. Cohen. Random utility systems - the infinite case. *Journal of Mathematical Psychology*, 22:1–23, 1980.

[14] A. Dawid. The well-calibrated bayesian. *Journal of the American Statistical Association*, 77(379):605–610, 1982.

[15] E. Dekel, B. Lipman, and A. Rustichini. Representing preferences with a unique subjective state space. *Econometrica*, 69(4):891–934, 2001.

[16] D. Dillenberger, J. Lleras, P. Sadowski, and N. Takeoka. A theory of subjective learning. Mimeo, 2012.

[17] A. Ellis. Foundations for optimal inattention. Mimeo, 2012.

[18] J. Falmagne. A representation theorem for finite random scale systems. *Journal of Mathematical Psychology*, 18:52–72, 1978.

[19] A. Finkelstein, E. Luttmer, and M. Notowidigdo. Approaches to estimating the health state dependence of the utility function. *American Economic Review: Papers and Proceedings*, 99(2):116–121, 2009.

[20] A. Finkelstein and K. McGarry. Multiple dimensions of private information: Evidence from the long-term care insurance market. *The American Economic Review*, 96(4):938–958, 2006.

[21] D. Fudenberg and T. Strzalecki. Dynamic logit with choice aversion. Mimeo, 2014.

[22] T. Gilovich, R. Vallone, and A. Tversky. The hot hand in basketball: On the misperception of random sequences. *Cognitive Psychology*, 17:295–314, 1985.

[23] S. Grant, A. Kajii, and B. Polak. Intrinsic preference for information. *Journal of Economic Theory*, 83:233–259, 1998.

[24] S. Grant, A. Kajii, and B. Polak. Temporal resolution of uncertainty and recursive non-expected utility models. *Econometrica*, 68(2):425–434, 2000.

[25] F. Gul, P. Natenzon, and W. Pesendorfer. Random choice as behavioral optimization. *Econometrica*, 82(5):1873–1912, 2014.

[26] F. Gul and W. Pesendorfer. Random expected utility. *Econometrica*, 74:121–146, 2006.

[27] N. Hendren. Private information and insurance rejections. *Econometrica*, 81(5):1713–1762, 2013.

[28] D. Kahneman and A. Tversky. An analysis of decision under risk. *Econometrica*, 47:263–291, 1979.

[29] E. Karni. Foundations of bayesian theory. *Journal of Economic Theory*, 132:167–188, 2007.

[30] E. Karni, D. Schmeidler, and K. Vind. On state dependent preferences and subjective probabilities. *Econometrica*, 51(4):1021–1031, 1983.

[31] D. Kreps. Representation theorem for 'preference for flexibility'. *Econometrica*, 47:565–578, 1979.

[32] D. Kreps and E. Porteus. Temporal resolution of uncertainty and dynamic choice theory. *Econometrica*, 46(1):185–200, 1978.

[33] F. Matějka and A. McKay. Rational inattention to discrete choices: A new foundation for the multinomial logit model. Mimeo, 2013.

[34] D. McFadden. Econometric models of probabilistic choice. In C. Manksi and D. McFadden, editors, *Structural Analysis of Discrete Data with Econometric Applications*, chapter 5. MIT Press, 1981.

[35] D. McFadden and M. Richter. Stochastic rationality and revealed stochastic preference. In J.S. Chipman, D. McFadden, and M. Richter, editors, *Preferences, Uncertainty and Optimality*, pages 161–186. Westview Press, Inc., 1991.

[36] A. Müller and D. Stoyan. *Comparison Methods for Stochastic Models and Risks*. John Wiley and Sons, Inc., 2002.

[37] P. Natenzon. Random choice and learning. Mimeo, 2013.

[38] M. Rabin. Inference by believers in the law of small numbers. *The Quarterly Journal of Economics*, 117(3):775–816, 2002.

[39] M. Rabin and J. Schrag. First impressions matter: A model of confirmation bias. *The Quarterly Journal of Economics*, 114(1):37–82, 1999.

[40] L. Rayo and I. Segal. Optimal information disclosure. *The Journal of Political Economy*, 118(5):949–987, 2010.

[41] D. Read and G. Loewenstein. Diversification bias: Explaining the discrepancy in variety seeking between combined and separated choices. *Journal of Experimental Psychology: Applied*, 1:34–49, 1995.

[42] K. Saito. Preference for flexibility and preference for randomization under ambiguity. Mimeo, 2014.

[43] J. Savage. *The Foundations of Statistics*. John Wiley and Sons, Inc., 1954.

[44] R. Schneider. *Convex Bodies: The Brunn-Minkowski Theory*. Cambridge University Press, 1993.

[45] R. Strotz. Myopia and inconsistency in dynamic utility maximization. *The Review of Economic Studies*, 23(3):165–180, 1955.

[46] K. Train. *Discrete Choice Methods with Simulation*. Cambridge University Press, 2009.

Appendix A

A1. Representation Theorem

In this section of Appendix A, we prove the main representation theorem. Given a non-empty collection $\mathcal{G}$ of subsets of H and some $F \in \mathcal{K}$, define

$$\mathcal{G} \cap F := \{G \cap F | G \in \mathcal{G}\}$$

Note that if $\mathcal{G}$ is a σ-algebra, then $\mathcal{G} \cap F$ is the *trace* of $\mathcal{G}$ on $F \in \mathcal{K}$. For $G \subset F \in \mathcal{K}$, let

$$G_F := \bigcap_{G \subset G' \in \mathcal{H}_F} G'$$

denote the smallest $\mathcal{H}_F$-measurable set containing G.

Lemma (A1)**.** *Let $G \subset F \in \mathcal{K}$.*

(1) *$\mathcal{H}_F \cap F = \mathcal{H} \cap F$.*

(2) *$G_F = \hat{G} \cap F \in \mathcal{H}_F$ for some $\hat{G} \in \mathcal{H}$.*

(3) *$F \subset F' \in \mathcal{K}$ implies $G_F = G_{F'} \cap F$.*

Proof. Let $G \subset F \in \mathcal{K}$.

(1) Recall that $\mathcal{H}_F := \sigma(\mathcal{H} \cup \{F\})$ so $\mathcal{H} \subset \mathcal{H}_F$ implies $\mathcal{H} \cap F \subset \mathcal{H}_F \cap F$. Let

$$\mathcal{G} := \{G \subset H | G \cap F \in \mathcal{H} \cap F\}$$

We first show that $\mathcal{G}$ is a σ-algebra. Let $G \in \mathcal{G}$ so $G \cap F \in \mathcal{H} \cap F$. Now

$$\begin{aligned} G^c \cap F &= (G^c \cup F^c) \cap F = (G \cap F)^c \cap F \\ &= F \backslash (G \cap F) \in \mathcal{H} \cap F \end{aligned}$$

as $\mathcal{H} \cap F$ is the trace σ-algebra on F. Thus, $G^c \in \mathcal{G}$. For $G_i \subset \mathcal{G}$, $G_i \cap F \in \mathcal{H} \cap F$ so

$$\left(\bigcup_i G_i\right) \cap F = \bigcup_i (G_i \cap F) \in \mathcal{H} \cap F$$

Hence, $\mathcal{G}$ is an σ-algebra

Note that $\mathcal{H} \subset \mathcal{G}$ and $F \in \mathcal{G}$ so $\mathcal{H} \cup \{F\} \subset \mathcal{G}$. Thus, $\mathcal{H}_F = \sigma(\mathcal{H} \cup \{F\}) \subset \mathcal{G}$. Hence,

$$\mathcal{H}_F \cap F \subset \mathcal{G} \cap F = \{G' \cap F | G' = G \cap F \in \mathcal{H} \cap F\} \subset \mathcal{H} \cap F$$

so $\mathcal{H}_F \cap F = \mathcal{H} \cap F$.

(2) Since $\mathcal{H}_F \cap F \subset \mathcal{H}_F$, we have

$$G_F := \bigcap_{G \subset G' \in \mathcal{H}_F} G' \subset \bigcap_{G \subset G' \in \mathcal{H}_F \cap F} G'$$

Suppose $g \in \bigcap_{G \subset G' \in \mathcal{H}_F \cap F} G'$. Let G' be such that $G \subset G' \in \mathcal{H}_F$. Now, $G \subset G' \cap F \in \mathcal{H}_F \cap F$ so by the definition of g, we have $g \in G' \cap F$. Since this is true for all such G', we have $g \in G_F$. Hence,

$$G_F = \bigcap_{G \subset G' \in \mathcal{H}_F \cap F} G' = \bigcap_{G \subset G' \in \mathcal{H} \cap F} G'$$

where the second equality follows from (1). Since F is finite, we can find $\hat{G}_i \in \mathcal{H}$ where $G \subset \hat{G}_i \cap F$ for $i \in \{1, \ldots, k\}$. Hence,

$$G_F = \bigcap_i \left(\hat{G}_i \cap F \right) = \hat{G} \cap F$$

where $\hat{G} := \bigcap_i \hat{G}_i \in \mathcal{H}$. Note that $G_F \in \mathcal{H}_F$ follows trivially.

(3) By (2), let $G_F = \hat{G} \cap F$ and $G_{F'} = \hat{G}' \cap F'$ for $\left\{ \hat{G}, \hat{G}' \right\} \subset \mathcal{H}$. Since $F \subset F'$,

$$G \subset G_{F'} \cap F = \hat{G}' \cap F \in \mathcal{H}_F$$

so $G_F \subset G_{F'} \cap F$ by the definition of G_F. Now, by the definition of $G_{F'}$, $G_{F'} \subset \hat{G} \cap F' \in \mathcal{H}_{F'}$ so

$$G_{F'} \cap F \subset \left(\hat{G} \cap F' \right) \cap F = \hat{G} \cap F = G_F$$

Hence, $G_F = G_{F'} \cap F$.

□

Let ρ be a RCR. By Lemma A1, we can now define

$$\rho_F^* (G) := \inf_{G \subset G' \in \mathcal{H}_F} \rho_F (G') = \rho_F (G_F)$$

for $G \subset F \in \mathcal{K}$. Going forward, we simply let ρ denote ρ^* without loss of generality. We also employ the notation

$$\rho (F, G) := \rho_{F \cup G} (F)$$

for $\{F,G\} \subset \mathcal{K}$. We say that two acts are *tied* iff they are indifferent.

Definition. f and g are *tied* iff $\rho(f,g) = \rho(g,f) = 1$.

Lemma (A2)**.** *For $\{f,g\} \subset F \in \mathcal{K}$, the following are equivalent:*

(1) *f and g are tied*

(2) $g \in f_F$

(3) $f_F = g_F$

Proof. We prove that (1) implies (2) implies (3) implies (1). Let $\{f,g\} \subset F \in \mathcal{K}$. First, suppose f and g are tied so $\rho(f,g) = \rho(g,f) = 1$. If $f_{f \cup g} = f$, then $g = (f \cup g) \backslash f_F \in \mathcal{H}_{f \cup g}$ so $g_{f \cup g} = g$. As a result, $\rho(f,g) + \rho(g,f) = 2 > 1$ a contradiction. Thus, $f_{f \cup g} = f \cup g$. Now, since $f \cup g \subset F$, by Lemma A1, $f \cup g = f_{f \cup g} = f_F \cap (f \cup g)$ so $g \in f_F$. Hence, (1) implies (2).

Now, suppose $g \in f_F$ so $g \in g_F \cap f_F$. By Lemma A1, $g_F \cap f_F \in \mathcal{H}_F$ so $g_F \subset g_F \cap f_F$ which implies $g_F \subset f_F$. If $f \notin g_F$, then $f \in f_F \backslash g_F \in \mathcal{H}_F$. As a result, $f_F \subset f_F \backslash g_F$ implying $g_F = \emptyset$ a contradiction. Thus, $f \in g_F$, so $f \in g_F \cap f_F$ which implies $f_F \subset g_F \cap f_F$ and $f_F \subset g_F$. Hence, $f_F = g_F$ so (2) implies (3).

Finally, assume $f_F = g_F$ so $f \cup g \subset f_F$ by definition. By Lemma A1 again,

$$f_{f \cup g} = f_F \cap (f \cup g) = f \cup g$$

so $\rho(f,g) = \rho_{f \cup g}(f \cup g) = 1$. By symmetric reasoning, $\rho(g,f) = 1$ so f and g are tied. Thus, (1), (2) and (3) are all equivalent. □

Lemma (A3)**.** *Let ρ be monotonic.*

(1) *For $f \in F \in \mathcal{K}$, $\rho_F(f) = \rho_{F \cup g}(f)$ if g is tied with some $g' \in F$.*

(2) *Let $F := \bigcup_i f_i$, $G := \bigcup_i g_i$ and assume f_i and g_i are tied for all $i \in \{1,\dots,n\}$. Then $\rho_F(f_i) = \rho_G(g_i)$ for all $i \in \{1,\dots,n\}$.*

Proof. We prove the lemma in order:

(1) By Lemma A2, we can find unique $h^i \in F$ for $i \in \{1,\dots,k\}$ such that $\{h_F^1,\dots h_F^k\}$ forms a partition on F. Without loss of generality, assume g is tied with some $g' \in h_F^1$.

By Lemma A2 again, $h^1_{F\cup g} = h^1_F \cup g$ and $h^i_{F\cup g} = h^i_F$ for $i > 1$. By monotonicity, for all i

$$\rho_F\left(h^i_F\right) = \rho_F\left(h^i\right) \geq \rho_{F\cup g}\left(h^i\right) = \rho_{F\cup g}\left(h^i_{F\cup g}\right)$$

Now, for any $f \in h^j_F$, $f \in h^j_{F\cup g}$ and

$$\rho_F\left(f\right) = 1 - \sum_{i\neq j} \rho_F\left(h^i_F\right) \leq 1 - \sum_{i\neq j} \rho_{F\cup g}\left(h^i_{F\cup g}\right) = \rho_{F\cup g}\left(f\right)$$

By monotonicity again, $\rho_F\left(f\right) = \rho_{F\cup g}\left(f\right)$.

(2) Let $F := \bigcup_i f_i$, $G := \bigcup_i g_i$ and assume f_i and g_i are tied for all $i \in \{1, \dots, n\}$. From (1), we have

$$\rho_F\left(f_i\right) = \rho_{F\cup g_i}\left(f_i\right) = \rho_{F\cup g_i}\left(g_i\right) = \rho_{(F\cup g_i)\setminus f_i}\left(g_i\right)$$

Repeating this argument yields $\rho_F\left(f_i\right) = \rho_G\left(g_i\right)$ for all i.

□

For $\{F, F'\} \subset \mathcal{K}$, we use the condensed notation $FaF' := aF + (1-a)F'$.

Lemma (A4)**.** *Let ρ be monotonic and linear. For $f \in F \in \mathcal{K}$, let $F' := Fah$ and $f' := fah$ for some $h \in H$ and $a \in (0,1)$. Then $\rho_F\left(f\right) = \rho_{F'}\left(f'\right)$ and $f'_{F'} = f_F ah$.*

Proof. Note that $\rho_F\left(f\right) = \rho_{F'}\left(f'\right)$ follows directly from linearity, so we just need to prove that $f'_{F'} = f_F ah$. Let $g' := gah \in f_F ah$ for $g \in F$ tied with f. By linearity, $\rho\left(f', g'\right) = \rho\left(g', f'\right) = 1$ so g' is tied with f'. Thus, $g' \in f'_{F'}$ by Lemma A2 and $f_F ah \subset f'_{F'}$. Now, let $g' \in f'_{F'}$ so $g' = gah$ is tied with fah. By linearity again, f and g are tied so $g' \in f_F ah$. Thus, $f'_{F'} = f_F ah$. □

We now associate each act $f \in H$ with the vector $f \in [0,1]^{S\times X}$ without loss of generality. Find $\{f_1, g_1, \dots, f_k, g_k\} \subset H$ such that $f_i \neq g_i$ are tied and $z_i \cdot z_j = 0$ for all $i \neq j$, where $z_i := \frac{f_i - g_i}{\|f_i - g_i\|}$. Let $Z := \text{lin}\{z_1, \dots, z_k\}$ be the linear space spanned by all z_i with $Z = 0$ if no such z_i exists. Let k be maximal in that for any $\{f, g\} \subset H$ that are tied, $f - g \in Z$. Note that Lemmas A3 and A4 ensure that k is well-defined. Define $\varphi : H \to \mathbb{R}^{S\times X}$ such that

$$\varphi\left(f\right) := f - \sum_{1\leq i\leq k} \left(f \cdot z_i\right) z_i$$

and let $W := \text{lin}\left(\varphi\left(H\right)\right)$. Lemma A5 below shows that φ projects H onto a space without ties.

Lemma (A5)**.** *Let ρ be monotonic and linear.*

(1) $\varphi(f) = \varphi(g)$ *iff f and g are tied.*

(2) $w \cdot \varphi(f) = w \cdot f$ *for all $w \in W$.*

Proof. We prove the lemma in order

(1) First, suppose f and g are tied so $f - g \in Z$ by the definition of Z. Thus,

$$f = g + \sum_{1 \leq i \leq k} \alpha_i z_i$$

for some $\alpha \in \mathbb{R}^k$. Hence,

$$\begin{aligned} \varphi(f) &= g + \sum_{1 \leq i \leq k} \alpha_i z_i - \sum_{1 \leq i \leq k} \left[\left(g + \sum_{1 \leq j \leq k} \alpha_j z_j \right) \cdot z_i \right] z_i \\ &= g - \sum_{1 \leq i \leq k} (g \cdot z_i) z_i = \varphi(g) \end{aligned}$$

For the converse, suppose $\varphi(f) = \varphi(g)$ so

$$\begin{aligned} f - \sum_{1 \leq i \leq k} (f \cdot z_i) z_i &= g - \sum_{1 \leq i \leq k} (g \cdot z_i) z_i \\ f - g &= \sum_{1 \leq i \leq k} ((f - g) \cdot z_i) z_i \in Z \end{aligned}$$

and f and g are tied.

(2) Note that for any $f \in H$,

$$\varphi(f) \cdot z_i = 0$$

Since $W = \text{lin}(\varphi(H))$ and φ is linear, $w \cdot z_i = 0$ for all $w \in W$. Thus,

$$w \cdot \varphi(f) = w \cdot \left(f - \sum_{1 \leq i \leq k} (f \cdot z_i) z_i \right) = w \cdot f$$

for all $w \in W$.

□

Lemma (A6)**.** *If ρ satisfies Axioms 1-4, then there exists a measure ν on W such that*

$$\rho_F(f) = \nu \{ w \in W | \, w \cdot f \geq w \cdot g \;\forall g \in F \}$$

Proof. Let $m := \dim(W)$. Note that if $m = 0$, then $W = \varphi(H)$ is a singleton so everything is tied by Lemma A5 and the result follows trivially. Thus, assume $m \geq 1$ and let $\Delta^m \subset \mathbb{R}^{S \times X}$ be the m-dimensional probability simplex. Now, there exists an affine transformation $T = \lambda A$ where $\lambda > 0$, A is an orthogonal matrix and $T \circ \varphi(H) \subset \Delta^m$. Let $V := \operatorname{lin}(\Delta^m)$ so $T(W) = V$. Now, for each finite set $D \subset \Delta^m$, we can find a $p^* \in \Delta^m$ and $a \in (0,1)$ such that $Dap^* \subset T \circ \varphi(H)$. Thus, we can define a RCR τ on Δ^m such that

$$\tau_D(p) := \rho_F(f)$$

where $T \circ \varphi(F) = Dap^*$ and $T \circ \varphi(f) = pap^*$. Linearity and Lemma A5 ensure that τ is well-defined.

Since the projection mapping φ is linear, Axioms 1-4 correspond exactly to the axioms of Gul and Pesendorfer [26] on Δ^m. Thus, by their Theorem 3, there exists a measure ν_T on V such that for $F \in \mathcal{K}_0$

$$\begin{aligned} \rho_F(f) &= \tau_{T \circ \varphi(F)}(T \circ \varphi(f)) \\ &= \nu_T \{ v \in V | \, v \cdot (T \circ \varphi(f)) \geq v \cdot (T \circ \varphi(g)) \;\; \forall g \in F \} \end{aligned}$$

Since $A^{-1} = A'$,

$$v \cdot (T \circ \varphi(f)) = v \cdot \lambda A(\varphi(f)) = \lambda \left(A^{-1} v\right) \cdot \varphi(f) = \lambda^2 T^{-1}(v) \cdot \varphi(f)$$

Thus,

$$\begin{aligned} \rho_F(f) &= \nu_T \left\{ v \in V | \, T^{-1}(v) \cdot \varphi(f) \geq T^{-1}(v) \cdot \varphi(g) \;\; \forall g \in F \right\} \\ &= \nu \{ w \in W | \, w \cdot \varphi(f) \geq w \cdot \varphi(g) \;\; \forall g \in F \} \\ &= \nu \{ w \in W | \, w \cdot f \geq w \cdot g \; \forall g \in F \} \end{aligned}$$

where $\nu := \nu_T \circ T$ is the measure on W induced by T. Note that the last equality follows from Lemma A5.

Finally, consider any generic $F \in \mathcal{K}$, and let $F_0 \subset F$ be such that $f \in F_0 \in \mathcal{K}_0$. By Lemma A3,

$$\rho_F(f) = \rho_{F_0}(f) = \nu \{ w \in W | \, w \cdot f \geq w \cdot g \; \forall g \in F_0 \}$$

By Lemma A5, if h and g are tied, then

$$w \cdot h = w \cdot \varphi(h) = w \cdot \varphi(g) = w \cdot g$$

for all $w \in W$. Thus,

$$\rho_F(f) = \nu\{w \in W | \, w \cdot f \geq w \cdot g \; \forall g \in F\}$$

□

Henceforth, assume ρ satisfies Axioms 1-4 and let ν be the measure on W as specified by Lemma A6. We let $w_s \in \mathbb{R}^X$ denote the vector corresponding to $w \in W$ and $s \in S$. For $u \in \mathbb{R}^X$, define $R(u) \subset \mathbb{R}^X$ as the set of all $\alpha u + \beta \mathbf{1}$ for some $\alpha > 0$ and $\beta \in \mathbb{R}$. Let $U := \left\{ u \in \mathbb{R}^X \middle| \, u \cdot \mathbf{1} = 0 \right\}$ and note that $R(u) \cap U$ is the set of all αu for some $\alpha > 0$.
A state $s^* \in S$ is null iff it satisfies the following.

Definition. $s^* \in S$ is *null* iff $f_s = g_s$ for all $s \neq s^*$ implies $\rho_{F \cup f}(f) = \rho_{F \cup g}(g)$ for all $F \in \mathcal{K}$

Lemma (A7)**.** *If ρ is non-degenerate, then there exists a non-null state.*

Proof. Suppose ρ is non-degenerate but all $s \in S$ are null and consider $\{f, g\} \subset H$. Let $S = \{s_1, \ldots, s_n\}$ and for $0 \leq i \leq n$, define $f^i \in H$ such that $f^i_{s_j} = g_{s_j}$ for $j \leq i$ and $f^i_{s_j} = f_{s_j}$ for $j > i$. Note that $f^0 = f$ and $f^n = g$. By the definition of nullity, we have $\rho(f^i, f^{i+1}) = 1 = \rho(f^{i+1}, f^i)$ for all $i < n$. Thus, f^i and f^{i+1} are tied for all $i < n$ so by Lemma A2, f and g are tied. This implies $\rho(f, g) = 1$ for all $\{f, g\} \subset H$ contradicting non-degeneracy so there must exist at least one non-null state. □

Lemma (A8)**.** *Let ρ satisfy Axioms 1-4 and 8. Suppose $\{s_1, s_2\} \subset S$ are non-null. Define $\phi : W \to U \times U$ such that*

$$\phi_i(w) := w_{s_i} - \left(\frac{w_{s_i} \cdot \mathbf{1}}{|X|} \right) \mathbf{1}$$

for $i \in \{1, 2\}$ and $\eta := \nu \circ \phi^{-1}$ as the measure on $U \times U$ induced by ϕ. Then

(1) $\eta(\{0\} \times U) = \eta(U \times \{0\}) = 0$

(2) $\eta\{(u_1, u_2) \in U \times U \,| u_1 \cdot r > 0 > u_2 \cdot r\} = 0$ for any $r \in U$

(3) $\eta\{(u_1, u_2) \in U \times U \,| u_2 \in R(u_1)\} = 1$

Proof. We prove the lemma in order.

(1) Since s_1 is non-null, we can find $\{f,g\} \subset H$ such that $f_s = g_s$ for all $s \neq s_1$ and f and g are not tied. Let $f_{s_1} = p$ and $g_{s_1} = q$ so

$$\begin{aligned} 1 &= \rho(f,g) + \rho(g,f) \\ &= \nu\{w \in W \,|w_{s_1} \cdot p \geq w_{s_1} \cdot q\} + \nu\{w \in W \,|w_{s_1} \cdot q \geq w_{s_1} \cdot p\} \\ 0 &= \nu\{w \in W \,|w_{s_1} \cdot r = 0\} = \eta(\{u_1 \in U|\, u_1 \cdot r = 0\} \times U) \end{aligned}$$

for $r := p - q$. Since we can assume η is complete, $\eta(\{0\} \times U) = 0$. The case for s_2 is symmetric.

(2) For any $\{p,q\} \subset \Delta X$, let $\{f,g,h\} \subset H$ be such that $f_{s_1} = f_{s_2} = h_{s_1} = p$, $g_{s_1} = g_{s_2} = h_{s_2} = q$ and $f_s = g_s = h_s$ for all $s \notin \{s_1, s_2\}$. First, suppose h is not tied with either f nor g. Hence, by S-independence,

$$\begin{aligned} 0 = \rho_{\{f,g,h\}}(h) &= \nu\{w \in W \,|w \cdot h \geq \max(w \cdot f,\ w \cdot g)\} \\ &= \nu\{w \in W \,|w_{s_2} \cdot q \geq w_{s_2} \cdot p \text{ and } w_{s_1} \cdot p \geq w_{s_1} \cdot q\} \\ &= \nu\{w \in W \,|w_{s_1} \cdot r \geq 0 \geq w_{s_2} \cdot r\} \end{aligned}$$

for $r := p - q \in U$. Note that if h is tied with g, then

$$\begin{aligned} 1 = \rho(g,h) = \rho(h,g) &= \nu\{w \in W \,|w \cdot h = w \cdot g\} \\ &= \nu\{w \in W \,|w_{s_1} \cdot r = 0\} \end{aligned}$$

Symmetrically, if h is tied with f, then $w_{s_2} \cdot r = 0$ ν-a.s., so we have

$$\begin{aligned} 0 &= \nu\{w \in W \,|w_{s_1} \cdot r > 0 > w_{s_2} \cdot r\} \\ &= \nu\{w \in W \,|\phi_1(w) \cdot r > 0 > \phi_2(w) \cdot r\} \\ &= \eta\{(u_1, u_2) \in U \times U \,|u_1 \cdot r > 0 > u_2 \cdot r\} \end{aligned}$$

for any $r \in U$ without loss of generality.

(3) First, define the closed halfspace corresponding to $r \in U$ as

$$H_r := \{u \in U|\, u \cdot r \geq 0\}$$

and let $\mathcal{E}$ be the set of all finite intersection of such halfspaces. Consider a partition $\mathcal{P} = \{0\} \cup \bigcup_i A_i$ of U where for each A_i, we can find two sequences $A_{ij} \in \mathcal{E}$ and $\bar{A}_{ij} \in \mathcal{E}$

such that $A_{ij} \nearrow A_i \cup \{0\}$, $A_{ij} \subset \text{int}\left(\bar{A}_{ij}\right) \cup \{0\}$ and $\bar{A}_{ij} \cap \bar{A}_{i'j} = \{0\}$ for all $i' \neq i$. Note that since sets in $\mathcal{E}$ are η-measurable, every $A_{ij} \times A_{i'j'}$ is η-measurable. By (1)

$$\begin{aligned}
1 = \eta\left(U \times U\right) &= \eta\left(\bigcup_i A_i \times \bigcup_i A_i\right) = \sum_{ii'} \eta\left(A_i \times A_{i'}\right) \\
&= \sum_i \eta\left(A_i \times A_i\right) + \sum_{i' \neq i} \eta\left(A_i \times A_{i'}\right) \\
&= \eta\left(\bigcup_i \left(A_i \times A_i\right)\right) + \sum_{i' \neq i} \lim_j \eta\left(A_{ij} \times A_{i'j}\right)
\end{aligned}$$

By a standard separating hyperplane argument (Theorem 1.3.8 of Schneider [44]), we can find some $r \in U$ such that $u_1 \cdot r \geq 0 \geq u_2 \cdot r$ for all $(u_1, u_2) \in \bar{A}_{ij} \times \bar{A}_{i'j}$. Since $A_{ij} \backslash \{0\} \subset \text{int}\left(\bar{A}_{ij}\right)$, we must have $u_1 \cdot r > 0 > u_2 \cdot r$ for all $(u_1, u_2) \in (A_{ij} \backslash \{0\}) \times (A_{i'j} \backslash \{0\})$. By (1) and (2),

$$\begin{aligned}
\eta\left(A_{ij} \times A_{i'j}\right) &= \eta\left(\left(A_{ij} \backslash \{0\}\right) \times \left(A_{i'j} \backslash \{0\}\right)\right) \\
&\leq \eta\left\{\left(u_1, u_2\right) \in U \times U \left| u_1 \cdot r > 0 > u_2 \cdot r\right.\right\} = 0
\end{aligned}$$

so $\eta\left(\bigcup_i \left(A_i \times A_i\right)\right) = 1$.

Now, consider a sequence of increasingly finer such partitions $\mathcal{P}^k := \{0\} \cup \bigcup_i A_i^k$ such that for any $(u_1, u_2) \in U \times U$ where $u_2 \notin R(u_1)$, there is some partition $\mathcal{P}^k$ where $(u_1, u_2) \in A_i^k \times A_{i'}^k$ for $i \neq i'$. Let

$$\begin{aligned}
C_k &:= \{0\} \cup \bigcup_i \left(A_i^k \times A_i^k\right) \\
C_0 &:= \left\{\left(u_1, u_2\right) \in U \times U \left| u_2 \in R\left(u_1\right)\right.\right\}
\end{aligned}$$

We show that $C_k \searrow C_0$. Since $\mathcal{P}^{k'} \subset \mathcal{P}^k$ for $k' \geq k$, $C_{k'} \subset C_k$. Note if $u_2 \in R(u_1)$, then $u_1 \in H_r$ iff $u_2 \in H_r$ for all $r \in U$ so $C_0 \subset C_k$ for all k. Suppose $(u_1, u_2) \in \left(\bigcap_k C_k\right) \backslash C_0$. Since $u_2 \notin R(u_1)$, there is some k such that $(u_1, u_2) \notin C_k$ a contradiction. Hence, $C_0 = \bigcap_k C_k$ so

$$\eta\left(C_0\right) = \lim_k \eta\left(C_k\right) = 1$$

□

Theorem (A9)**.** *If ρ satisfies Axioms 1-5 and 8, then it has a RSEU representation.*

Proof. Let ρ satisfy Axioms 1-5 and 8, and ν be the measure on W as specified by Lemma

A6. Let $S^* \subset S$ be the set of non-null states with some $s^* \in S^*$ as guaranteed by Lemma A7. Define

$$W_0 := \{w \in W \,|w_s \in R(w_{s^*}) \;\; \forall s \in S^*\}$$

and note that by Lemma A8,

$$\eta(W_0) = \eta\left(\bigcap_{s \in S^*} \{w \in W \,|w_s \in R(w_{s^*})\}\right) = 1$$

Let $Q : W_0 \to \Delta S$ be such that $Q_s(w) := 0$ for $s \in S \backslash S^*$ and

$$Q_s(w) := \frac{\alpha_s(w)}{\sum_{s \in S^*} \alpha_s(w)}$$

for $s \in S^*$ where $w_s = \alpha_s(w) w_{s^*} + \beta_s(w) \mathbf{1}$ for $\alpha_s(w) > 0$ and $\beta_s(w) \in \mathbb{R}$. Define $\hat{Q} : W_0 \to \Delta S \times \mathbb{R}^X$ such that

$$\hat{Q}(w) := (Q(w), w_{s^*})$$

and let $\pi := \eta \circ \hat{Q}^{-1}$ be the measure on $\Delta S \times \mathbb{R}^X$ induced by $\hat{Q}$.

For $s \in S \backslash S^*$, let $\{f, h\} \subset H$ be such that $h_s = \frac{1}{|X|}\mathbf{1}$ and $f_{s'} = h_{s'}$ for all $s' \neq s$. By the definition of nullity, f and g are tied so

$$1 = \rho(f, h) = \rho(h, f) = \nu\left\{w \in W \,\middle|\, w_s \cdot f(s) = \frac{1}{|X|}(w_s \cdot \mathbf{1})\right\}$$

Thus

$$\begin{aligned} \rho_F(f) &= \nu\left\{w \in W \,\middle|\, \sum_{s \in S} w_s \cdot f(s) \geq \sum_{s \in S} w_s \cdot g(s) \;\; \forall g \in F\right\} \\ &= \nu\left\{w \in W_0 \,\middle|\, \sum_{s \in S^*} w_s \cdot f(s) \geq \sum_{s \in S^*} w_s \cdot g(s) \;\; \forall g \in F\right\} \\ &= \pi\left\{(q, u) \in \Delta S \times \mathbb{R}^X \,|q \cdot (u \circ f) \geq q \cdot (u \circ g) \;\; \forall g \in F\right\} \end{aligned}$$

Note that Lemma A8 implies that u is non-constant. Finally, we show that π is regular. Suppose $\exists \{f, g\} \subset H$ such that

$$\pi\left\{(q, u) \in \Delta S \times \mathbb{R}^X \,\middle|\, q \cdot (u \circ f) = q \cdot (u \circ g)\right\} \in (0, 1)$$

If f and g are tied, then $q \cdot (u \circ f) = q \cdot (u \circ g)$ π-a.s. yielding a contradiction. Since f and

g are not tied, then

$$\pi\left\{(q,u)\in\Delta S\times\mathbb{R}^X \,\middle|\, q\cdot(u\circ f)=q\cdot(u\circ g)\right\}=\rho(f,g)-(1-\rho(g,f))=0$$

a contradiction. Thus, ρ is represented by π. □

Theorem (A10)**.** *If ρ has a RSEU representation, then it satisfies Axioms 1-5 and 8.*

Proof. Note that monotonicity, linearity and extremeness all follow trivially from the representation. Note that if ρ is degenerate, then for any constant $\{f,g\}\subset H$,

$$1=\rho(f,g)=\rho(g,f)=\pi\left\{(q,u)\in\Delta S\times\mathbb{R}^X \,\middle|\, u\circ f=u\circ g\right\}$$

so u is constant π-a.s. a contradiction. Thus, non-degeneracy is satisfied.

To show S-independence, suppose $f_{s_1}=f_{s_2}=h_{s_1}$, $g_{s_1}=g_{s_2}=h_{s_2}$ and $f_s=g_s=h_s$ for all $s\notin\{s_1,s_2\}$. Note that if h is tied with f or g, then the result follows immediately, so assume h is tied to neither. Thus,

$$\begin{aligned}\rho_{\{f,g,h\}}(h)&=\pi\left\{(q,u)\in\Delta S\times\mathbb{R}^X \,\middle|\, q\cdot(u\circ h)\geq\max\left(q\cdot(u\circ g),q\cdot(u\circ g)\right)\right\}\\&=\pi\left\{(q,u)\in\Delta S\times\mathbb{R}^X \,\middle|\, u(h_{s_2})\geq u(h_{s_1})\text{ and }u(h_{s_1})\geq u(h_{s_2})\right\}\end{aligned}$$

Note that if $u(h_{s_2})=u(h_{s_1})$ π-a.s., then h is tied with both, so by the regularity of π, $\rho_{\{f,g,h\}}(h)=0$.

Finally, we show continuity. First, consider $\{f,g\}\subset F_k\in\mathcal{K}_0$ such that $f\neq g$ and suppose $q\cdot(u\circ f)=q\cdot(u\circ g)$ π-a.s.. Thus, $\rho(f,g)=\rho(g,f)=1$ so f and g are tied. As ρ is monotonic, Lemma A2 implies $g\in f_{F_k}$ contradicting the fact that $F_k\in\mathcal{K}_0$. As μ is regular, $q\cdot(u\circ f)=q\cdot(u\circ g)$ with π-measure zero and the same holds for any $\{f,g\}\subset F\in\mathcal{K}_0$. Now, for $G\in\mathcal{K}$, let

$$Q_G:=\bigcup_{\{f,g\}\subset G,\ f\neq g}\left\{(q,u)\in\Delta S\times\mathbb{R}^X \,\middle|\, q\cdot(u\circ f)=q\cdot(u\circ g)\right\}$$

and let

$$\bar{Q}:=Q_F\cup\bigcup_k Q_{F_k}$$

Thus, $\mu(\bar{Q})=0$ so $\mu(Q)=1$ for $Q:=\Delta S\backslash\bar{Q}$. Let $\hat{\pi}(A)=\pi(A)$ for $A\in\mathcal{B}\left(\Delta S\times\mathbb{R}^X\right)\cap Q$. Thus, $\hat{\pi}$ is the restriction of π to Q (see Exercise I.3.11 of Çinlar [12]).

Now, for each F_k, let $\xi_k : Q \to H$ be such that

$$\xi_k(q,u) := \arg\max_{f \in F_k} q \cdot (u \circ f)$$

and define ξ similarly for F. Note that both ξ_k and ξ are well-defined as they have domain Q. For any $B \in \mathcal{B}(H)$,

$$\begin{aligned}
\xi_k^{-1}(B) &= \{(q,u) \in Q \mid \xi_k(q,u) \in B \cap F_k\} \\
&= \bigcup_{f \in B \cap F_k} \left\{(q,u) \in \Delta S \times \mathbb{R}^X \middle| q \cdot (u \circ f) > q \cdot (u \circ g) \ \ \forall g \in F_k\right\} \cap Q \\
&\in \mathcal{B}\left(\Delta S \times \mathbb{R}^X\right) \cap Q
\end{aligned}$$

Hence, ξ_k and ξ are random variables. Moreover,

$$\begin{aligned}
\hat{\pi} \circ \xi_k^{-1}(B) &= \sum_{f \in B \cap F_k} \hat{\pi}\{(q,u) \in Q \mid q \cdot (u \circ f) > q \cdot (u \circ g) \ \ \forall g \in F_k\} \\
&= \sum_{f \in B \cap F_k} \pi\left\{(q,u) \in \Delta S \times \mathbb{R}^X \middle| q \cdot (u \circ f) \geq q \cdot (u \circ g) \ \ \forall g \in F_k\right\} \\
&= \rho_{F_k}(B \cap F_k) = \rho_{F_k}(B)
\end{aligned}$$

so ρ_{F_k} and ρ_F are the distributions of ξ_k and ξ respectively. Finally, let $F_k \to F$ and fix $(q,u) \in Q$. Let $f := \xi(q,u)$ so $q \cdot (u \circ f) > q \cdot (u \circ g)$ for all $g \in F$. Since linear functions are continuous, there is some $l \in \mathbb{N}$ such that $q \cdot (u \circ f_k) > q \cdot (u \circ g_k)$ for all $k > l$. Thus, $\xi_k(q,u) = f_k \to f = \xi(q,u)$ so ξ_k converges to ξ $\hat{\pi}$-a.s.. Since almost sure convergence implies convergence in distribution (see Exercise III.5.29 of Çinlar [12]), $\rho_{F_k} \to \rho_F$ and continuity is satisfied. □

Corollary (A11)**.** *ρ satisfies Axioms 1-7 iff it has an information representation.*

Proof. We first prove sufficiency. Note that if ρ satisfies Axioms 1-6 and 8, then by Theorem A9, ρ has an information representation. We show that Axioms 1-7 imply Axiom 8. Suppose $f_{s_1} = f_{s_2} = h_{s_1}$, $g_{s_1} = g_{s_2} = h_{s_2}$ and $f_s = g_s = h_s$ for all $s \notin \{s_1, s_2\}$. Note that if h is tied with f or g, then the result follows immediately, so assume h is tied to neither. Note that if h_{s_1} and h_{s_2} are tied, then S-monotonicity implies h is tied to both, so assume $\rho(h_{s_1}, h_{s_2}) = 1$ without loss of generality. By S-monotonicity again, $\rho(f,h) = 1$ implying $\rho(h,f) = 0$. Thus, $\rho_{\{f,g,h\}}(h) = 0$ so Axiom 8 is satisfied.

For necessity, note that Axioms 1-5 all follow from Theorem A10. C-determinism follows trivially from the representation. To show S-monotonicity, suppose $\rho_{F_s}(f_s) = 1$ for all $s \in S$. Thus, $u(f_s) \geq u(g_s)$ for all $g \in F$ and $s \in S$ which implies $q \cdot (u \circ f) \geq q \cdot (u \circ g)$ for all $g \in F$. Hence, $\rho_F(f) = 1$ from the representation yielding S-monotonicity. □

A2. Uniqueness

In this section of Appendix A, we use test functions to prove the uniqueness properties of information representations. Let $H_c \subset H$ denote the set of all constant acts.

Lemma (A12)**.** *Let ρ be represented by (μ, u). Then for any measurable $\phi : \mathbb{R} \to \mathbb{R}$,*

$$\int_{[0,1]} \phi dF_\rho = \int_{\Delta S} \phi \left(\frac{u(\overline{f}) - \sup_{f \in F} q \cdot (u \circ f)}{u(\overline{f}) - u(\underline{f})} \right) \mu(dq)$$

Proof. For $F \in \mathcal{K}$, let $\psi_F : \Delta S \to [0,1]$ be such that $\psi_F(q) = \frac{u(\overline{f}) - \sup_{f \in F} q \cdot (u \circ f)}{u(\overline{f}) - u(\underline{f})}$ which is measurable. Let $\lambda^F := \mu \circ \psi_F^{-1}$ be the image measure on $[0,1]$. By a standard change of variables (Theorem I.5.2 of Çinlar [12]),

$$\int_{[0,1]} \phi(x) \lambda^F(dx) = \int_{\Delta S} \phi(\psi_F(q)) \mu(dq)$$

We now show that the cumulative distribution function of λ^F is exactly F_ρ. For $a \in [0,1]$, let $f^a := \underline{f} a \overline{f} \in H_c$. Now,

$$\begin{aligned}
\lambda^F[0,a] &= \mu \circ \psi_F^{-1}[0,a] = \mu\{q \in \Delta S \mid a \geq \psi_F(q) \geq 0\} \\
&= \mu \left\{ q \in \Delta S \;\middle|\; \sup_{f \in F} q \cdot (u \circ f) \geq u(f^a) \right\}
\end{aligned}$$

First, assume f^a is tied with nothing in F. Since μ is regular, $\mu\{q \in \Delta S | u(f_a) = q \cdot (u \circ f)\} = 0$ for all $f \in F$. Thus,

$$\begin{aligned}
\lambda^F[0,a] &= 1 - \mu\{q \in \Delta S | u(f_a) \geq q \cdot (u \circ f) \;\; \forall f \in F\} \\
&= 1 - \rho(f^a, F) = \rho(F, f^a) = F_\rho(a)
\end{aligned}$$

Now, assume f^a is tied with some $g \in F$ so $u(f^a) = q \cdot (u \circ g)$ μ-a.s.. Thus, $f^a \in g_{F \cup f^a}$ so

$$F_\rho(a) = \rho(F, f^a) = 1 = \lambda^F[0,a]$$

Hence, $\lambda^F[0,a] = F_\rho(a)$ for all $a \in [0,1]$. Note that $\lambda^F[0,1] = 1 = F_\rho(1)$ so F_ρ is the cumulative distribution function of λ^F. $\square$

For convenience, we define the following.

Definition. $F \geq_m G$ iff $\int_{\mathbb{R}} x dF(x) \geq \int_{\mathbb{R}} x dG(x)$.

Lemma (A13)**.** *Let ρ and τ be represented by (μ, u) and (ν, v) respectively. Then the following are equivalent:*

(1) $u = \alpha v + \beta$ *for* $\alpha > 0$

(2) $f_\rho = f_\tau$ *for all* $f \in H_c$

(3) $f_\rho =_m f_\tau$ *for all* $f \in H_c$

Proof. For $f \in H_c$, let $\hat{u}(f) := \frac{u(\overline{f}) - u(f)}{u(\overline{f}) - u(\underline{f})}$ and note that

$$f_\rho(a) = \rho\left(f,\ \underline{f} a \overline{f}\right) = \mathbf{1}_{[\hat{u}(f),1]}(a)$$

Thus, the distribution of f_ρ is a Dirac measure at $\{\hat{u}(f)\}$ so

$$\int_{[0,1]} a\ df_\rho(a) = \hat{u}(f)$$

and $\lambda^f_\rho = \delta_{\left\{\int_{[0,1]} df_\rho(a) a\right\}}$. Hence, $\lambda^f_\rho = \lambda^f_\tau$ iff $f_\rho =_m f_\tau$ so (2) and (3) are equivalent.

We now show that (1) and (3) are equivalent. Let $\succeq^c_\rho$ and $\succeq^c_\tau$ be the two preference relations induced on H_c by ρ and τ respectively, and let $\left(\underline{f}, \overline{f}\right)$ and $\left(\underline{g}, \overline{g}\right)$ denote their respective worst and best acts. If (1) is true, then we can take $\left(\underline{f}, \overline{f}\right) = \left(\underline{g}, \overline{g}\right)$. Thus, for $f \in H_c$

$$\int_{[0,1]} a\ df_\rho(a) = \hat{u}(f) = \hat{v}(f) = \int_{[0,1]} a\ df_\tau(a)$$

so (3) is true. Now, suppose (3) is true. For any $f \in H_c$, we can find $\{\alpha, \beta\} \subset [0,1]$ such that $\underline{f} \alpha \overline{f} \sim^c_\rho f \sim^c_\tau \underline{g} \beta \overline{g}$. Note that

$$\alpha = \hat{u}(f) = \int_{[0,1]} a\ df_\rho(a) = \int_{[0,1]} a\ df_\tau(a) = \hat{v}(f) = \beta$$

so $f \sim^c_\rho \underline{f} \alpha \overline{f}$ iff $f \sim^c_\tau \underline{g} \alpha \overline{g}$. As a result, $f \succeq^c_\rho g$ iff $\underline{f} \alpha \overline{f} \succeq^c_\rho \underline{f} \beta \overline{f}$ iff $\beta \geq \alpha$ iff $\underline{g} \alpha \overline{g} \succeq^c_\tau \underline{g} \beta \overline{g}$ iff $f \succeq^c_\tau g$. Thus, $\rho = \tau$ on H_c so $u = \alpha v + \beta$ for $\alpha > 0$. Hence, (1), (2) and (3) are all equivalent. $\square$

Theorem (A14)**.** *Let ρ and τ be represented by (μ, u) and (ν, v) respectively. Then the following are equivalent:*

(1) $(\mu, u) = (\nu, \alpha v + \beta)$ *for* $\alpha > 0$

(2) $\rho = \tau$

(3) $\rho(f, g) = \tau(f, g)$ *for all* $\{f, g\} \subset H$

(4) $f_\rho = f_\tau$ *for all* $f \in H$

Proof. Let ρ and τ be represented by (μ, u) and (ν, v) respectively. If (1) is true, then $\rho_F(f) = \tau_F(f)$ for all $f \in H$ from the representation. Moreover, since $\rho(f, g) = \rho(g, f) = 1$ iff $\tau(f, g) = \tau(g, f) = 1$ iff f and g are tied, the partitions $\{f_F\}_{f \in F}$ agree under both ρ and τ. Thus, $\mathcal{H}_F^\rho = \mathcal{H}_F^\tau$ for all $F \in \mathcal{K}$ so $\rho = \tau$ and (2) is true. Note that (2) implies (3) implies (4) trivially.

Hence, all that remains is to prove that (4) implies (1). Assume (4) is true so $f_\rho = f_\tau$ for all $f \in H$. By Lemma A13, this implies $u = \alpha v + \beta$ for $\alpha > 0$. Thus, without loss of generality, we can assume $1 = u\left(\overline{f}\right) = v\left(\overline{f}\right)$ and $0 = u\left(\underline{f}\right) = v\left(\underline{f}\right)$ so $u = v$. Now,

$$\psi_f(q) := 1 - q \cdot (u \circ f) = 1 - q \cdot (v \circ f)$$

where $\psi_f : \Delta S \to [0, 1]$. Let $\lambda_\rho^f = \mu \circ \psi_f^{-1}$ and $\lambda_\tau^f = \nu \circ \psi_f^{-1}$, so by the lemma above, they correspond to the cumulatives f_ρ and f_τ. Now, by Ionescu-Tulcea's extension (Theorem IV.4.7 of Çinlar [12]), we can create a probability space on Ω with two independent random variables $X : \Omega \to \Delta S$ and $Y : \Omega \to \Delta S$ such that they have distributions μ and ν respectively. Let $\phi(a) = e^{-a}$, and since $f_\rho = f_\tau$, by Lemma A12,

$$\begin{aligned}
\mathbb{E}\left[e^{-\psi_f(X)}\right] &= \int_{\Delta S} e^{-\psi_f(q)} \mu(dq) \\
&= \int_{[0,1]} e^{-a}\, df_\rho(a) = \int_{[0,1]} e^{-a}\, df_\tau(a) \\
&= \int_{\Delta S} e^{-\psi_f(q)} \nu(dq) = \mathbb{E}\left[e^{-\psi_f(Y)}\right]
\end{aligned}$$

for all $f \in H$. Let $w_f \in [0, 1]^S$ be such that $w_f = \mathbf{1} - u \circ f$ so $\psi_f(q) = q \cdot w_f$. Since this is true for all $f \in H$, we have $\mathbb{E}\left[e^{-w \cdot X}\right] = \mathbb{E}\left[e^{-w \cdot Y}\right]$ for all $w \in [0, 1]^S$. Since Laplace transforms completely characterize distributions (see Exercise II.2.36 of Çinlar [12]), X and Y have the same distribution, so $\mu = \nu$. Thus, $(\mu, u) = (\nu, \alpha v + \beta)$ for $\alpha > 0$ and (1) is true. Hence, (1) to (4) are all equivalent. □

Lemma A15 below shows that (1) every decision-problem is arbitrarily (Hausdorff) close to some decision-problem in $\mathcal{K}_0$, and (2) ρ is discontinuous at precisely those decision-problems that contain ties (indifferences).

Lemma (A15)**.** *Let ρ have an information representation.*

(1) *$\mathcal{K}_0$ is dense in $\mathcal{K}$.*

(2) *f and g are not tied iff $f_k \to f$ and $g_k \to g$ imply $\rho(f_k, g_k) \to \rho(f, g)$.*

Proof. Let ρ be represented by (μ, u). We prove the lemma in order:

(1) Consider $F \in \mathcal{K}$. For each $\{f_i, g_i\} \subset F$ tied and $f_i \neq g_i$, let

$$z_i := u \circ f_i - u \circ g_i$$

so $q \cdot z_i = 0$ μ-a.s.. Let $q^* \in \Delta S$ be in the support of μ so $q^* \cdot z_i = 0$ for all i. Now, for every $f \in \hat{F} := \{ f \in F | \, f_F \neq f\}$, let $\varepsilon_f > 0$ and $f' \in H$ be such that

$$u \circ f' = u \circ f + \varepsilon_f q^*$$

Since F is finite, we can assume $\varepsilon_f \neq \varepsilon_g$ for all $\{f, g\} \subset \hat{F}$ such that $f \neq g$. Suppose f' and g' are tied, so μ-a.s.

$$0 = q \cdot (u \circ f' - u \circ g') = q \cdot (z_i + (\varepsilon_f - \varepsilon_g) q^*) = (\varepsilon_f - \varepsilon_g) q \cdot q^*$$

Thus, $q \cdot q^* = 0$ μ-a.s.. Since $q^* \cdot q^* \neq 0$, q^* is not in the support of μ yielding a contradiction. If we let $f' := f$ for $f \in F \backslash \hat{F}$, then $F' := \bigcup_{f \in F} f' \in \mathcal{K}_0$. Setting $\varepsilon_f^k \to 0$ for all $f \in \hat{F}$ yields that $F'_k \to F$. Thus, $\mathcal{K}_0$ is dense in $\mathcal{K}$.

(2) First, let f and g not be tied and $f_k \to f$ and $g_k \to g$. Suppose there is some subsequence j such that all f_j and g_j are tied. Let

$$z_j := u \circ f_j - u \circ g_j$$

and $\bar{Z} := \text{lin}\left(\bigcup_j z_j\right) \cap [0,1]^S$. Let $z := u \circ f - u \circ g$ and since f and g are not tied, $z \notin \bar{Z}$ by linearity. Thus, z and $\bar{Z}$ can be strongly separated (see Theorem 1.3.7 of Schneider [44]), but $z_j \to z$ yielding a contradiction. Hence, there is some $m \in \mathbb{N}$ such that f_k and g_k are not tied for all $k > m$. Continuity yields $\rho(f_k, g_k) \to \rho(f, g)$.

Finally, suppose f and g are tied. Without loss of generality, let $f_\varepsilon \in H$ be such that

$$u \circ f_\varepsilon = u \circ f - \varepsilon \mathbf{1}$$

for some $\varepsilon > 0$. By S-monotonicity, $\rho(f_\varepsilon, f) = \rho(f_\varepsilon, g) = 0$. Thus, if we let $\varepsilon \to 0$ and $f_\varepsilon \to f$, then

$$\rho(f_\varepsilon, g) \to 0 < 1 = \rho(f, g)$$

violating continuity.

□

Appendix B

In Appendix B, we prove our results relating valuations with random choice. In this section, consider RCRs ρ such that there are $\left\{\underline{f}, \bar{f}\right\} \subset H_c$ where $\rho\left(\overline{f}, f\right) = \rho\left(f, \underline{f}\right) = 1$ for all $f \in H$ and F_ρ is a cumulative distribution function for all $F \in \mathcal{K}$. For $a \in [0,1]$, define $f^a := \underline{f} a \overline{f}$.

Lemma (B1)**.** *For any cumulative F on $[0,1]$,*

$$\int_{[0,1]} F(a)\, da = 1 - \int_{[0,1]} a \; dF(a)$$

Proof. By Theorem 18.4 of Billingsley [4], we have

$$\int_{(0,1]} a \; dF(a) = F(1) - \int_{(0,1]} F(a)\, da$$

The result then follows immediately. □

Lemma (B2)**.** *For cumulatives F and G on $[0,1]$, $F = G$ iff $F = G$ a.e..*

Proof. Note that sufficiency is trivial so we prove necessity. Let λ be the Lebesgue measure and $D := \{b \in [0,1] | \, F(b) \neq F(G)\}$ so $\lambda(D) = 0$. For each $a < 1$ and $\varepsilon > 0$ such that $a + \varepsilon \leq 1$, let $B_{a,\varepsilon} := (a, a+\varepsilon)$. Suppose $F(b) \neq G(b)$ for all $b \in B_{a,\varepsilon}$. Thus, $B_{a,\varepsilon} \subset D$ so

$$0 < \varepsilon = \lambda(B_{a,\varepsilon}) \leq \lambda(D)$$

a contradiction. Thus, there is some $b \in B_{a,\varepsilon}$ such that $F(b) = G(b)$ for all such a and ε. Since both F and G are cumulatives, they are right-continuous so $F(a) = G(a)$ for all $a < 1$. Since $F(1) = 1 = G(1)$, $F = G$. □

Lemma (B3)**.** *Let ρ be monotonic and linear. Then $\left(F \cup f^b\right)_\rho = F_\rho \vee f^b_\rho$ for all $b \in [0,1]$.*

Proof. Let ρ be monotonic and linear. Note that if $\rho\left(\underline{f}, \overline{f}\right) > 0$, then $\underline{f}$ and $\overline{f}$ are tied so by Lemma A3, $\rho\left(\underline{f}, f\right) = \rho\left(f, \underline{f}\right) = 1$ for all $f \in H$. Thus, all acts are tied, so $\left(F \cup f^b\right)_\rho = 1 = F_\rho \vee f^b_\rho$ trivially.

Assume $\rho\left(\underline{f}, \overline{f}\right) = 0$, so linearity implies $\rho\left(f^b, f^a\right) = 1$ for $a \geq b$ and $\rho\left(f^b, f^a\right) = 0$ otherwise. Hence $f^b_\rho = \mathbf{1}_{[b,1]}$, so for any $F \in \mathcal{K}$,

$$\left(F_\rho \vee f^b_\rho\right)(a) = \left(F_\rho \vee \mathbf{1}_{[b,1]}\right)(a) = \begin{cases} 1 & \text{if } a \geq b \\ F_\rho(a) & \text{otherwise} \end{cases}$$

Let $G := F \cup f^b \cup f^a$ so

$$\left(F \cup f^b\right)_\rho(a) = \rho_G\left(F \cup f^b\right)$$

First, suppose $a \geq b$. If $a > b$, then $\rho\left(f^a, f^b\right) = 0$ so $\rho_G(f^a) = 0$ by monotonicity. Hence, $\rho_G\left(F \cup f^b\right) = 1$. If $a = b$, then $\rho_G\left(F \cup f^b\right) = 1$ trivially. Thus, $\left(F \cup f^b\right)_\rho(a) = 1$ for all $a \geq b$. Now consider $a < b$ so $\rho\left(f^b, f^a\right) = 0$ which implies $\rho_G\left(f^b\right) = 0$ by monotonicity. First, suppose f^a is tied with nothing in F. Thus, by Lemma A2, $f^a_G = f^a_{F \cup f^a} = f^a$ so

$$\rho_{F \cup f^a}(F) + \rho_{F \cup f^a}(f^a) = 1 = \rho_G(F) + \rho_G(f^a)$$

By monotonicity, $\rho_{F \cup f^a}(F) \geq \rho_G(F)$ and $\rho_{F \cup f^a}(f^a) \geq \rho_G(f^a)$ so $\rho_G(F) = \rho_{F \cup f^a}(F)$. Hence,

$$\rho_G\left(F \cup f^b\right) = \rho_G(F) = \rho_{F \cup f^a}(F) = F_\rho(a)$$

Finally, suppose f^a is tied with some $f' \in F$. Thus, by Lemma A3,

$$\rho_G\left(F \cup f^b\right) = \rho_{F \cup f^b}\left(F \cup f^b\right) = 1 = F_\rho(a)$$

so $\left(F \cup f^b\right)_\rho(a) = F_\rho(a)$ for all $a < b$. Thus, $\left(F \cup f^b\right)_\rho = F_\rho \vee f^b_\rho$. □

Definition. u is *normalized* iff $u\left(\underline{f}\right) = 0$ and $u\left(\overline{f}\right) = 1$.

Lemma (B4). *Let ρ be monotonic and linear. Suppose $\succeq_\rho$ and τ are represented by (μ, u). Then $F_\rho = F_\tau$ for all $F \in \mathcal{K}$.*

Proof. Let ρ be monotonic and linear, and suppose $\succeq_\rho$ and τ are represented by (μ, u). By Theorem A14, we can assume u is normalized without loss of generality. Let

$$V(F) := \int_{\Delta S} \sup_{f \in F} q \cdot (u \circ F) \mu(dq)$$

so V represents $\succeq_\rho$. Since test functions are well-defined under ρ, let $\overline{f}$ and $\underline{f}$ be the best and worst acts respectively. We first show that $\rho\left(\underline{f}, \overline{f}\right) = 0$. Suppose otherwise so $\underline{f}$ and $\overline{f}$ must be tied. By Lemma A4, f^b and f^a are tied for all $\{a, b\} \subset [0,1]$. Thus, $f^b(a) = 1$ for all $\{a, b\} \subset [0,1]$. Hence $V_\rho\left(f^b\right) = V_\rho(f^a)$ so $V\left(f^b\right) = V(f^a)$ for all $\{a, b\} \subset [0,1]$. This implies

$$u\left(\underline{f}\right) = V\left(f^1\right) = V\left(f^b\right) = u\left(f^b\right)$$

for all $b \in [0,1]$ contradicting the fact that u is non-constant. Thus, $\rho\left(\underline{f}, \overline{f}\right) = 0$ so

$$\int_{[0,1]} \underline{f}_\rho(a)\, da = 0 \leq \int_{[0,1]} f_\rho(a)\, da \leq 1 = \int_{[0,1]} \overline{f}_\rho(a)\, da$$

which implies $\underline{f} \preceq_\rho f \preceq_\rho \overline{f}$. Thus, $V\left(\underline{f}\right) \leq V(f) \leq V\left(\overline{f}\right)$ for all $f \in H$ so $u\left(\underline{f}\right) \leq u(f) \leq u\left(\overline{f}\right)$ for all $f \in H_c$ and $\left\{\underline{f}, \overline{f}\right\} \subset H_c$. Hence, we can let $\underline{f}$ and $\overline{f}$ be the worst and best acts of τ.

Since $\succeq_\rho$ is represented by V, we have $V_\rho(F) = \phi(V(F))$ for some monotonic transformation $\phi : \mathbb{R} \to \mathbb{R}$. Now, for $b \in [0,1]$,

$$1 - b = \int_{[0,1]} f^b_\rho(a)\, da = V_\rho\left(f^b\right) = \phi\left(V\left(f^b\right)\right) = \phi(1 - b)$$

so $\phi(a) = a$ for all $a \in [0,1]$. Now, by Lemmas A12 and B1,

$$\begin{aligned}\int_{[0,1]} F_\rho(a)\, da &= V_\rho(F) = V(F) \\ &= 1 - \int_{[0,1]} a\, dF_\tau(a) = \int_{[0,1]} F_\tau(a)\, da\end{aligned}$$

for all $F \in \mathcal{K}$.

By Lemma B3, for all $b \in [0,1]$,

$$\int_{[0,1]} \left(F \cup f^b\right)_\rho (a)\, da = \int_{[0,1]} \left(F_\rho \vee f^b_\rho\right)(a)\, da = \int_{[0,1]} \left(F_\rho \vee \mathbf{1}_{[b,1]}\right)(a)\, da$$
$$= \int_{[0,b]} F_\rho (a)\, da + 1 - b$$

Thus, for all $b \in [0,1]$,

$$G(b) := \int_{[0,b]} F_\rho (a)\, da = \int_{[0,b]} F_\tau (a)\, da$$

Let λ be the measure corresponding to G so $\lambda[0,b] = G(b)$. Thus, by the Radon-Nikodym Theorem (see Theorem I.5.11 of Çinlar [12]), we have a.e.

$$F_\rho (a) = \frac{d\lambda}{da} = F_\tau (a)$$

Lemma B2 then establishes that $F_\rho = F_\tau$ for all $F \in \mathcal{K}$. □

Lemma (B5)**.** *Let ρ be monotonic, linear and continuous. Suppose τ is represented by (μ, u). Then $F_\rho = F_\tau$ for all $F \in \mathcal{K}$ iff $\rho = \tau$.*

Proof. Note that necessity is trivial so we prove sufficiency. Assume u is normalized without loss of generality. Suppose $F_\rho = F_\tau$ for all $F \in \mathcal{K}$. Let $\left\{\underline{f}, \overline{f}, \underline{g}, \overline{g}\right\} \subset H_c$ be such that for all $f \in H$,

$$\rho\left(\overline{f}, f\right) = \rho\left(f, \underline{f}\right) = \tau\left(\overline{g}, f\right) = \tau\left(f, \underline{g}\right) = 1$$

Note that

$$\tau\left(\overline{f}, \overline{g}\right) = \overline{f}_\tau (0) = \overline{f}_\rho (0) = 1$$

so $\overline{f}$ and $\overline{g}$ are τ-tied. Thus, by Lemma A3, we can assume $\overline{f} = \overline{g}$ without loss of generality. Now, suppose $u\left(\underline{f}\right) > u\left(\underline{g}\right)$ so we can find some $f \in H_c$ such that $u\left(\underline{f}\right) > u(f)$ and $\underline{f} = \overline{f} b f$ for some $b \in (0,1)$. Now,

$$1 = \tau\left(f, \underline{g}\right) = f_\tau (1) = f_\rho (1) = \rho\left(f, \underline{f}\right)$$

violating linearity. Thus, $u\left(\underline{f}\right) = u\left(\underline{g}\right)$, so $\underline{f}$ and $\underline{g}$ are also τ-tied and we assume $\underline{f} = \underline{g}$ without loss of generality.

Suppose $f \in H$ and f^b are τ-tied for some $b \in [0,1]$. We show that f^b and f are also ρ-tied. Note that

$$\mathbf{1}_{[b,1]} (a) = f_\tau (a) = f_\rho (a) = \rho\left(f, f^a\right)$$

Suppose f^b is not ρ-tied with g. Thus, $\rho\left(f^b, f\right) = 0$. Now, for $a < b$, $\rho\left(f, f^a\right) = 0$ implying $\rho\left(f^a, f\right) = 1$. This violates the continuity of ρ. Thus, f^b is ρ-tied with f.

Consider any $\{f, g\} \subset H$ such that f and g are τ-tied As both ρ and τ are linear, we can assume $g \in H_c$ without loss of generality by Lemma A4. Let f^b be τ-tied with g, so it is also τ-tied with f. From above, we have f^b is ρ-tied with both f and g, so both f and g are ρ-tied by Lemma A2.

Now, suppose f and g are ρ-tied and we assume $g \in H_c$ again without loss of generality. Let f^b be τ-tied with g. From above, f^b is ρ-tied with g are thus also with f. Hence

$$\tau\left(f, g\right) = \tau\left(f, f^b\right) = f_\tau\left(b\right) = f_\rho\left(b\right) = 1$$

Now, let $h \in H$ be such that $g = fah$ for some $a \in (0, 1)$. By linearity, we have h is ρ-tied with g and thus also with f^b. Hence

$$\tau\left(h, g\right) = \tau\left(h, f^b\right) = h_\tau\left(b\right) = h_\rho\left(b\right) = 1$$

By linearity, f and g are τ-tied. Hence, f and g are ρ-tied iff they are τ-tied, so ties agree on both ρ and τ and $\mathcal{H}_F^\rho = \mathcal{H}_F^\tau$ for all $F \in \mathcal{K}$.

Now, consider $f \in G$. Note that by linearity and Lemma A3, we can assume $f = f^a$ for some $a \in [0, 1]$ without loss of generality. First, suppose f^a is tied with nothing in $F := G \backslash f^a$. Thus,

$$\rho_G\left(f\right) = 1 - \rho_G\left(F\right) = 1 - F_\rho\left(a\right) = 1 - F_\tau\left(a\right) = \tau_G\left(f\right)$$

Now, if f^a is tied with some act in G, then let $F' := F \backslash f_G^a$. By Lemma A3, $\rho_G\left(f\right) = \rho\left(f, F'\right)$ and $\tau_G\left(f\right) = \tau\left(f, F'\right)$ where f is tied with nothing in F'. Applying the above on F' yields $\rho_G\left(f\right) = \tau_G\left(f\right)$ for all $f \in G \in \mathcal{K}$. Hence, $\rho = \tau$. $\square$

Theorem (B6)**.** *Let ρ be monotonic, linear and continuous. Then the following are equivalent:*

(1) *ρ is represented by (μ, u)*

(2) *$\succeq_\rho$ is represented by (μ, u)*

Proof. First suppose (1) is true and assume u is normalized without loss of generality. Let

$$V\left(F\right) := \int_{\Delta S} \sup_{f \in F} q \cdot \left(u \circ F\right) \mu\left(dq\right)$$

so from Lemmas A12 and B1,

$$V_\rho(F) = 1 - \int_{[0,1]} a \, dF_\rho(a) = 1 - (1 - V(F)) = V(F)$$

so (2) is true. Now, suppose (2) is true and let τ be represented by (μ, u) with u normalized. By Lemma B4, $F_\rho = F_\tau$ for all $F \in \mathcal{K}$. By Lemma B5, $\rho = \tau$ so (1) is true. $\square$

Lemma (B7)**.** *Let $\succeq$ be dominant and $\rho = \rho_\succeq$. Then for all $F \in \mathcal{K}$*

(1) $\overline{f} \succeq F \succeq \underline{f}$

(2) $F \cup \overline{f} \sim \overline{f}$ *and* $F \cup \underline{f} \sim F$

Proof. Let $\succeq$ be dominant and $\rho = \rho_\succeq$. We prove the lemma in order:

(1) Since $\rho = \rho_\succeq$, let $V : \mathcal{K} \to [0,1]$ represent $\succeq$ and $\rho(f_a, F) = \frac{dV(F \cup f_a)}{da}$ for $f_a := a\overline{f} + (1-a)\underline{f}$. Thus,

$$V(F \cup f_1) - V(F \cup f_0) = \int_{[0,1]} \frac{dV(F \cup f_a)}{da} da = \int_{[0,1]} \rho(f_a, F) \, da$$

Now, for $F = \underline{f}$,

$$V\left(\underline{f} \cup \overline{f}\right) - V\left(\underline{f}\right) = \int_{[0,1]} \rho\left(f_a, \underline{f}\right) da = 1$$

Thus, $V\left(\underline{f}\right) = 0$ and $V\left(\underline{f} \cup \overline{f}\right) = 1$. Since $\overline{f} \succeq \underline{f}$, by dominance,

$$V\left(\overline{f}\right) = V\left(\underline{f} \cup \overline{f}\right) = 1$$

so $V\left(\overline{f}\right) = 1 \geq V(F) \geq 0 = V\left(\underline{f}\right)$ for all $F \in \mathcal{K}$.

(2) From (1), $\overline{f} \succeq f \succeq \underline{f}$ for all $f \in H$. Let $F = \{f_1, \dots, f_k\}$. By iteration,

$$\overline{f} \sim \overline{f} \cup f_1 \sim \overline{f} \cup f_1 \cup f_2 \sim \overline{f} \cup F$$

Now, for any $f \in F$, $f_s \succeq \underline{f}$ for all $s \in S$ so $F \sim F \cup \underline{f}$.

$\square$

Lemma (B8)**.** *Let ρ be monotone, linear and $\rho\left(\underline{f}, \overline{f}\right) = 0$. Then a.e.*

$$\rho(f_a, F) = 1 - F_\rho(1-a) = \frac{dV_\rho(F \cup f_a)}{da}$$

Proof. Let ρ be monotone, linear and $\rho\left(\underline{f},\overline{f}\right)=0$ and let $f^{b}:=f_{1-b}$. We first show that a.e.

$$1=\rho\left(f^{b},F\right)+F_{\rho}\left(b\right)=\rho\left(f^{b},F\right)+\rho\left(F,f^{b}\right)$$

By Lemma A2, this is violated iff $\rho\left(f^{b},F\right)>0$ and there is some act in $f\in F$ tied with f^{b}. Note that if f is tied with f^{b}, then f cannot be tied with f^{a} for some $a\neq b$ as $\rho\left(\underline{f},\overline{f}\right)=0$. Thus, $\rho\left(f^{b},F\right)+F_{\rho}\left(b\right)\neq 1$ at most a finite number of points as F is finite. The result follows.

Now, by Lemma B3,

$$\begin{aligned} V_{\rho}\left(F\cup f_{b}\right)=V_{\rho}\left(F\cup f^{1-b}\right) &= \int_{[0,1]}\left(F_{\rho}\left(a\right)\vee\left(f^{1-b}\right)_{\rho}\left(a\right)\right)da \\ &= \int_{[0,1-b]}F_{\rho}\left(a\right)da+b=\int_{[b,1]}F_{\rho}\left(1-a\right)da+b \end{aligned}$$

Since $V_{\rho}\left(F\cup f_{0}\right)=\int_{[0,1]}F_{\rho}\left(1-a\right)da$, we have

$$\begin{aligned} V_{\rho}\left(F\cup f_{b}\right)-V_{\rho}\left(F\cup f_{0}\right) &= b-\int_{[0,b]}F_{\rho}\left(1-a\right)da \\ &= \int_{[0,b]}\left(1-F_{\rho}\left(1-a\right)\right)da \end{aligned}$$

Thus, we have a.e.

$$\frac{dV_{\rho}\left(F\cup f_{a}\right)}{da}=1-F_{\rho}\left(1-a\right)=\rho\left(f_{a},F\right)$$

□

Theorem (B9)**.** *Let $\succeq$ be dominant. Then the following are equivalent:*

(1) *$\succeq$ is represented by (μ,u)*

(2) *$\rho_{\succeq}$ is represented by (μ,u)*

Proof. Assume u is normalized without loss of generality and let

$$V\left(F\right):=\int_{\Delta S}\sup_{f\in F}q\cdot\left(u\circ f\right)\mu\left(dq\right)$$

First, suppose (1) is true and let $\rho=\rho_{\succeq}$ where $W:\mathcal{K}\to[0,1]$ represents $\succeq$ and $\rho\left(f_{a},F\right)=\frac{dW\left(F\cup f_{a}\right)}{da}$ for $f_{a}:=a\overline{f}+\left(1-a\right)\underline{f}$. Since V also represents $\succeq$, $W=\phi\circ V$ for some monotonic $\phi:\mathbb{R}\to\mathbb{R}$. By Lemma B7, $\overline{f}\succeq F\succeq\underline{f}$ so $u\left(\overline{f}\right)\geq u\left(f\right)\geq u\left(\underline{f}\right)$ for all $f\in H$. Let τ be represented by (μ,u) so $\underline{f}$ and $\overline{f}$ are the worst and best acts of τ as well.

By Lemmas A12 and B1,

$$V_\tau(F) = 1 - \int_{[0,1]} a \; dF_\tau(a) = 1 - (1 - V(F)) = V(F)$$

so by Lemma B8, $\tau(f_a, F) = \frac{dV(F \cup f_a)}{da}$.

Suppose $\rho\left(\underline{f}, \overline{f}\right) > 0$ so $\underline{f}$ and $\overline{f}$ are ρ-tied. Thus, by Lemma A2, $\rho\left(\underline{f}, f\right) = \rho\left(f, \underline{f}\right) = 1$ so all acts are tied under ρ. Thus,

$$W(f_1) - W(f_1 \cup f_0) = \int_{[0,1]} \rho(f_a, f_1)\, da = 1$$

so $\overline{f} \succ \overline{f} \cup \underline{f} \sim \overline{f}$ by Lemma B7 a contradiction. Thus, $\rho\left(\underline{f}, \overline{f}\right) = 0$.

Now,

$$W\left(\underline{f} \cup \overline{f}\right) - W\left(\underline{f}\right) = \int_{[0,1]} \rho\left(f_a, \underline{f}\right) da = 1$$

so $W\left(\underline{f}\right) = 0$ and $W\left(\overline{f}\right) = 1$ by dominance. By dominance, for $b \geq 0$,

$$W(f_b) = W(f_0 \cup f_b) - W(f_0 \cup f_0) = \int_{[0,b]} \rho(f_a, f_0)\, da = b$$

By the same argument, $V(f_b) = b$ so

$$b = W(f_b) = \phi(V(f_b)) = \phi(b)$$

so $W = V$. By Lemma B8, we have a.e.

$$1 - F_\tau(1-a) = \tau(f_a, F) = \frac{dW(F \cup f_a)}{da} = \frac{dV(F \cup f_a)}{da} = 1 - F_\rho(1-a)$$

so $F_\tau = F_\rho$ a.e.. By Lemma B2, $F_\tau = F_\rho$ so by Lemma B5, $\rho_{\succeq} = \rho = \tau$ and (2) holds.

Now, suppose (2) is true and let $\rho = \rho_{\succeq}$ where $W : \mathcal{K} \to [0,1]$ represents $\succeq$ and $\rho(f_a, F) = \frac{dW(F \cup f_a)}{da}$ for $f_a := a\overline{f} + (1-a)\underline{f}$. Suppose ρ is represented by (μ, u) and since $V_\rho = V$, we have $\rho(f_a, F) = \frac{dV(F \cup f_a)}{da}$ by Lemma B8. Now, by dominance,

$$\begin{aligned}1 - W(F) = W(F \cup f_1) - W(F \cup f_0) &= \int_{[0,1]} \rho(f_c, F)\, da \\ &= V(F \cup f_1) - V(F \cup f_0) = 1 - V(F)\end{aligned}$$

so $W = V$ proving (1). □

Appendix C

C1. Assessing Informativeness

In this section of Appendix C, we prove our result on assessing informativeness.

Theorem (C1)**.** *Let ρ and τ be represented by (μ, u) and (ν, u) respectively. Then the following are equivalent:*

(1) *μ is more informative than ν*

(2) *$F_\tau \geq_{SOSD} F_\rho$ for all $F \in \mathcal{K}$*

(3) *$F_\tau \geq_m F_\rho$ for all $F \in \mathcal{K}$*

Proof. Let ρ and τ be represented by (μ, u) and (ν, u) respectively and we assume u is normalized without loss of generality. We show that (1) implies (2) implies (3) implies (1). First, suppose μ is more informative than ν. Fix $F \in \mathcal{K}$ and let $U := u \circ F$ and $h(U, q)$ denote the support function of U at $q \in \Delta S$. Let $\psi_F(q) := 1 - h(U, q)$, and since support functions are convex, ψ_F is concave in $q \in \Delta S$.[34] Let $\phi : \mathbb{R} \to \mathbb{R}$ be increasing concave, and note that by Lemma A12,

$$\int_{[0,1]} \phi dF_\rho = \int_{\Delta S} \phi \circ \psi_F(q) \, \mu(dq)$$

Now for $\alpha \in [0, 1]$, $\psi_F(q\alpha r) \geq \alpha \psi_F(q) + (1 - \alpha) \psi_F(r)$ so

$$\begin{aligned} \phi(\psi_F(q\alpha r)) &\geq \phi(\alpha \psi_F(q) + (1 - \alpha) \psi_F(r)) \\ &\geq \alpha \phi(\psi_F(q)) + (1 - \alpha) \phi(\psi_F(r)) \end{aligned}$$

[34] See Theorem 1.7.5 of Schneider [44] for elementary properties of support functions.

so $\phi \circ \psi_F$ is concave. By Jensen's inequality,

$$\begin{aligned}\int_{\Delta S} \phi \circ \psi_F (q) \mu (dq) &= \int_{\Delta S} \int_{\Delta S} \phi \circ \psi_F (p) K (q, dp) \nu (dq) \\ &\leq \int_{\Delta S} \phi \circ \psi_F \left(\int_{\Delta S} p \ K (q, dp) \right) \nu (dq) \\ &\leq \int_{\Delta S} \phi \circ \psi_F (q) \nu (dq)\end{aligned}$$

so $\int_{[0,1]} \phi dF_\rho \leq \int_{[0,1]} \phi dF_\tau$ and $F_\tau \geq_{SOSD} F_\rho$ for all $F \in \mathcal{K}$.

Since $\geq_{SOSD}$ implies $\geq_m$, (2) implies (3) is trivially. Now, suppose $F_\tau \geq_m F_\rho$ for all $F \in \mathcal{K}$. Thus, if we let $\phi(x) = x$, then

$$\begin{aligned}\int_{\Delta S} \psi_F (q) \mu (dq) &= \int_{[0,1]} a \ dF_\rho (a) \\ &\leq \int_{[0,1]} a \ dF_\tau (a) = \int_{\Delta S} \psi_F (q) \nu (dq)\end{aligned}$$

Thus,

$$\int_{\Delta S} h (u \circ F, q) \mu (dq) \geq \int_{\Delta S} h (u \circ F, q) \nu (dq)$$

for all $F \in \mathcal{K}$. Hence, by Blackwell [5, 6], μ is more informative than ν . □

Lemma (C2)**.** *Let ρ and τ be represented by (μ, u) and (ν, v) respectively. Then $f_\rho =_m f_\tau$ for all $f \in H$ iff μ and ν share average beliefs and $u = \alpha v + \beta$ for $\alpha > 0$.*

Proof. Let ρ and τ be represented by (u, μ) and (v, ν) respectively. We assume u is normalized without loss of generality. Let $\psi_f (q) := 1 - q \cdot (u \circ f)$ so by Lemma A12,

$$\int_{[0,1]} a \ df_\rho (a) = \int_{\Delta S} \psi_f (q) \mu (dq)$$

First, suppose μ and ν share average beliefs and $u = v$ without loss of generality. Thus,

$$\begin{aligned}\int_{\Delta S} \psi_f (q) \mu (dq) &= \psi_f \left(\int_{\Delta S} q \ \mu (dq) \right) \\ &= \psi_f \left(\int_{\Delta S} q \ \nu (dq) \right) = \int_{\Delta S} \psi_f (q) \nu (dq)\end{aligned}$$

so $f_\rho =_m f_\tau$ for all $f \in H$. Now assume $f_\rho =_m f_\tau$ for all $f \in H$ so by Lemma A13, $u = \alpha v + \beta$

for $\alpha > 0$. We assume $u = v$ without loss of generality so

$$\psi_f\left(\int_{\Delta S} q \ \mu(dq)\right) = \int_{\Delta S} \psi_f(q)\,\mu(dq)$$
$$= \int_{\Delta S} \psi_f(q)\,\nu(dq) = \psi_f\left(\int_{\Delta S} q \ \nu(dq)\right)$$

If we let $r_\mu = \int_{\Delta S} q \ \mu(dq)$ and $r_\nu = \int_{\Delta S} q \ \nu(dq)$, then

$$1 - r_\mu \cdot (u \circ f) = 1 - r_\nu \cdot (u \circ f)$$
$$0 = (r_\mu - r_\nu) \cdot (u \circ f)$$

for all $f \in H$. Thus, $w \cdot (r_\mu - r_\nu) = 0$ for all $w \in [0,1]^S$ implying $r_\mu = r_\nu$. Thus, μ and ν share average beliefs. □

Lemma C3 below shows that our definition of "more preference for flexibility than" coincides with that of DLST.

Lemma (C3)**.** *Let $\succeq_1$ and $\succeq_2$ have subjective learning representations. Then $\succeq_1$ has more preference for flexibility than $\succeq_2$ iff $F \succ_2 f$ implies $F \succ_1 f$.*

Proof. Let $\succeq_1$ and $\succeq_2$ be represented by V_1 and V_2 respectively. Suppose $g \succeq_2 f$ implies $g \succeq_1 f$. Let $\underline{f}$ and $\overline{f}$ be the worst and best acts under V_2 and assume $V_2(\underline{f}) = 0$ and $V_2(\overline{f}) = 1$ without loss of generality. Now, $g \sim_2 f$ implies $g \sim_1 f$. If we let $g \sim_2 \overline{f} a \underline{f}$ for some $a \in [0,1]$, then $V_2(g) = a$ and

$$V_1(g) = aV_1(\overline{f}) + (1-a)V_1(\underline{f}) = \left(V_1(\overline{f}) - V_1(\underline{f})\right)V_2(g) + V_1(\underline{f})$$

for all $g \in H$. Thus, $\succeq_1$ and $\succeq_2$ coincide on singletons. Note that the case for $g \succeq_1 f$ implies $g \succeq_2 f$ is symmetric.

First, suppose $\succeq_1$ has more preference for flexibility than $\succeq_2$. Let $F \succ_2 f$ and $F \sim_2 g$ for some $g \in H$. Thus, $F \succeq_1 g$ and since $g \succ_2 f$, $g \succ_1 f$. Hence, $F \succ_1 f$. For the converse, suppose $F \succ_2 f$ implies $F \succ_1 f$. Let $F \succeq_2 f$ and note that if $F \succ_2 f$, then the result follows so assume $F \sim_2 f$. Let $g \sim_1 F$ for some $g \in H$ so $g \succeq_2 F$. Thus, $g \succeq_2 f$ so $g \succeq_1 f$ which implies $F \succeq_1 f$ so $\succeq_1$ has more preference for flexibility than $\succeq_2$. □

C2. Partitional Information Representations

In this section of Appendix C, we consider partitional information representations. Given an algebra $\mathcal{F}$ on S, let $Q_{\mathcal{F}}(S) := \bigcup_{s \in S} Q_{\mathcal{F}}(s)$. For each $q \in Q_{\mathcal{F}}(S)$, let

$$E_q^{\mathcal{F}} := \{ s \in S | \, Q_{\mathcal{F}}(s) = q \}$$

and let $\mathcal{P}_{\mathcal{F}} := \left\{ E_q^{\mathcal{F}} \right\}_{q \in Q_{\mathcal{F}}(S)}$ be a partition on S. Also define

$$C_{\mathcal{F}} := \operatorname{conv}(Q_{\mathcal{F}}(S))$$

Lemma (C4)**.** *$\sigma(\mathcal{P}_{\mathcal{F}}) = \mathcal{F}$.*

Proof. Let $q \in Q_{\mathcal{F}}(S)$ and note that since $Q_{\mathcal{F}}(\cdot, \{s'\})$ is $\mathcal{F}$-measurable for all $s' \in S$,

$$E_q^{\mathcal{F}} = \bigcap_{s' \in S} \{ s \in S | \, Q_{\mathcal{F}}(s, \{s'\}) = q_{s'} \} \in \mathcal{F}$$

Thus, $E_q^{\mathcal{F}} \in \mathcal{F}$ for all $q \in Q_{\mathcal{F}}(S)$ so $\mathcal{P}_{\mathcal{F}} \subset \mathcal{F}$. Now, let $A \in \mathcal{F}$ and note that

$$\mathbf{1}_A(s) = \mathbb{E}_{\mathcal{F}}[\mathbf{1}_A] = Q_{\mathcal{F}}(s, A)$$

so $Q_{\mathcal{F}}(s, A) = 1$ for $s \in A$ and $Q_{\mathcal{F}}(s, A) = 0$ for $s \notin A$. Since $\mathcal{P}_{\mathcal{F}}$ is a partition of S, let $\mathcal{P}_A \subset \mathcal{P}_{\mathcal{F}}$ be such that

$$A \subset \overline{E} := \bigcup_{E \in \mathcal{P}_A} E$$

Suppose $\exists s \in E \backslash A$ for some $E \in \mathcal{P}_A$. Thus, we can find an $s' \in A \cap E$ so $Q_{\mathcal{F}}(s) = Q_{\mathcal{F}}(s')$. However, $Q_{\mathcal{F}}(s', A) = 1 > 0 = Q_{\mathcal{F}}(s, A)$ a contradiction. Thus, $A = \overline{E} \in \sigma(\mathcal{P}_{\mathcal{F}})$ so $\mathcal{P}_{\mathcal{F}} \subset \mathcal{F} \subset \sigma(\mathcal{P}_{\mathcal{F}})$. This proves that $\sigma(\mathcal{P}_{\mathcal{F}}) = \mathcal{F}$ (see Exercise I.1.10 of Çinlar [12]). $\square$

Lemma (C5)**.** *$E_q^{\mathcal{F}} = \{ s \in S | \, q_s > 0 \}$ for $q \in Q_{\mathcal{F}}(S)$.*

Proof. Let $q = Q_{\mathcal{F}}(s)$ for some $s \in S$. Since $E_q^{\mathcal{F}} \in \mathcal{F}$, $Q_{\mathcal{F}}\left(s', E_q^{\mathcal{F}}\right) = \mathbf{1}_{E_q^{\mathcal{F}}}(s')$ for all $s' \in S$. Note that since $s \in E_q^{\mathcal{F}}$, $q\left(E_q^{\mathcal{F}}\right) = 1$ for all $q \in Q_{\mathcal{F}}(S)$. Thus, $q_s > 0$ implies $s \in E_q^{\mathcal{F}}$. Suppose $s \in E_q^{\mathcal{F}}$ but $q_s = 0$. Now,

$$\begin{aligned} r_s &= \mathbb{E}\left[\mathbb{E}_{\mathcal{F}}\left[\mathbf{1}_{\{s\}}\right]\right] = \mathbb{E}\left[Q_{\mathcal{F}}(s', \{s\})\right] \\ &= \sum_{q' \in Q_{\mathcal{F}}(S)} r\left(E_{q'}^{\mathcal{F}}\right) q'_s = r\left(E_q^{\mathcal{F}}\right) q_s = 0 \end{aligned}$$

contradicting the fact that r has full support. Thus, $E_q^{\mathcal{F}} = \{ s \in S | \, q_s > 0 \}$. $\square$

Lemma (C6)**.** *ext* $(C_{\mathcal{F}}) \subset Q_{\mathcal{F}}(S)$.

Proof. Suppose $q \in \bar{Q}_{\mathcal{F}} := \text{ext}(C_{\mathcal{F}}) \subset C_{\mathcal{F}}$ but $q \notin Q_{\mathcal{F}}(S)$. If $q \in \text{conv}(Q_{\mathcal{F}}(S))$, then $q = \sum_i \alpha_i p_i$ for $\alpha_i \in (0,1)$, $\sum_i \alpha_i = 1$ and $p_i \in Q_{\mathcal{F}}(S) \subset C_{\mathcal{F}}$. However, this contradicts the fact that $q \in \text{ext}(C_{\mathcal{F}})$, so $q \notin \text{conv}(Q_{\mathcal{F}}(S)) = C_{\mathcal{F}}$ another contradiction. Thus, $\bar{Q}_{\mathcal{F}} \subset Q_{\mathcal{F}}(S)$. □

Proposition (C7)**.** *Let ρ and τ be represented by $(\mathcal{F}, u)$ and $(\mathcal{G}, u)$ respectively. Then the following are equivalent:*

(1) $\mathcal{D}_\tau \subset \mathcal{D}_\rho$

(2) $C_{\mathcal{F}} \subset C_{\mathcal{G}}$

(3) $\mathcal{F} \subset \mathcal{G}$

Proof. Let ρ and τ be represented by $(\mathcal{F}, u)$ and $(\mathcal{G}, u)$ respectively. Assume u is normalized without loss of generality. We show that (1) implies (2) implies (3) implies (1).

First, suppose (1) is true but $C_{\mathcal{F}} \not\subset C_{\mathcal{G}}$ and let $p \in C_{\mathcal{F}} \backslash C_{\mathcal{G}}$. Note that $C_{\mathcal{G}}$ is compact (see Theorem 1.1.10 of Schneider [44]). Thus, by a separating hyperplane argument (Theorem 1.3.4 of Schneider [44]), there is a $a \in \mathbb{R}$, $\varepsilon > 0$ and $v \in \mathbb{R}^S$ such that for all $q \in C_{\mathcal{G}}$,

$$q \cdot v \geq a + \varepsilon > a - \varepsilon \geq p \cdot v$$

Note that since $C_{\mathcal{G}} \subset \Delta S$ and $p \in \Delta S$, we can assume $v \in [0,1]^S$ without loss of generality. Let $f \in H$ be such that $u \circ f = v$. Note that $(a - \varepsilon, a + \varepsilon) \subset [0,1]$, and since both $Q_{\mathcal{F}}(S)$ and $Q_{\mathcal{G}}(S)$ are finite, we can find

$$b \in (a - \varepsilon, a + \varepsilon) \setminus \bigcup_{q \in Q_{\mathcal{F}}(S) \cup Q_{\mathcal{G}}(S)} q \cdot v$$

Thus, $b \neq q \cdot v$ for all $q \in Q_{\mathcal{F}}(S) \cup Q_{\mathcal{G}}(S)$. Let $h \in H_c$ such that $u(h) = b$ so

$$u(h) \neq q \cdot (u \circ f)$$

for all $q \in Q_{\mathcal{F}}(S) \cup Q_{\mathcal{G}}(S)$. Thus, $\{f, h\}$ is generic under both $\mathcal{F}$ and $\mathcal{G}$. Since $q \cdot (u \circ f) > b = u(h)$ for all $q \in C_{\mathcal{G}}$, we have

$$\tau(f, h) = \mu_{\mathcal{G}} \{q \in \Delta S | \, q \cdot (u \circ f) \geq u(h)\} = 1$$

so $\{f,h\} \in \mathcal{D}_\tau$. However, $u(h) > p \cdot (u \circ f)$ for some $p \in C_{\mathcal{F}}$. If $q \cdot (u \circ f) \geq u(h)$ for all $q \in Q_{\mathcal{F}}(S)$, then $q \cdot (u \circ f) \geq u(h)$ for all $q \in C_{\mathcal{F}}$ a contradiction. Thus, $\exists q \in Q_{\mathcal{F}}(S)$ such that $u(h) > q \cdot (u \circ f)$. On the other hand, if $u(h) > q \cdot (u \circ f)$ for all $q \in Q_{\mathcal{F}}(S)$, then

$$Q_{\mathcal{G}}(s) \cdot (u \circ f) > b > Q_{\mathcal{F}}(s) \cdot (u \circ f)$$

for all $s \in S$. Thus, $\mathbb{E}_{\mathcal{G}}[u \circ f] - \mathbb{E}_{\mathcal{F}}[u \circ f] > \varepsilon'$ for some $\varepsilon' > 0$. Taking expectations yield

$$\mathbb{E}[\mathbb{E}_{\mathcal{G}}[u \circ f] - \mathbb{E}_{\mathcal{F}}[u \circ f]] = \mathbb{E}[u \circ f] - \mathbb{E}[u \circ f] = 0$$

a contradiction. Thus, $\exists \{s, s'\} \subset S$ such that $u(h) > Q_{\mathcal{F}}(s) \cdot (u \circ f)$ and $Q_{\mathcal{F}}(s') \cdot (u \circ f) \geq u(h)$. Since we assume r has full support,

$$\rho(h,f) = \mu_{\mathcal{F}}\{q \in \Delta S | \, u(h) \geq q \cdot (u \circ f)\} \geq r_s > 0$$
$$\rho(f,h) = \mu_{\mathcal{F}}\{q \in \Delta S | \, q \cdot (u \circ f) \geq u(h)\} \geq r_{s'} > 0$$

so $\{f,h\} \notin \mathcal{D}_\rho$ contradicting (1). Thus, (1) implies (2).

Now, assume (2) is true. Let $q \in Q_{\mathcal{F}}(S)$ so $q \in C_{\mathcal{F}} \subset C_{\mathcal{G}}$. Since $\mathrm{ext}(C_{\mathcal{G}}) \subset Q_{\mathcal{G}}(S)$ from Lemma C6, Minkowski's Theorem (Corollary 1.4.5 of Schneider [44]) yields that $q = \sum_i \alpha_i p^i$ for $\alpha_i > 0$, $\sum_i \alpha_i = 1$ and $p^i = Q_{\mathcal{G}}(s_i)$. Note that by Lemma C5, $\sum_{s \in E_q^{\mathcal{F}}} q_s = 1$. If $p_s^i > 0$ for some $s \notin E_q^{\mathcal{F}}$, then $q_s > 0$ a contradiction. Thus, $\sum_{s \in E_q^{\mathcal{F}}} p_s^i = 1$ for all p^i. Now, by Lemma C5 again, for each p^i,

$$E_{p^i}^{\mathcal{G}} = \left\{s \in S | \, p_s^i > 0\right\} \subset E_q^{\mathcal{F}}$$

so $\bigcup_i E_{p^i}^{\mathcal{G}} \subset E_q^{\mathcal{F}}$. Moreover, if $s \in E_q^{\mathcal{F}}$ then $q_s > 0$ so $\exists p^i$ such that $p_s^i > 0$ which implies $s \in \bigcup_i E_{p^i}^{\mathcal{G}}$. Thus, $E_q^{\mathcal{F}} = \bigcup_i E_{p^i}^{\mathcal{G}} \in \mathcal{G}$ so $\mathcal{P}_{\mathcal{F}} \subset \mathcal{G}$. Hence $\mathcal{F} \subset \mathcal{G}$ so (2) implies (3).

Finally, assume (3) is true so $\mathcal{F} \subset \mathcal{G}$ and let $F \in \mathcal{D}_\tau$. Since F is generic under τ, for all $\{f,g\} \subset F$,

$$r\{s \in S | \, \mathbb{E}_{\mathcal{G}}[u \circ f] = \mathbb{E}_{\mathcal{G}}[u \circ g]\} \in \{0,1\}$$

Thus, $\mathbb{E}_{\mathcal{G}}[u \circ f - u \circ g] = 0$ or $\mathbb{E}_{\mathcal{G}}[u \circ f - u \circ g] \neq 0$. Since $\mathcal{F} \subset \mathcal{G}$, by repeated conditioning (see Theorem IV.1.10 of Çinlar [12]),

$$\mathbb{E}_{\mathcal{F}}[\mathbb{E}_{\mathcal{G}}[u \circ f - u \circ g]] = \mathbb{E}_{\mathcal{F}}[u \circ f - u \circ g]$$

so $\mathbb{E}_{\mathcal{F}}[u \circ f - u \circ g] = 0$ or $\mathbb{E}_{\mathcal{F}}[u \circ f - u \circ g] \neq 0$. Thus, F is generic under ρ. Since F is deterministic under τ, we can find a $f \in F$ such that

$$1 = \tau_F(f) = r\{s \in S | \mathbb{E}_{\mathcal{G}}[u \circ f] \geq \mathbb{E}_{\mathcal{G}}[u \circ g] \ \forall g \in F\}$$

By repeated conditioning again, $\mathbb{E}_{\mathcal{F}}[u \circ f - u \circ g] \geq 0$ for all $g \in F$ so

$$1 = \rho_F(f) = r\{s \in S | \mathbb{E}_{\mathcal{F}}[u \circ f] \geq \mathbb{E}_{\mathcal{F}}[u \circ g] \ \forall g \in F\}$$

so $F \in \mathcal{D}_\rho$. Hence $\mathcal{D}_\tau \subset \mathcal{D}_\rho$ so (3) implies (1). □

Appendix D

In Appendix D, we prove our results for calibrating beliefs.

Lemma (D1)**.** *Let ρ_s be represented by (μ_s, u) and $\rho_s(\underline{f}^s, \overline{f}) = 0$.*

(1) $q_s > 0$ *μ_s-a.s..*

(2) *For $F \in \mathcal{K}_s$,*

$$\int_{[0,p_s]} a \ dF^s_\rho(a) = \int_{\Delta S} \frac{r_s}{q_s}\left(1 - \sup_{f \in F} q \cdot (u \circ f)\right) \mu_s(dq)$$

Proof. Assume u is normalized without loss of generality. We prove the lemma in order:

(1) Note that

$$0 = \rho_s(\underline{f}^s, \overline{f}) = \mu_s\{q \in \Delta S | q \cdot (u \circ \underline{f}^s) \geq 1\}$$
$$= \mu_s\{q \in \Delta S | 1 - q_s \geq 1\} = \mu_s\{q \in \Delta S | 0 \geq q_s\}$$

Thus, $q_s > 0$ μ_s-a.s..

(2) Define $\psi^s_F(q) := \frac{r_s}{q_s}\left(1 - \sup_{f \in F} q \cdot (u \circ f)\right)$ and let $\lambda^F_s := \mu_s \circ (\psi^s_F)^{-1}$ be the image measure on $\mathbb{R}$. By a change of variables,

$$\int_{\mathbb{R}} x \lambda^F_s(dx) = \int_{\Delta S} \psi^s_F(q) \mu_s(dq)$$

Note that by (1), the right integral is well-defined. We now show that the cumulative distribution function of λ^F_s is exactly F^s_ρ. For $a \in [0,1]$, let $f^a_s := \underline{f}^s a \overline{f}$ and first assume

f_s^a is tied with nothing in F. Thus,

$$\begin{aligned}
\lambda_s^F[0,r_sa] &= \mu_s\circ\left(\psi_F^s\right)^{-1}[0,r_sa] = \mu_s\left\{q\in\Delta S|\, r_sa\geq\psi_F^s(q)\right\} \\
&= \mu_s\left\{q\in\Delta S|\sup_{f\in F} q\cdot(u\circ f)\geq 1-aq_s\right\} \\
&= \mu_s\left\{q\in\Delta S|\sup_{f\in F} q\cdot(u\circ f)\geq q\cdot(u\circ f_s^a)\right\} = \rho_s(F,f_s^a) = F_\rho^s(r_sa)
\end{aligned}$$

Now, if f_s^a is tied with some $g\in F$, then

$$F_\rho^s(r_sa) = \rho_s(F,f_a^s) = 1 = \mu_s\left\{q\in\Delta S|\sup_{f\in F} q\cdot(u\circ f)\geq q\cdot(u\circ f_a^s)\right\} = \lambda_s^F[0,r_sa]$$

Thus, $\lambda_s^F[0,r_sa] = F_\rho^s(r_sa)$ for all $a\in[0,1]$. Since $F\in\mathcal{K}_s$,

$$1 = F_\rho^s(r_s) = \lambda_s^F[0,r_s]$$

so F_ρ^s is the cumulative distribution function of λ_s^F.

□

Lemma (D2)**.** *Let ρ be represented by (μ,u).*

(1) *$\bar\rho$ is represented by $(\bar\mu,u)$ where $\bar\mu := \sum_s r_s\mu_s$.*

(2) *For $s\in S$, $q_s>0$ $\bar\mu$-a.s. iff $q_s>0$ $\mu_{s'}$-a.s. for all $s'\in S$.*

Proof. Let ρ be represented by (μ,u). We prove the lemma in order:

(1) Recall that the measurable sets of $\rho_{s,F}$ and $\bar\rho_F$ coincide for each $F\in\mathcal{K}$. Note that ρ_s is represented by (μ_s,u_s) for all $s\in S$. Since the ties coincide, we can assume $u_s=u$ without loss of generality. For $f\in F\in\mathcal{K}$, let

$$Q_{f,F} := \left\{q\in\Delta S|\, q\cdot(u\circ f)\geq q\cdot(u\circ f)\;\;\forall g\in F\right\}$$

Thus

$$\bar\rho_F(f) = \bar\rho_F(f_F) = \sum_s r_s\rho_{s,F}(f_F) = \sum_s r_s\mu_s(Q_{f,F}) = \bar\mu(Q_{f,F})$$

so $\bar\rho$ is represented by $(\bar\mu,u)$.

(2) Let $s \in S$ and

$$Q := \left\{ q \in \Delta S |\, q \cdot \left(u \circ \underline{f}^{s}\right) \geq u\left(\overline{f}\right)\right\} = \left\{ q \in \Delta S |\, 1 - q_{s} \geq 1\right\}$$
$$= \left\{ q \in \Delta S |\, q_{s} \leq 0\right\}$$

For any $s' \in S$, we have $\rho_{s'}\left(\overline{f}, \underline{f}^{s}\right) = 1 = \bar{\rho}\left(\overline{f}, \underline{f}^{s}\right)$ where the second inequality follows from (1). Thus, $\underline{f}^{s}$ is either tied with $\overline{f}$ or $\mu_{s'}(Q) = \mu(Q) = 0$. In the case of the former, $\mu_{s'}(Q) = \mu(Q) = 1$. The result thus follows.

□

Theorem (D3)**.** *Let ρ be represented by (μ, u). If $F_{\rho}^{s} =_{m} F_{\bar{\rho}}$, then μ is well-calibrated.*

Proof. Let $S_{+} := \left\{ s \in S |\, \rho_{s}\left(\underline{f}^{s}, \overline{f}\right) = 0\right\} \subset S$. Let $s \in S_{+}$ so $q_{s} > 0$ μ_{s}-a.s. by Lemma D1. Define the measure ν_{s} on ΔS such that for all $Q \in \mathcal{B}(\Delta S)$,

$$\nu_{s}(Q) := \int_{Q} \frac{r_{s}}{q_{s}} \mu_{s}(dq)$$

We show that $\mu = \nu_{s}$. Since $F_{\rho}^{s} =_{m} F_{\bar{\rho}}$ and by Lemmas D1 and D2, we have

$$\int_{[0,1]} a dF_{\bar{\rho}}(a) = \int_{[0,p_{s}]} a dF_{\rho}^{s}(a)$$
$$\int_{\Delta S} \left(1 - \sup_{f \in F} q \cdot (u \circ f)\right) \bar{\mu}(dq) = \int_{\Delta S} \frac{r_{s}}{q_{s}} \left(1 - \sup_{f \in F} q \cdot (u \circ f)\right) \mu_{s}(dq)$$
$$= \int_{\Delta S} \left(1 - \sup_{f \in F} q \cdot (u \circ f)\right) \nu_{s}(dq)$$

for all $F \in \mathcal{K}_{s}$.

Let $G \in \mathcal{K}$ and $F_{a} := \left(G a \overline{f}\right) \cup \underline{f}^{s}$ for $a \in (0,1)$. Since $\underline{f}^{s} \in F$, $\rho_{s}\left(F_{a}, \underline{f}^{s}\right) = 1$ so $F_{a} \in \mathcal{K}_{s}$. Let

$$Q_{a} := \left\{ q \in \Delta S \;\middle|\; \sup_{f \in G a \overline{f}} q \cdot (u \circ f) \geq q \cdot \left(u \circ \underline{f}^{s}\right)\right\}$$

and note that

$$\sup_{f \in G a \overline{f}} q \cdot (u \circ f) = h\left(a (u \circ G) + (1 - a) u\left(\overline{f}\right), q\right)$$
$$= 1 - a\left(1 - h(u \circ G, q)\right)$$

where $h(U,q)$ denotes the support function of the set U at q. Thus,

$$\int_{\Delta S}\left[1-\sup_{f\in F_a} q\cdot(u\circ f)\right]\bar{\mu}(dq)=\int_{Q_a}(a(1-h(u\circ G,q)))\,\bar{\mu}(dq)+\int_{Q_a^c} q_s\bar{\mu}(dq)$$

so for all $a\in(0,1)$,

$$\int_{Q_a}(1-h(u\circ G,q))\,\bar{\mu}(dq)+\int_{Q_a^c}\frac{q_s}{a}\bar{\mu}(dq)=\int_{Q_a}(1-h(u\circ G,q))\,\nu_s(dq)+\int_{Q_a^c}\frac{q_s}{a}\nu_s(dq)$$

Note that $q_s>0$ $\bar{\mu}$-a.s. by Lemma D2, so by dominated convergence

$$\begin{aligned}\lim_{a\to 0}\int_{Q_a}(1-h(u\circ G,q))\,\bar{\mu}(dq)&=\lim_{a\to 0}\int_{\Delta S}(1-h(u\circ G,q))\,\mathbf{1}_{Q_a\cap\{q_s>0\}}(q)\,\bar{\mu}(dq)\\&=\int_{\Delta S}(1-h(u\circ G,q))\lim_{a\to 0}\mathbf{1}_{\{q_s\geq a(1-h(u\circ G,q))\}\cap\{q_s>0\}}(q)\,\bar{\mu}(dq)\\&=\int_{\Delta S}(1-h(u\circ G,q))\,\mathbf{1}_{\{q_s>0\}}(q)\,\bar{\mu}(dq)\\&=\int_{\Delta S}(1-h(u\circ G,q))\,\bar{\mu}(dq)\end{aligned}$$

For $q\in Q_a^c$,

$$\begin{aligned}1-q_s=q\cdot(u\circ f^s)&>1-a(1-h(u\circ G,q))\\\frac{q_s}{a}&<1-h(u\circ G,q)\leq 1\end{aligned}$$

so $\int_{Q_a^c}\frac{q_s}{a}\bar{\mu}(dq)\leq\int_{\Delta S}\mathbf{1}_{Q_a^c}(q)\,\bar{\mu}(dq)$. By dominated convergence again,

$$\begin{aligned}\lim_{a\to 0}\int_{Q_a^c}\frac{q_s}{a}\bar{\mu}(dq)&\leq\lim_{a\to 0}\int_{\Delta S}\mathbf{1}_{Q_a^c}(q)\,\bar{\mu}(dq)\\&\leq\int_{\Delta S}\lim_{a\to 0}\mathbf{1}_{\{q_s<a(1-h(u\circ G,q))\}}(q)\,\bar{\mu}(dq)\\&\leq\int_{\Delta S}\mathbf{1}_{\{q_s=0\}}(q)\,\bar{\mu}(dq)=0\end{aligned}$$

By a symmetric argument for ν_s, we have

$$\int_{\Delta S}(1-h(u\circ G,q))\,\bar{\mu}(dq)=\int_{\Delta S}(1-h(u\circ G,q))\,\nu_s(dq)$$

for all $G\in\mathcal{K}$. Letting $G=\underline{f}$ yields $1=\bar{\mu}(\Delta S)=\nu_s(\Delta S)$ so ν_s is a probability measure on ΔS and

$$\int_{\Delta S}\sup_{f\in G} q\cdot(u\circ f)\,\bar{\mu}(dq)=\int_{\Delta S}\sup_{f\in G} q\cdot(u\circ f)\,\nu_s(dq)$$

Thus, $\bar{\mu} = \nu_s$ for all $s \in S$ by the uniqueness properties of the subjective learning representation (Theorem 1 of DLST). As a result,

$$\int_Q \frac{q_s}{r_s} \bar{\mu}(dq) = \int_Q \frac{q_s}{r_s} \nu_s(dq) = \mu_s(Q)$$

for all $Q \in \mathcal{B}(\Delta S)$ and $s \in S_+$.

Finally, for $s \notin S_+$, $\rho_s\left(\underline{f}^s, \overline{f}\right) = 1$ so $q_s = 0$ μ_s-a.s.. By Lemma D2, $q_s = 0$ μ-a.s.. Let

$$Q_0 := \left\{ q \in \Delta S \;\middle|\; \sum_{s \notin S_+} q_s = 0 \right\}$$

and note that $\mu(Q_0) = 1$. Now,

$$\begin{aligned} \sum_{s \in S_+} r_s = \sum_{s \in S_+} \int_{\Delta S} q_s \bar{\mu}(dq) &= \int_{Q_0} \sum_{s \in S_+} q_s \bar{\mu}(dq) \\ &= \int_{Q_0} \left(\sum_{s \in S} q_s \right) \bar{\mu}(dq) = \bar{\mu}(Q_0) = 1 \end{aligned}$$

which implies $\sum_{s \notin S_+} r_s = 0$ a contradiction. Thus, $S_+ = S$ and μ is well-calibrated. □

Theorem (D4)**.** *Let ρ be represented by (μ, u). If μ is well-calibrated, then $F_\rho^s =_m F_{\bar{\rho}}$.*

Proof. Note that the measurable sets and ties of ρ_s and $\bar{\rho}$ coincide by definition. As above, let $S_+ := \left\{ s \in S \middle| \rho_s\left(\underline{f}^s, \overline{f}\right) = 0 \right\} \subset S$. Thus, $s \notin S_+$ implies $\underline{f}^s$ and $\overline{f}$ are tied and $q_s = 0$ a.s. under all measures. By the same argument as the sufficiency proof above, letting $Q_0 := \left\{ q \in \Delta S \middle| \sum_{s \notin S_+} q_s = 0 \right\}$ yields

$$\sum_{s \in S_+} r_s = \sum_{s \in S_+} \int_{\Delta S} q_s \bar{\mu}(dq) = \int_{Q_0} \left(\sum_{s \in S} q_s \right) \bar{\mu}(dq) = 1$$

a contradiction. Thus, $S_+ = S$.

Let $F \in \mathcal{K}_s$ and $s \in S$. Since $\rho_s\left(\underline{f}^s, \overline{f}\right) = 0$, by Lemmas A12 and D1 and the fact that

μ is well-calibrated,

$$\begin{aligned}\int_{[0,p_s]} a dF^s_\rho (a) &= \int_{\Delta S} \frac{r_s}{q_s}\left(1-\sup_{f\in F} q\cdot(u\circ f)\right)\mu_s(dq)\\ &= \int_{\Delta S} \frac{r_s}{q_s}\left(1-\sup_{f\in F} q\cdot(u\circ f)\right)\frac{q_s}{r_s}\bar{\mu}(dq)\\ &= \int_{\Delta S}\left(1-\sup_{f\in F} q\cdot(u\circ f)\right)\bar{\mu}(dq) = \int_{[0,1]} a dF_{\bar{\rho}}(a)\end{aligned}$$

so $F^s_\rho =_m F_{\bar{\rho}}$. □

Appendix E

In this section, we relate our model with that of Ahn and Sarver [1]. We focus on the individual interpretation for ease of comparison. Ahn and Sarver introduce a condition called consequentialism to link choice behavior from the two time periods.[35] Consequentialism translates into the following in our setting.

Axiom (Consequentialism)**.** *If $\rho_F = \rho_G$, then $F \sim G$.*

However, consequentialism fails as a sufficient condition for linking the two choice behaviors in our setup. This is demonstrated in the following.

Example. Let $S = \{s_1, s_2\}$, $X = \{x, y\}$ and $u(a\delta_x + (1-a)\delta_y) = a$. Associate each $q \in \Delta S$ with $t \in [0,1]$ such that $t = q_{s_1}$. Let μ have the uniform distribution and ν have density $6t(1-t)$. Thus, μ is more informative than ν. Let $\succeq$ be represented by (μ, u) and ρ be represented by (ν, u). We show that $(\succeq, \rho)$ satisfies consequentialism. Let $F^+ \subset F \cap G$ denote the support of $\rho_F = \rho_G$. Since $f \in F \backslash F^+$ implies it is dominated by F^+ μ-a.s., it is also dominated by F^+ ν-a.s. so $F \sim F^+$. A symmetric analysis for G yields $F \sim F^+ \sim G$. Thus, consequentialism is satisfied, but $\mu \neq \nu$.

The reason for why consequentialism fails in the Anscombe-Aumann setup is that the representation of DLR is more permissive than that of DLST. In the lottery setup, if consequentialism is satisfied, then this extra freedom allows us to construct an ex-ante representation that is completely consistent with that of ex-post random choice. On the other hand,

[35] Their second axiom deals with indifferences which we resolve using non-measurability.

information is uniquely identified in the representation of DLST, so this lack of flexibility prevents us from performing this construction even when consequentialism is satisfied. A stronger condition is needed to perfectly equate choice behavior from the two time periods.

Axiom (Strong Consequentialism). *If F_ρ and G_ρ share the same mean, then $F \sim G$.*

The following demonstrates why this is a strengthening of consequentialism.

Lemma (E1). *For ρ monotonic, $\rho_F = \rho_G$ implies $F_\rho = G_\rho$.*

Proof. Let ρ be monotonic and define $F^+ := \{f \in H | \, \rho_F(f) > 0\}$. We first show that $F_\rho^+ = F_\rho$. Let $F^0 := F \backslash F^+$ and for $a \in [0,1]$, monotonicity yields

$$0 = \rho_F\left(F^0\right) \geq \rho_{F \cup f^a}\left(F^0\right)$$

Note that by Lemma A2, $\{F^0, F^+\} \in \mathcal{H}_F$. First, suppose f^a is tied with nothing in F. Hence,

$$\rho_{F^+ \cup f^a}\left(F^+\right) + \rho_{F^+ \cup f^a}\left(f^a\right) = 1 = \rho_{F \cup f^a}\left(F^+\right) + \rho_{F \cup f^a}\left(f^a\right)$$

By monotonicity, $\rho_{F^+ \cup f^a}\left(F^+\right) \geq \rho_{F \cup f^a}\left(F^+\right)$ and $\rho_{F^+ \cup f^a}\left(f^a\right) \geq \rho_{F \cup f^a}\left(f^a\right)$ so

$$F_\rho^+(a) = \rho_{F^+ \cup f^a}\left(F^+\right) = \rho_{F \cup f^a}\left(F^+\right) = \rho_{F \cup f^a}(F) = F_\rho(a)$$

Now, if f^a is tied with some act in F, then by Lemma A3 and monotonicity,

$$1 = \rho_F\left(F^+\right) = \rho_{F \cup f^a}\left(F^+\right) \leq \rho_{F^+ \cup f^a}\left(F^+\right)$$

Thus, $F_\rho^+(a) = 1 = F_\rho(a)$ so $F_\rho^+ = F_\rho$.

Now, suppose $\rho_F = \rho_G$ for some $\{F, G\} \subset \mathcal{K}$. Since $\rho_F(f) > 0$ iff $\rho_G(f) > 0$, $F^+ = G^+$. We thus have

$$F_\rho = F_\rho^+ = G_\rho^+ = G_\rho$$

□

Thus, if strong consequentialism is satisfied, then consequentialism must also be satisfied as $\rho_F = \rho_G$ implies $F_\rho = G_\rho$ which implies that F_ρ and G_ρ must have the same mean. Strong consequentialism delivers the corresponding connection between ex-ante and ex-post choice behaviors that consequentialism delivered in the lottery setup.

Proposition (E2). *Let $\succeq$ and ρ be represented by (μ, u) and (ν, v) respectively. Then the following are equivalent:*

(1) $(\succeq, \rho)$ *satisfies strong consequentialism*

(2) $F \succeq G$ *iff* $F \succeq_\rho G$

(3) $(\mu, u) = (\nu, \alpha v + \beta)$ *for* $\alpha > 0$

Proof. Note that the equivalence of (2) and (3) follows from Theorem B6 and the uniqueness properties of the subjective learning representation (see Theorem 1 of DLST). That (2) implies (1) is immediate, so we only need to prove that (1) implies (2).

Assume (1) is true. Since $\succeq_\rho$ is represented by (ν, v), we have $F \sim_\rho G$ implies $F \sim G$. Without loss of generality, we assume both u and v are normalized. First, consider only constant acts and let $\underline{f}$ and $\overline{f}$ be the worst and best acts under v. Now, for any $f \in H_c$, we can find $a \in [0,1]$ such that $\underline{f} a \overline{f} \sim_\rho f$ which implies $\underline{f} a \overline{f} \sim f$. Thus

$$v(f) = v\left(\underline{f} a \overline{f}\right) = 1 - a$$

and

$$\begin{aligned} u(f) &= au\left(\underline{f}\right) + (1-a)\,u\left(\overline{f}\right) = (1 - v(f))\,u\left(\underline{f}\right) + v(f)\,u\left(\overline{f}\right) \\ &= \left(u\left(\overline{f}\right) - u\left(\underline{f}\right)\right) v(f) + u\left(\underline{f}\right) \end{aligned}$$

for all $f \in H_c$. Thus, $u = \alpha v + \beta$ where $\alpha := u\left(\overline{f}\right) - u\left(\underline{f}\right)$ and $\beta := u\left(\underline{f}\right)$. Since $\underline{f} \cup \overline{f} \sim_\rho \overline{f}$ implies $\underline{f} \cup \overline{f} \sim \overline{f}$, we have $u\left(\overline{f}\right) \geq u\left(\underline{f}\right)$ so $\alpha \geq 0$. If $\alpha = 0$, then $u = \beta$ contradicting the fact that u is non-constant. Thus, $\alpha > 0$.

We can now assume without loss of generality that $\succeq_\rho$ is represented by (ν, u). Now, given any $F \in \mathcal{K}$, we can find $f \in H_c$ such that $F \sim_\rho f$ which implies $F \sim g$. Thus,

$$\int_{\Delta S} \sup_{f \in F} q \cdot (u \circ f)\, \nu(dq) = u(g) = \int_{\Delta S} \sup_{f \in F} q \cdot (u \circ f)\, \mu(dq)$$

so $\succeq_\rho$ and $\succeq$ represent the same preference which implies (2). Thus, (1), (2) and (3) are all equivalent. $\square$