Random Choice and Private Information*

Jay Lu†

January 2015

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 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, Kota Saito and seminar audiences at Yale, UCLA, Northwestern, Kellogg MEDS, Cornell, MIT, UBC, BU, UPenn, 14th SAET Conference at Waseda, USC, Caltech, Duke-UNC, NYU and Wisconsin for helpful suggestions and constructive feedback.

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

1 Introduction

1.1 Overview and motivation

In many economic situations, an agent's private information is not observable. For example, consider an agent choosing whether to invest in a project. Before deciding, the agent obtains some private information (i.e. a signal) that influences her beliefs about the prospects of the project. As a result, her investment decision depends on the realization of her signal. An analyst (i.e. an outside observer) does not observe the agent's signal so from the former's perspective, the agent's choice is probabilistic or *random*. Call this the *individual interpretation* of random choice. Alternatively, consider a group of agents choosing whether to invest in a project where each agent has some private information about the project that is unobserved by the analyst. From the analyst's perspective, an agent's choice within the group is random, and we call this the *group interpretation* of random choice.[1]

In this paper, we provide a theory for identifying private information from random choice. We model a standard expected utility maximizer who chooses optimally after receiving a signal that is *strictly private*, that is, it is completely unobservable to the analyst. Many problems in information economics and decision theory fit in this framework. These include specialists providing expert advice[2], consumers choosing health insurance[3] and users clicking on online ads.[4]

Our main results provide a methodology for identifying information and performing standard exercises of inference. First, the analyst can *evaluate option sets* by directly computing ex-ante valuations of option sets from random choice and vice-versa. Second, the analyst can use choice dispersion to *assess the informativeness* of agents' signals. Third, if valuations of option sets are known, then the analyst can compare them with random choice to *detect biases*, that is, when beliefs about information are dynamically inconsistent. Finally, from the joint distribution of choices and payoff-relevant states, the analyst can *calibrate beliefs*

[1] Note that all agents in the group are initially observationally identical to the analyst. We can think of this as the end result after applying all possible econometric analysis (parametric and non-parametric) on the observable data to differentiate agents.

[2] For example, see Chambers and Lambert [10].

[3] For example, see Finkelstein and McGarry [18] and Hendren [26] who provide an alternative approach by directly eliciting private information from survey data. Respondents however may not accurately report their true beliefs or the data may be subject to other complications (such as excess concentrations at focal points). In contrast, our approach follows the original spirit of Savage [43] by inferring beliefs from choice behavior.

[4] For example, see Rayo and Segal [40].

and test if agents' beliefs are rational.

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.

Consider an objective state space. Each choice option corresponds to a state-contingent payoff or an Anscome-Aumann [2] *act*. A set of acts is a *decision-problem*. Since the agent's private information is unobservable to the analyst, the analyst only observes a *random choice rule (RCR)* that specifies a choice distribution over acts for each decision-problem. We consider a random utility maximization (RUM)[5] model where the utilities are subjective expected utilities that depend only on the distribution of the agent's beliefs. As in Savage [43] and Anscombe and Aumann [2], we assume that beliefs are independent of the decision-problem.[6] The probability that an act is chosen is then the probability that the act attains the highest subjective expected utility in the decision-problem. Call this an *information representation* of the RCR.[7]

Theorem 1 shows that the analyst can completely identify the agent's private information from binary choices. We introduce a key tool that will feature prominently in the subsequent analysis. Given a decision-problem, consider the addition of a *test act* (for example, an act that gives 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.

We first *evaluate option sets*. In the individual interpretation, the valuation of an option

[5] For more about RUM, see Block and Marschak [7], Falmagne [17], McFadden and Richter [35], Gul and Pesendorfer [24] and Gul, Natenzon and Pesendorfer [25]. RUM is also used extensively in discrete choice estimation where most models assume specific parametrizations such as the logit, the probit, the nested logit, etc. (see Train [46]).

[6] In the group interpretation, the assumption that the distribution of beliefs is independent of the decision-problem corresponds exactly to the assumption that each agent's prior in the Anscombe and Aumann [2] model is independent of the decision-problem.

[7] In the Supplementary Appendix, we provide a full axiomatic characterization of an information representation along with other more general representations.

set is the ex-ante utility of the set before any information is received and corresponds to the subjective learning representation of Dillenberger, Lleras, Sadowski and Takeoka [15] (henceforth DLST). In the group interpretation, the valuation of an option set is the total utility of the set for all agents in the group.[8] Theorem 2 shows that computing integrals of test functions recovers valuations. Conversely, Theorem 3 shows that computing the marginal valuations of decision-problems with respect to test acts recovers random choice. These operations are mathematical inverses; the analyst can directly compute valuations from random choice and vice-versa. This provides a precise connection between menu (i.e. option set) choice and random choice and offers a methodology for elicitation that is similar to classical results from consumer and producer theory (Theorem 3 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 4 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 relations. Intuitively, a more informative signal structure (or more private information in a group of agents)[9] is characterized by greater dispersion or randomness in choice.

We then apply these results to *detect biases*. We address a form of informational dynamic inconsistency where the ex-ante preference relation (reflecting valuations) suggests a more (or less) informative signal than that implied by random choice. In *prospective overconfidence*, the agent initially prefers large option sets in anticipation of an informative signal but subsequently exhibits deterministic choice reflecting a less informative signal. In *prospective underconfidence*, the ordering is reversed. In either case, the agent exhibits *subjective misconfidence*. These biases also apply in the group interpretation. For example, if a firm chooses option sets based on total employee welfare (i.e valuation), then misconfidence suggests that the firm has an incorrect assessment of the distribution of employee beliefs. Our tools allow the analyst to detect these biases even when information is not directly observable.

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

[9] In the group interpretation, Theorem 4 is the revealed preference analog of Hendren [26] who uses elicited beliefs from survey data to test whether there is more private information in one group of agents (insurance rejectees) than another group (non-rejectees).

Finally, we *calibrate beliefs*. We show that given joint data on choices *and* actual state realizations, the analyst can test whether the agent has well-calibrated (i.e. consistent) beliefs. 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, indicating that there is genuine private information.[10] Define a *conditional test function* using a *conditional test act* with payoffs that vary only in a given state. Theorem 5 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 which can be combined with the results on detecting biases to obtain measures of objective misconfidence.

In general, RUM models have difficulty dealing with indifferences in the random utility. We address this issue by drawing an analogy with deterministic choice. Under deterministic choice, if two acts are indifferent (i.e. they have the same utility), 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: (1) 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, and (2) 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 and Aumann [2] is a special case of our model when choice is completely deterministic.

1.2 Related literature

This paper is related to a long literature on stochastic choice. Recent papers that have specifically studied the relationship between stochastic choice and information include Natenzon [37], Caplin and Dean [8], Matějka and McKay [33] and Ellis [16]. In these models, the information structure varies with the decision-problem so the resulting RCR may not have a RUM representation. In contrast, the information structure in our model is fixed and is independent of the decision-problem, which conforms to the benchmark model of information processing and choice. Caplin and Martin [9] also study a RUM model with a fixed infor-

[10] See Chiappori and Salanié [11] and Finkelstein and McGarry [18] for empirical tests for the presence of private information.

mation structure. If we recast their model in our Anscombe-Aumann setup, our use of test functions to calibrate beliefs coincides with checking their NIAS inequalities (see Proposition 1). Note that by working in our richer setup, information can be uniquely identified from the RCR.

This paper is also related to the literature on menu choice which includes Kreps' [31], Dekel, Lipman and Rustichini [14] (henceforth DLR) and DLST. Our main contribution to this literature is showing that there is an intimate connection between ex-ante choice *over* option sets and ex-post random choice *from* option sets. Ahn and Sarver [1] (henceforth AS) also study this relationship but in the lottery space, and their work connecting DLR with Gul and Pesendorfer [24] random expected utility is analogous to our results connecting DLST with our model. 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 (see Appendix E). Fudenberg and Strzalecki [20] also analyze the relationship between preference for flexibility and random choice but in a dynamic setting with a generalized logit model. Saito [42] 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 [22, 23] study decision-theoretic models involving information. 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, the agent in our model prefers information only due to its instrumental value as in the classical sense of Blackwell.

Finally, 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. Kamenica and Gentzkow [28] and Rayo and Segal [40] characterize optimal information structures where senders commit to an information disclosure policy. In these models, the sender's ex-ante utility is a function of the receiver's random choice rule, so our results relating random choice with valuations provide a technique for expressing the sender's utility in terms of the receiver's utility and vice-versa.

2 An Informational Model of Random Choice

Let S be a finite objective state space and X be a finite set of prizes. 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 an Anscombe-Aumann [2] *act*, that is, a mapping $f : S \to \Delta X$. Let H be the set of all acts. A finite set of acts is called a *decision-problem* and let $\mathcal{K}$ be the set of all decision-problems endowed with the Hausdorff metric.[11] Call an act f *constant* iff $f(s)$ is the same for all $s \in S$. For notational convenience, let f denote the singleton set $\{f\}$ whenever there is no risk of confusion.

The primitive of our model 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.

Under classic deterministic 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 the random choice rule is unable to specify choice probabilities for each act in the decision-problem. As in deterministic choice, we interpret indifference as choice behavior that is beyond the scope of the model. This provides the analyst with additional freedom to interpret data. It also allows for just enough flexibility so that we can include the deterministic Anscombe-Aumann [2] model as a special case of ours.

Formally, indifferences correspond to non-measurability with respect to a 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. Indifferences occur when $\mathcal{H}$ is coarser than the Borel algebra. Since the agent will be choosing some act in a decision-problem, the decision-problem itself must be measurable. Hence, given any decision-problem

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

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

F, the corresponding choice distribution must be a measure on the σ-algebra generated by $\mathcal{H} \cup \{F\}$ which we denote by $\mathcal{H}_F$.[12] Let Π be the set of all probability measures on H.

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

For each decision-problem F, the RCR ρ assigns a probability measure on $(H, \mathcal{H}_F)$ such that $\rho_F(F) = 1$. Interpret $\rho_F(G)$ as the probability that some act in G will be chosen in $F \in \mathcal{K}$. For ease 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 may not be $\mathcal{H}_F$-measurable, define the outer measure[13]

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

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

An 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) how classic deterministic choice is a special case of random choice.

Example 1. Let $S = \{s_1, s_2\}$ and $X = \{x, y\}$. Without loss of generality, we can associate H with $[0,1] \times [0,1]$ where $f_i = f(s_i)(x)$ for $i \in \{1,2\}$. 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 and they are indifferent. Observe that ρ corresponds exactly to classic deterministic choice where f is preferred to g iff $f_1 \geq g_1$.

We now describe an *information representation* of an RCR. Recall the timing of our model. At time 1, the agent receives some private information about the underlying state. At time 2, she chooses the best act in the decision-problem given her updated belief. Since her private information is unobservable, to the analyst, choice is probabilistic and can be modeled as an RCR.

[12] This definition imposes a common measurability across all decision-problems which can be relaxed if our axioms (see Supplementary Appendix) are strengthened.

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

Since each signal realization corresponds to a posterior belief $q \in \Delta S$, we model private information as a *signal distribution* μ over the canonical signal space ΔS. This approach allows us to be agnostic about updating and work directly with posterior beliefs. To illustrate, 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 never changes her prior. In the group interpretation, this corresponds to the case where all agents in the group share a common belief. Note that in either interpretation, the resulting RCR is deterministic.

Let $u : \Delta X \to \mathbb{R}$ be an affine utility function. An agent's subjective expected utility of an act f given her belief q is $q \cdot (u \circ f)$.[14] Given a utility function, a signal distribution is *regular* iff the subjective expected utilities of two acts are either always or never equal. This a relaxation of the standard restriction in traditional RUM where utilities are never equal and allows us to handle indifferences.

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 the agent's private information. The probability of choosing an act f is exactly the measure of beliefs that rank f higher than every other act in the decision-problem. Although both μ and u are unobserved, the analyst can infer about the agent's private information and taste utility by studying her RCR. Note that when $\mu = \delta_q$, this reduced to the standard subjective expected utility model of Anscombe and Aumann [2].

Since we are interesting in studying the role of information in random choice, the taste utility u is fixed in an information representation. In the individual interpretation, this implies that signals only affect beliefs but not tastes. In the group interpretation, this implies that agents have unobserved beliefs but observed tastes (e.g. risk aversion). To the analyst, choice is random only as a result of unobserved belief shocks.[15]

[14] 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 $s \in S$.

[15] In the Supplementary Appendix, we provide an axiomatic characterization of a more a general model that allows for unobserved utility shocks as well.

One of the classic critiques of subjective expected utility (especially in the context of choosing health insurance for example) is the state independence of the (taste) utility. This can be addressed in our model by considering random choice generalizations of classic solutions to state-dependent utility as in Karni, Schmeidler and Vind [30] and Karni [29]. Note that in practice however, the empirical literature on health insurance has largely assumed state independence due to lack of better empirical evidence.[16]

Theorem 1 below 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.[17]

Theorem 1 (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. □

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. Our definition of regularity circumvents this by allowing for just enough flexibility so that we can model indifferences using non-measurability. Note that our definition still imposes certain restrictions on μ. For example, multiple mass points are not allowed if μ is regular.[18]

[16] See Finkelstein, Luttmer and Notowidigdo [19].

[17] 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.

[18] 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 [24] 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.

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.

An act is 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 constant best and worst acts.[19] *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 $a \in [0,1]$.

Note that test acts are also constant acts. 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

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

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. Test functions 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. □

What follows is an example of a test function.

Example 2. Let $S = \{s_1, s_2\}$, $X = \{x, y\}$ and $u\left(a\delta_x + (1-a)\,\delta_y\right) = a \in [0,1]$. Consider the decision-problem $F = \{f, g\}$ where $u \circ f = \left(\frac{2}{5}, \frac{2}{5}\right)$ and $u \circ g = \left(\frac{1}{4}, \frac{3}{4}\right)$. Set μ such that

[19] To see this, note that ρ induces a preference relation over constant acts that is represented by u. Since u is affine and X is finite, we can always find a best and worst act.

the density of q_{s_1} is $6q_{s_1}(1-q_{s_1})$. Let ρ be represented by (μ, u). The test function of F is

$$
\begin{aligned}
F_\rho(a) &= \mu \left\{ q_{s_1} \in [0,1] \;\middle|\; \max\left\{ \frac{2}{5},\ \frac{1}{4} q_{s_1} + \frac{3}{4}(1-q_{s_1}) \right\} \geq 1-a \right\} \\
&= \begin{cases} 0 & \text{if } a < \frac{1}{4} \\ (4a-1)^2 (1-a) & \text{if } \frac{1}{4} \leq a < \frac{3}{5} \\ 1 & \text{if } \frac{3}{5} \leq a \end{cases}
\end{aligned}
$$

which is a cumulative distribution function.

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 they are sufficient statistics 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 immediately from Theorem 1. □

4 Evaluating Option Sets

We now address our first exercise of inference and 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, then she will exhibit a strict preference for flexibility. This is the subjective learning representation axiomatized by 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. For instance, consider an entity (e.g. a firm) that completely internalizes the utilities of all agents (e.g. employees) in a group and makes decisions based on total welfare. If the entity thinks that agent beliefs are very dispersed, then it would prefer more flexible (i.e. larger) option sets.

In this section, assume that ρ has a best and worst act and F_ρ is a well-defined cumulative for all $F \in \mathcal{K}$. 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.[20] Consider the following properties of RCRs.

Definition. ρ is *monotone* iff $G \subset F$ implies $\rho_G(f) \geq \rho_F(f)$.

Definition. ρ is *linear* iff $\rho_F(f) = \rho_{aF+(1-a)g}(af + (1-a)g)$ for $a \in (0,1)$.

Definition. ρ is *continuous* iff it is continuous on $\mathcal{K}_0$.[21]

An RCR is *standard* iff it is monotone, linear and continuous. Monotonicity is necessary for any RUM while linearity is the random choice analog of the standard independence axiom. Continuity is the usual continuity adjusted for indifferences. Any RCR that has an information representation is standard, although the condition is a relatively weak restriction; in fact, it is insufficient to guarantee that a random utility representation even exists.[22]

Consider evaluating decision-problems as follows.

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

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

A decision-problem that is valuable ex-ante will have acts that are chosen more frequently over a potential test act, yielding a test function that takes on high values. Theorem 2 confirms that $\succeq_\rho$ is the valuation preference relation corresponding to the RCR ρ.

Theorem 2. *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. □

[20] That is, $f \in \mathcal{H}_F$ for all $f \in F \in \mathcal{K}_0$.

[21] In other words, $\rho : \mathcal{K}_0 \to \Pi_0$ is continuous where Π_0 is the set of all Borel measures on H endowed with the topology of weak convergence. Note that $\rho_F \in \Pi_0$ for all $F \in \mathcal{K}_0$ without loss of generality.

[22] See the axiomatic treatment in the Supplementary Appendix. In particular, extremeness is necessary in order to ensure the existence of a random expected utility representation.

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

Theorem 2 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, given any act f and state s, let f_s denote the constant act that yields $f(s)$ in every state.

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. Consider an 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 3 shows that these issues are irrelevant; if $\succeq$ has a subjective learning representation, then $\rho_\succeq$ exists and is the unique RCR corresponding to $\succeq$.

Theorem 3. *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. ∎

The probability that an act f_a is chosen is exactly its marginal contribution to the valuation of the decision-problem.[23] The more often the act is chosen, the greater its effect on the decision-problem's overall valuation. For instance, if the act is never chosen, then it will have no effect on valuations (this is a cardinal version of Axiom 1 in AS).[24] Any violation of this would indicate some form of inconsistency, which we explore in Section 6.

The analyst can now use $\rho_{\succeq}$ to directly compute random choice from valuations. To see how, first define ρ so that it coincides with $\succeq$ 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 $F \in \mathcal{K}$. Linearity then extends ρ to all decision-problems. By Theorem 3, the ρ so constructed is represented by (μ, u).

The other implication is that if a dominant preference relation induces an RCR that has an information representation, then that preference relation has a subjective learning representation. As in Theorem 2, this is an alternate characterization of subjective learning representations using properties of its induced RCR.

Theorem 3 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".[25] Similar to how Hotelling's Lemma is used to compute firm outputs from the profit function, Theorem 3 can be used to compute random choice from valuations.

Thus, similar to how classical results from consumer and producer theory (e.g. Hotelling's Lemma) provide a methodology for relating data, Theorems 3 and 4 allow an analyst to relate valuations with random choice and vice-versa. Integrating test functions yields valuations, while differentiating valuations yields random choice. This is a method of direct computation that completely bypasses the need to identify the signal distribution or taste utility. Observing choice data in one time period allows the analyst to directly compute choice data in the other. We summarize these results below.

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

[23] In the econometrics literature, this is related to the Williams-Daly-Zachary Theorem that follows from an envelope argument (see McFadden [34]). The presence of constant acts in the Anscombe-Aumann setup however means Theorem 2 has no counterpart.

[24] See Appendix E for a more detailed discussion.

[25] 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 $\mathcal{Q}_y$.

Proof. Follows immediately from Theorems 3 and 4. □

5 Assessing Informativeness

We now show how the analyst can infer who gets better information even when information is not directly observable. First, consider the classic methodology when information *is* observable. A transition kernel[26] on ΔS is *mean-preserving* iff it preserves average beliefs.

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] ordering of informativeness based on signal sufficiency. If K is the identity kernel for example, 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. We now show how to characterize better information using random choice. 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 an act corresponds to a singleton mass point. Another agent (or group of agents) with more information will have test functions that have a more dispersed distribution. This is captured exactly by *second-order stochastic dominance*, that is, $F \geq_{SOSD} G$ iff $\int_{\mathbb{R}} \phi dF \geq \int_{\mathbb{R}} \phi dG$ for all increasing concave $\phi : \mathbb{R} \to \mathbb{R}$.

[26] $K : \Delta S \times \mathcal{B}(\Delta S) \to [0,1]$ 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 $q \in \Delta S$.

Theorem 4. *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. $\square$

Theorem 4 equates an unobservable multi-dimensional information ordering 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. We illustrate with an example.

Example 3. Recall Example 2 from above and now let ν be the uniform distribution. Let τ be represented by (ν, u). The test function of $F = \{f, g\}$ under τ is

$$F_\tau(a) = \begin{cases} 0 & \text{if } a < \frac{1}{4} \\ 2a - \frac{1}{2} & \text{if } \frac{1}{4} \leq a < \frac{3}{5} \\ 1 & \text{if } \frac{3}{5} \leq a \end{cases}$$

Note that ν is more informative than μ and $F_\rho \geq_{SOSD} F_\tau$ as desired.

In DLST, better information is characterized by a greater preference for flexibility in the valuation preference relation. This is the choice-theoretic 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. Corollary 3 relates our random choice characterization of better information with more preference for flexibility.

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

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 4, (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). $\square$

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 choice. 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 from the Blackwell ordering that the two distributions must have the same average belief. Combined with Theorem 4, this implies that test functions of singleton acts under the more informative signal is a mean-preserving spread of those under the less informative signal. This condition however is insufficient for assessing informativeness; it corresponds to a strictly weaker stochastic dominance relation known as the linear concave order.[27] Note that an analyst cannot assess informativeness by studying valuations of singleton acts since all singletons have the same ex-ante valuation under two signals with the same average belief. However, looking at test functions will allow the analyst to assess informativeness so the random choice characterization of better information may sometimes be richer than the valuation one.

6 Detecting Biases

In this section, we study situations when the information inferred from valuations is inconsistent with that inferred from random choice. 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 dynamic inconsistency as in Strotz [45]. In the group interpretation, this describes the situation when valuations of option sets indicate a more (or less) dispersed distribution of beliefs in a group than that implied by random choice.

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 option sets containing a large

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

variety of foods, in the end, she always chooses the same food from the set.[28] 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 behavior 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 a confirmatory bias where consecutive good and consecutive bad signals generate posterior beliefs that are more dispersed.[29] Call this *prospective underconfidence.*

Since beliefs in our model are subjective, we are silent as to which period's 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 is in some sense more fundamental.[30]

Let $(\succeq, \rho)$ denote the valuation preference relation $\succeq$ and the RCR ρ. Motivated by Theorem 4, define subjective misconfidence as follows.[31]

Definition. $(\succeq, \rho)$ exhibits *prospective overconfidence (underconfidence) iff* $F_\rho \geq_{SOSD} F_{\rho_\succeq}$ $(F_\rho \leq_{SOSD} F_{\rho_\succeq})$ for all $F \in \mathcal{K}$.

Corollary 4. *Let $\succeq$ and ρ be represented by (μ, u) and (ν, u) respectively. Then $(\succeq, \rho)$ exhibits prospective overconfidence (underconfidence) iff μ is more (less) informative than ν.*

Proof. Follows immediately from Corollary 2 and Theorem 4. □

Corollary 4 provides a choice-theoretic foundation for subjective misconfidence. We can also apply it to other behavioral biases involving information processing such as the hot-hand

[28] 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).

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

[30] In Section 7, we show how an analyst can discern which period's choice behavior is correct by studying a richer data set (e.g. the joint data over choices and state realizations).

[31] Note that by Corollary 3, we could redefine prospective overconfidence (underconfidence) solely in terms of more (less) preference for flexibility.

fallacy and the gambler's fallacy.[32] If we assume that the agent is unaffected by these biases ex-ante but she becomes afflicted ex-post, then the hot-hand and gambler's fallacies correspond to prospective underconfidence and overconfidence respectively. Corollary 4 also allows us to rank the severity of these biases via the Blackwell ordering of information structures and provides a unifying methodology to study a wide variety of behavioral biases.

7 Calibrating Beliefs

Following 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.[33] 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, the 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.[34] We show how the analyst can use test functions to calibrate beliefs even when information is not observable.

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 *state-dependent random choice rule (sRCR)* $\rho := (\rho_s)_{s \in S}$ that specifies a RCR for each state.[35] Let $\rho_{s,F}(f)$ denote the probability of choosing $f \in F$ given state $s \in S$. The unconditional RCR is

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

Note that r in conjunction with the sRCR ρ completely specify the joint distribution over

[32] See Gilovich, Vallone and Tversky [21] and Rabin [38] respectively.

[33] In the Supplementary Appendix, we also provide an axiomatic treatment that incorporates the observed distribution of states as part of the primitive.

[34] For example, see Dawid [13].

[35] Formally, an sRCR consists of $(\rho, \mathcal{H})$ where $\rho : S \times \mathcal{K} \to \Pi$ and $(\rho_s, \mathcal{H})$ is an RCR for all $s \in S$.

choices and state realizations.[36]

Information now corresponds to a joint distribution over beliefs and state realizations. In this section, let $\mu := (\mu_s)_{s\in S}$ be a *state-dependent signal distribution* where μ_s is the signal distribution conditional on $s \in S$. Let (μ, u) denote a state-dependent signal distribution μ and a non-constant u. We now define a *state-dependent information representation* and say ρ is represented by (μ, u) iff ρ_s is represented by (μ_s, u) for all $s \in S$. Note that this does not imply beliefs are well-calibrated since μ may not be consistent with the observed frequency r. Define the unconditional signal distribution as

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

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 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 have rational (i.e. correct) beliefs about their payoff-relevant state so there is genuine private information in the group.

We now show how the analyst can test for well-calibrated beliefs using test functions. Let ρ be represented by (μ, u). Since u is fixed, both the best and worst acts are well-defined for ρ. Given a state $s \in S$, define a *conditional worst act* $\underline{f}^s$ as the act that coincides with the worst act if s occurs and with the best act otherwise.[37] Let $f_s^a := a\underline{f}^s + (1-a)\overline{f}$ be a *conditional test act* and define a *conditional test function* 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

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

A conditional test function specifies the conditional choice probability as we vary the conditional test act from the best to the conditional worst act. As with unconditional

[36] State-dependent stochastic choice was studied by Caplin and Martin [9] and Caplin and Dean [8] who demonstrate the feasibility of collecting such data for individuals. In the group interpretation, this data is also readily available (see Chiappori and Salanié [11]).

[37] That is, $\underline{f}^s(s) = \underline{f}(s)$ and $\underline{f}^s(s') = \bar{f}(s')$ for all $s' \neq s$.

test functions, conditional test functions are increasing functions that are cumulatives if $F_{\rho}^{s}(r_s) = 1$. Let $\mathcal{K}_s$ denote all decision-problems with conditional test functions that are cumulatives.

Theorem 5. *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 5 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 using test functions.

Suppose that in addition to the sRCR 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. By enriching choice behavior with data on state realizations, the analyst can now make claims about objective belief misalignment.

Finally, we relate Theorem 5 to Caplin and Martin [9] who characterize state-dependent random choice with a restriction called *No Improving Action Switches (NIAS)*. We adapt NIAS to our setting below and the proposition below relates NIAS to belief calibration.

Definition. Let ρ be represented by (μ, u). Then ρ *satisfies NIAS* iff for all $g \in F \in \mathcal{K}$,

$$\sum_{s \in S} r_s \rho_{s,F}(f)\, u(f(s)) \geq \sum_{s \in S} r_s \rho_{s,F}(f)\, u(g(s))$$

Proposition 1. *Let ρ be represented by (μ, u). Then μ is well-calibrated iff ρ satisfies NIAS.*

Proof. See Appendix. □

Under state-dependent information representations, using test functions to calibrate beliefs is equivalent to checking for NIAS. Hence, the application of NIAS extends to our model.

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. Revealed preference, rational inattention, and costly information acquisition. Mimeo, 2014.

[9] A. Caplin and D. Martin. A testable theory of imperfect perception. *Economic Journal*, Forthcoming.

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

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

[15] D. Dillenberger, J. Lleras, P. Sadowski, and N. Takeoka. A theory of subjective learning. *Journal of Economic Theory*, 153:287–312, 2014.

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

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

[18] 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.

[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] D. Fudenberg and T. Strzalecki. Dynamic logit with choice aversion. *Econometrica*. Forthcoming.

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

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

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

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

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

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

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

[28] E. Kamenica and M. Gentzkow. Bayesian persuasion. *The American Economic Review*, 101(6):2590–2615, 2011.

[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. Preliminaries and identification

Given a non-empty collection $\mathcal{G}$ of subsets of H and some $F \in \mathcal{K}$, define

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

If $\mathcal{G}$ is a σ-algebra, then $\mathcal{G} \cap F$ is the *trace* of $\mathcal{G}$ on $F \in \mathcal{K}$. For $G \subset F \in \mathcal{K}$, define the smallest $\mathcal{H}_F$-measurable set containing G as follows

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

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

$\square$

Let ρ be an 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, 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. $\square$

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 $h^i \in F$ for $i \in \{1, \dots, k\}$ such that $\left\{h_F^1, \dots h_F^k\right\}$ forms a unique 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

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

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

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

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

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

□

For $\{F, F'\} \subset \mathcal{K}$, we use the condensed notation $FaF' := aF + (1-a)F'$. Let $H_c \subset H$ denote the set of all constant acts.

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 $\rho_F(f) = \rho_{F'}(f')$ and $f'_{F'} = f_F ah$.*

Proof. Note that $\rho_F(f) = \rho_{F'}(f')$ 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, $\rho(f', g') = \rho(g', f') = 1$ 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, $f'_{F'} = f_F ah$. □

Lemma (A5)**.** *Let ρ be represented by (μ, u). Then for any measurable $\phi : \mathbb{R} \to \mathbb{R}$,*

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

Proof. For $F \in \mathcal{K}$, let $\psi_F : \Delta S \to [0,1]$ be such that $\psi_F(q) = \frac{u\left(\overline{f}\right) - \sup_{f\in F} q\cdot(u\circ f)}{u\left(\overline{f}\right) - u\left(\underline{f}\right)}$ 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, define the following.

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

Lemma (A6)**.** *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 $f \in H_c$

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

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

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

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. □

Theorem (A7)**.** *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) $\rho\left(f, g\right) = \tau\left(f, g\right)$ *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 $\rho_F\left(f\right) = \tau_F\left(f\right)$ for all $f \in H$ from the representation. Moreover, since $\rho\left(f, g\right) = \rho\left(g, f\right) = 1$ iff $\tau\left(f, g\right) = \tau\left(g, f\right) = 1$ iff f and g are tied, the partitions $\{f_F\}_{f \in F}$ agree under both ρ and τ. Thus, $\mathcal{H}^\rho_F = \mathcal{H}^\tau_F$ 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 A6, 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,

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

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 A5,

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

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. □

B. Evaluating option sets

In this section, consider RCRs ρ such that there are $\{\underline{f}, \bar{f}\} \subset H_c$ where $\rho(\bar{f}, f) = \rho(f, \underline{f}) = 1$ 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 $f^a := \underline{f} a \bar{f}$.

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

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

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 $\rho_G\left(f^a\right) = 0$ 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

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

By monotonicity, $\rho_{F \cup f^a}(F) \geq \rho_G(F)$ and $\rho_{F \cup f^a}\left(f^a\right) \geq \rho_G\left(f^a\right)$ 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 A7, we can 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 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, $f^b\left(a\right)=1$ for all $\{a,b\}\subset[0,1]$. Hence $V_\rho\left(f^b\right)=V_\rho\left(f^a\right)$ so $V\left(f^b\right)=V\left(f^a\right)$ 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

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

which implies $\underline{f}\preceq_\rho f\preceq_\rho\overline{f}$. Thus, $V\left(\underline{f}\right)\leq V\left(f\right)\leq V\left(\overline{f}\right)$ for all $f\in H$ so $u\left(\underline{f}\right)\leq u\left(f\right)\leq u\left(\overline{f}\right)$ 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 $V_\rho\left(F\right)=\phi\left(V\left(F\right)\right)$ for some monotonic transformation $\phi:\mathbb{R}\to\mathbb{R}$. Now, for $b\in[0,1]$,

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

so $\phi\left(a\right)=a$ for all $a\in[0,1]$. Now, by Lemmas A5 and B1,

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

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

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

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

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 $F \in \mathcal{K}$. $\square$

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 A5 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.

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

$\square$

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 $\rho\left(f_a,F\right)=\frac{dW(F\cup f_a)}{da}$ for $f_a:=a\overline{f}+(1-a)\underline{f}$. 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 A5 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). $\square$

C. 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 $q \in \Delta S$.[38] Let $\phi : \mathbb{R} \to \mathbb{R}$ be increasing concave, and note that by Lemma A5,

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

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

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

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

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

$F \in \mathcal{K}$. Thus, if we let $\phi(x) = x$, then

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

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 . $\square$

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 A5,

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

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

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

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 A6, $u = \alpha v + \beta$ 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. $\square$

D. Calibrating beliefs

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

(1) $q_s > 0$ μ_s*-a.s..*

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

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

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

(1) Note that

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

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 $\lambda^F_s := \mu_s \circ \left(\psi^s_F\right)^{-1}$ 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^a_s is tied with nothing in F. Thus,

$$\begin{aligned} \lambda^F_s[0, r_s a] = \mu_s \circ \left(\psi^s_F\right)^{-1}[0, r_s a] &= \mu_s\left\{q \in \Delta S | \, r_s a \geq \psi^s_F(q)\right\} \\ &= \mu_s\left\{q \in \Delta S | \sup_{f \in F} q \cdot (u \circ f) \geq 1 - a q_s\right\} \\ &= \mu_s\left\{q \in \Delta S | \sup_{f \in F} q \cdot (u \circ f) \geq q \cdot \left(u \circ f^a_s\right)\right\} = \rho_s\left(F, f^a_s\right) = F^s_\rho(r_s a) \end{aligned}$$

Now, if f^a_s is tied with some $g \in F$, then

$$F^s_\rho(r_s a) = \rho_s\left(F, f^s_a\right) = 1 = \mu_s\left\{q \in \Delta S | \sup_{f \in F} q \cdot (u \circ f) \geq q \cdot \left(u \circ f^s_a\right)\right\} = \lambda^F_s[0, r_s a]$$

Thus, $\lambda^F_s[0, r_s a] = F^s_\rho(r_s a)$ for all $a \in [0,1]$. Since $F \in \mathcal{K}_s$,

$$1 = F^s_\rho(r_s) = \lambda^F_s[0, r_s]$$

so F^s_ρ is the cumulative distribution function of λ^F_s.

$\square$

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

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

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

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

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.

$\square$

Theorem (D3)**.** *Let ρ be represented by (μ, u). If $F^s_\rho =_m F_{\bar{\rho}}$, 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 $F^s_\rho =_m F_{\bar{\rho}}$ and by Lemmas D1 and D2, we have

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

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

Let $G \in \mathcal{K}$ and $F_a := \left(Ga\overline{f}\right)\cup \underline{f}^s$ for $a \in (0,1)$. Since $\underline{f}^s \in F_a$, $\rho_s\left(F_a, \underline{f}^s\right) = 1$ so $F_a \in \mathcal{K}_s$. Let

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

and note that

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

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

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

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

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

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

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

For $q \in Q_a^c$,

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

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

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

By a symmetric argument for ν_s, we have

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

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

$$\int_{\Delta S} \sup_{f \in G} q \cdot \left(u \circ f\right) \bar{\mu}\left(dq\right) = \int_{\Delta S} \sup_{f \in G} q \cdot \left(u \circ f\right) \nu_s\left(dq\right)$$

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,

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

for all $Q \in \mathcal{B}\left(\Delta S\right)$ 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 $\mu\left(Q_0\right) = 1$. Now,

$$\begin{aligned}
\sum_{s \in S_+} r_s &= \sum_{s \in S_+} \int_{\Delta S} q_s \bar{\mu}\left(dq\right) = \int_{Q_0} \sum_{s \in S_+} q_s \bar{\mu}\left(dq\right) \\
&= \int_{Q_0} \left(\sum_{s \in S} q_s \right) \bar{\mu}\left(dq\right) = \bar{\mu}\left(Q_0\right) = 1
\end{aligned}$$

which implies $\sum_{s \notin S_+} r_s = 0$ a contradiction. Thus, $S_+ = S$ and μ is well-calibrated. $\square$

Theorem (D4)**.** *Let ρ be represented by (μ, u). If μ is well-calibrated, then $F^s_\rho =_m F_{\bar{\rho}}$.*

Proof. Note that the measurable sets and ties of ρ_s and $\bar{\rho}$ coincide by definition. As above, let $S_+ := \left\{ s \in S | \, \rho_s \left(\underline{f}^s, \overline{f} \right) = 0 \right\} \subset S$. 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 $Q_0 := \left\{ q \in \Delta S | \sum_{s \notin S_+} q_s = 0 \right\}$ yields

$$\sum_{s \in S_+} r_s = \sum_{s \in S_+} \int_{\Delta S} q_s \bar{\mu} \left(dq \right) = \int_{Q_0} \left(\sum_{s \in S} q_s \right) \bar{\mu} \left(dq \right) = 1$$

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 A5 and D1 and the fact that μ is well-calibrated,

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

so $F^s_\rho =_m F_{\bar{\rho}}$. $\square$

Proposition (D5)**.** *Let ρ be represented by (μ, u). Then μ is well-calibrated iff ρ satisfies NIAS.*

Proof. Let ρ be represented by (μ, u). First, suppose μ is well-calibrated. For $f \in F \in \mathcal{K}$, define

$$Q^f_F := \{ q \in \Delta S \; | \; q \cdot (u \circ f) \geq q \cdot (u \circ g) \;\; \forall g \in F \}$$

so $\rho_{s,F}(f) = \mu_s\left(Q_F^f\right)$ for all $s \in S$ and $\bar{\rho}_F(f) = \bar{\mu}\left(Q_F^f\right)$. Hence for any $g \in F$,

$$\begin{aligned}
\int_{Q_F^f} q \cdot (u \circ f)\,\bar{\mu}(dq) &\geq \int_{Q_F^f} q \cdot (u \circ g)\,\bar{\mu}(dq) \\
\sum_s r_s \int_{Q_F^f} \frac{q_s}{r_s} u(f(s))\,\bar{\mu}(dq) &\geq \sum_s r_s \int_{Q_F^f} \frac{q_s}{r_s} u(g(s))\,\bar{\mu}(dq) \\
\sum_s r_s \int_{Q_F^f} \mu_s(dq)\, u(f(s)) &\geq \sum_s r_s \int_{Q_F^f} \mu_s(dq)\, u(g(s)) \\
\sum_s r_s \rho_{s,F}(f)\, u(f(s)) &\geq \sum_s r_s \rho_{s,F}(f)\, u(g(s))
\end{aligned}$$

so ρ satisfies NIAS.

Now, suppose ρ satisfies NIAS so for all $g \in F \in \mathcal{K}$ and $\bar{\rho}_F(f) > 0$,

$$\begin{aligned}
\sum_s r_s \rho_{s,F}(f)\, u(f(s)) &\geq \sum_s r_s \rho_{s,F}(f)\, u(g(s)) \\
\sum_s \frac{r_s \mu_s\left(Q_F^f\right)}{\bar{\mu}\left(Q_F^f\right)} u(f(s)) &\geq \sum_s \frac{r_s \mu_s\left(Q_F^f\right)}{\bar{\mu}\left(Q_F^f\right)} u(g(s)) \\
q_F(f) \cdot (u \circ f) &\geq q_F(f) \cdot (u \circ g)
\end{aligned}$$

where $q_F : F \to \Delta S$ is such that $q_F(f)(s) := \frac{r_s \mu_s\left(Q_F^f\right)}{\bar{\mu}\left(Q_F^f\right)}$. Hence, $q_F(f) \in Q_F^f$ for all $f \in F$. For each $s \in S$, define the measure $\nu_s(Q) := \int_Q \frac{q_s}{r_s} \bar{\mu}(dq)$ and note that $\sum_s r_s \nu_s(Q) = \bar{\mu}(Q) = \sum_s r_s \mu_s(Q)$. Now, we also have for all $g \in F$,

$$\begin{aligned}
\int_{Q_F^f} q \cdot (u \circ f)\,\bar{\mu}(dq) &\geq \int_{Q_F^f} q \cdot (u \circ g)\,\bar{\mu}(dq) \\
\sum_s r_s \int_{Q_F^f} \frac{q_s}{r_s} u(f(s))\,\bar{\mu}(dq) &\geq \sum_s r_s \int_{Q_F^f} \frac{q_s}{r_s} u(g(s))\,\bar{\mu}(dq) \\
\sum_s \frac{r_s \nu_s\left(Q_F^f\right)}{\bar{\mu}\left(Q_F^f\right)} u(f(s)) &\geq \sum_s \frac{r_s \nu_s\left(Q_F^f\right)}{\bar{\mu}\left(Q_F^f\right)} u(g(s)) \\
p_F(f) \cdot (u \circ f) &\geq p_F(f) \cdot (u \circ g)
\end{aligned}$$

where $p_F : F \to \Delta S$ is such that $p_F(f)(s) := \frac{r_s \nu_s\left(Q_F^f\right)}{\bar{\mu}\left(Q_F^f\right)}$. Hence, $p_F(f) \in Q_F^f$ for all $f \in F$. Consider a partition $\mathcal{P}^n$ of ΔS such that for every $P_i^n \in \mathcal{P}^n$, $P_i^n = Q_F^f$ for some $f \in F$ and

$\sup_{\{p,q\}\subset P_i} |p-q| \leq \frac{1}{n}$ for every $i \in \{1,\ldots,n\}$. Since $\{p_F(f), q_F(f)\} \subset P_i^n$,

$$\left| \frac{r_s \mu_s (P_i^n)}{\bar{\mu}(P_i^n)} - \frac{r_s \nu_s (P_i^n)}{\bar{\mu}(P_i^n)} \right| \leq \frac{1}{n}$$

$$|\mu_s (P_i^n) - \nu_s (P_i^n)| \leq \bar{\mu}(P_i^n) \frac{1}{n r_s}$$

Now, for any $\psi^n : \Delta S \to \mathbb{R}$ that is $\mathcal{P}^n$-measurable, we have

$$\sum_i \nu_s (P_i^n) \psi_i^n - \frac{1}{n} \sum_i \frac{1}{r_s} \bar{\mu}(P_i^n) \psi_i^n \leq \sum_i \mu_s (P_i^n) \psi_i^n \leq \sum_i \nu_s (P_i^n) \psi_i^n + \frac{1}{n} \sum_i \frac{1}{r_s} \bar{\mu}(P_i^n) \psi_i^n$$

For any measurable $\psi : \Delta S \to \mathbb{R}$, we can find a sequence of $\mathcal{P}^n$-measurable functions such that $\psi^n \to \psi$. Hence by dominated convergence,

$$\begin{aligned} \lim_{n\to\infty} \left[\int_{\Delta S} \psi^n (q) \nu_s (dq) - \frac{1}{n} \int_{\Delta S} \frac{1}{r_s} \psi^n (q) \bar{\mu}(dq) \right] &\leq \lim_{n\to\infty} \int_{\Delta S} \psi^n (q) \mu_s (dq) \\ &\leq \lim_{n\to\infty} \left[\int_{\Delta S} \psi^n (q) \nu_s (dq) + \frac{1}{n} \int_{\Delta S} \frac{1}{r_s} \psi^n (q) \bar{\mu}(dq) \right] \\ \int_{\Delta S} \psi (q) \mu_s (dq) &= \int_{\Delta S} \psi (q) \nu_s (dq) \end{aligned}$$

Hence, $\mu_s = \nu_s$ and μ is well-calibrated. □

E. Relation to AS

In this section, we relate our results to that of AS. We focus on the individual interpretation for ease of comparison. AS introduces a condition called consequentialism to link choice behavior from the two time periods.[39] 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 4. 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 $F^+ \subset F \cap G$

[39] Their second axiom deals with indifferences which we resolve using non-measurability.

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, 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

$$F_\rho = F^+_\rho = G^+_\rho = G_\rho$$

$\square$

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

$$\begin{aligned} u(f) &= a u\left(\underline{f}\right) + (1-a) u\left(\overline{f}\right) = (1 - v(f)) u\left(\underline{f}\right) + v(f) u\left(\overline{f}\right) \\ &= \left(u\left(\overline{f}\right) - u\left(\underline{f}\right)\right) v(f) + u\left(\underline{f}\right) \end{aligned}$$

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. $\square$

Supplementary Appendix

In this supplementary appendix, we provide an axiomatic treatment of our model. We first introduce a more general representation. 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.. The corresponding regularity condition on π is as follows.

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

We now define a random subjective expected utility (RSEU) representation.

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 $\pi(\Delta S \times \{u\}) = 1$, this reduces to an information representation where μ is the marginal distribution of π on ΔS.

We now introduce the axioms. The first four are familiar restrictions on RCRs. Let ext$F$ denote the set of extreme acts of $F \in \mathcal{K}$.[40] Recall that $\mathcal{K}_0$ is the set of decision-problems where every act in the decision-problem is measurable with respect to the RCR.

Axiom 1 (Monotonicity)**.** *ρ is monotone*

Axiom 2 (Linearity)**.** *ρ is linear*

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

Axiom 4 (Continuity)**.** *ρ is continuous on $\mathcal{K}_0$.*

Monotonicity follows from the fact that when decision-problems are enlarged, new acts could become dominant so the probability of choosing old acts can only decrease. Linearity and extremeness follow from the fact that the random utilities in our model are linear (i.e.

[40] 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 $a \in (0,1)$.

agents are subjective expected utility maximizers). In fact, linearity is *the* version of the independence axiom tested in many experimental settings (see Kahneman and Tversky [27] for example). Extremeness follows from the fact that when linear utilities are used for evaluation, mixtures of acts in a decision-problem are never chosen (aside from indifferences). Note that it rules out behaviors associated with random non-linear utilities (such as ambiguity aversion).

For continuity, note that if $\mathcal{H}$ is the Borel algebra, then $\mathcal{K}_0 = \mathcal{K}$ and our continuity axiom condenses to the usual continuity. In general though, the RCR is not continuous over all decision-problems and is in fact discontinuous at precisely those decision-problems that contain indifferences.[41]

These first four axioms are necessary and sufficient for random expected utility (Gul and Pesendorfer [24]). We now present new axioms. For any act f and states s_1 and s_2, define $f_{s_1}^{s_2}$ 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 5 (S-independence)**.** $\rho_F\left(f_{s_1}^{s_2} \cup f_{s_2}^{s_1}\right) = 1$ *for* $F = \left\{f, f_{s_1}^{s_2}, f_{s_2}^{s_1}\right\}$.

S-independence states that if two acts are constant over two states and each coincides with a third act in each of the two states, then the first two will be chosen for sure over the third provided all three coincide on all other states. This follows from the fact that an agent will "hedge" by choosing the non-constant third act only when her taste utility is state-dependent. This is the random choice version of the state-by-state independence axiom.[42] Finally, non-degeneracy rules out the trivial case of universal indifference.

Axiom 6 (Non-degeneracy)**.** $\rho_F(f) < 1$ *for some* F *and* $f \in F$.

Axioms 1-6 are necessary and sufficient for a RSEU representation (Theorems S5). This demonstrates that the 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. For an information representation, we need one additional restriction.

---

[41] Every decision-problem is in fact arbitrarily (Hausdorff) close to some decision-problem in $\mathcal{K}_0$, so continuity is preserved over almost all decision-problems.

[42] 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 S5 implies that this is equivalent to state-by-state independence axiom in the presence of the other standard axioms.

Axiom 7 (S-determinism)**.** *$\rho_F(f) \in \{0,1\}$ if $f(s) = g(s)$ for all $s \neq s'$ and $g \in F$.*

S-determinism states that the RCR is deterministic over decision-problems consisting of acts that differ only on a single state. This is because in an information representation, choice is stochastic only as a result of varying beliefs. For acts that only vary in one state but are otherwise the same, only payoffs in that state matter so choice must be deterministic.

Taken together, Axioms 1-7 are necessary and sufficient for an information representation (Corollary S7). Note that if we allow the utility u to be constant, then the non-degeneracy axiom can be dropped without loss of generality. However, the uniqueness of μ in the representation would obviously fail in Theorem 1.

Our treatment so far assumes beliefs are completely subjective. Finally, similar to Section 7, we present an axiomatization that includes as a primitive the observed frequency of states $r \in \Delta S$ (assume r has full support as before). This provides us with additional restrictions on the RCR and also allows us to address state-dependent utilities.[43] Let $u := (u_s)_{s \in S}$ be a collection of state-dependent utilities with at least one u_s non-constant. Call μ *well-calibrated* iff $r = \int_{\Delta S} q \; \mu(dq)$.

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

$$\rho_F(f) = \mu \left\{ q \in \Delta S \;\middle|\; \sum_{s \in S} q_s u_s(f) \geq \sum_{s \in S} q_s u_s(g) \;\; \forall g \in F \right\}$$

Since a calibrated information representation allows for state-dependent utilities, S-independence must be relaxed. We now introduce a consistency axiom that relates r with RCR. Define a *conditional best act* $\bar{f}^s$ as the act that coincides with the best act if s occurs and with the worst act otherwise. Let $\bar{f}^s_\rho$ be the test function of $\bar{f}^s$ under ρ.[44]

Axiom 8 (Consistency)**.** *The mean of $\bar{f}^s_\rho$ is r_s for all $s \in S$.*

Axioms 1-4, 6-8 are necessary and sufficient for a calibrated information representation (Theorem S8). Moreover, calculating means of test functions allows an analyst to check if the RCR is consistent with the objective frequency r.

We now present the proofs. 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

[43] This is analogous to that of Karni, Schmeidler and Vind [30] under deterministic choice.

[44] Axioms 1-4, 6 and 7 ensure that a best and worst act exist and that test functions are well-defined.

$z_i \cdot z_j = 0$ for all $i \neq j$, where $z_i := \frac{f_i - g_i}{\|f_i - g_i\|}$. Let $Z := \text{lin}\{z_1, \ldots, z_k\}$ 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

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

and let $W := \text{lin}(\varphi(H))$. Lemma S1 below shows that φ projects H onto a space without ties.

Lemma (S1)**.** *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 $W = \mathrm{lin}\left(\varphi\left(H\right)\right)$ and φ is linear, $w \cdot z_i = 0$ for all $w \in W$. Thus,

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

for all $w \in W$.

$\square$

Lemma (S2)**.** *If ρ satisfies Axioms 1-4, then there exists a measure ν on W such that*

$$\rho_F\left(f\right) = \nu\left\{w \in W \middle| w \cdot f \geq w \cdot g \ \forall g \in F\right\}$$

Proof. Let $m := \dim\left(W\right)$. Note that if $m = 0$, then $W = \varphi\left(H\right)$ is a singleton so everything is tied by Lemma S1 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. Let $V := \mathrm{lin}\left(\Delta^m - \mathbf{1}\right)$ so there exists an linear transformation $A : W \to V$ corresponding to an orthogonal matrix (i.e. $A^{-1} = A'$) where $v = Aw$. Define $T\left(w\right) = \lambda A w + \beta \mathbf{1}$ such that $T \circ \varphi\left(H\right) \subset \Delta^m$ where $\lambda > 0$ and $\beta > 0$. Now, for each finite set $D \subset \Delta^m$, we can find a $p^* \in \Delta^m$ and $a \in (0,1)$ such that $Dap^* \subset T \circ \varphi\left(H\right)$. Thus, we can define an RCR τ on Δ^m such that

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

where $T \circ \varphi\left(F\right) = Dap^*$ and $T \circ \varphi\left(f\right) = pap^*$. Linearity and Lemma S1 ensure that τ is well-defined.

Since the projection mapping φ and T are both affine, Axioms 1-4 correspond exactly to the axioms of Gul and Pesendorfer [24] on Δ^m. Thus, by their Theorem 3, there exists a measure ν_T on V such that for $F \in \mathcal{K}_0$

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

Since $v \cdot \mathbf{1} = 0$ and A is orthogonal,

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

Hence

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

where $\nu := \nu_T \circ A$ is the measure on W induced by A. Note that the last equality follows from Lemma S1.

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

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

By Lemma S1, 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,

$$\rho_F(f) = \nu\left\{w \in W \middle| w\cdot f \geq w\cdot g\ \forall g \in F\right\}$$

□

Henceforth, assume ρ satisfies Axioms 1-4 and let ν be the measure on W as specified by Lemma S2. 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 $\alpha > 0$. We also employ the notation $\rho(F,G) := \rho_{F\cup G}(F)$. 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 (S3)**.** *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. $\square$

Lemma (S4)**.** *Let ρ satisfy Axioms 1-5. 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 $A_{ij} \nearrow A_i \cup \{0\}$, $A_{ij} \subset \operatorname{int} \left(\bar{A}_{ij} \right) \cup \{0\}$ 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)

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

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 $A_{ij} \setminus \{0\} \subset \operatorname{int} \left(\bar{A}_{ij} \right)$, we must have $u_1 \cdot r > 0 > u_2 \cdot r$ for all $(u_1, u_2) \in (A_{ij} \setminus \{0\}) \times (A_{i'j} \setminus \{0\})$. By (1) and (2),

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

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

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

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

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 $(u_1, u_2) \in (\bigcap_k C_k) \backslash C_0$. 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

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

□

Theorem (S5.1)**.** *If ρ satisfies Axioms 1-6, then it has a RSEU representation.*

Proof. Let ρ satisfy Axioms 1-6, and ν be the measure on W as specified by Lemma S2. Let $S^* \subset S$ be the set of non-null states with some $s^* \in S^*$ as guaranteed by Lemma S3. Define

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

and note that by Lemma S4,

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

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 h 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

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

Note that Lemma S4 implies that u is non-constant. Finally, we show that π is regular. Suppose there are $\{f, g\} \subset H$ 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 (S5.2)**.** *If ρ has a RSEU representation, then it satisfies Axioms 1-6.*

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, $\mu\left(\bar{Q}\right)=0$ 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 $B\in\mathcal{B}(H)$,

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

Hence, ξ_k and ξ are random variables. Moreover,

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

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. $\square$

Lemma (S6). *ρ satisfy Axioms 1-4, 6 and 7 iff there exists a regular μ and $u = (u_s)_{s\in S}$ with at least one u_s non-constant such that*

$$\rho_F(f) = \mu\left\{ q \in \Delta S \;\middle|\; \sum_s q_s u_s(f(s)) \geq \sum_s q_s u_s(g(s)) \;\; \forall g \in F \right\}$$

Proof. We first prove sufficiency. Since ρ satisfies Axioms 1-4, by Lemma S2, there exists a measure ν on W such that

$$\rho_F(f) = \nu\left\{ w \in W \;\middle|\; \sum_s w_s(f(s)) \geq \sum_s w_s(g(s)) \;\; \forall g \in F \right\}$$

Fix some $s \in S$ and $y \in X$. For each $x \in X$, let $f^x \in H$ be such that $f^x(s) = \delta_x$ and $f^x(s') = \delta_y$ for all $s' \neq s$. Let $F := \bigcup_{x\in X} f^x$ so by S-determinism,

$$\rho_F(f^y) = \nu\{ w \in W \mid w_s(y) \geq w_s(x) \;\; \forall x \in X \} \in \{0, 1\}$$

Hence, we can find some $x_1^s \in X$ such that $w_s(x_1^s) \geq w_s(x)$ ν-a.s. for all $x \in X$ and by iteration some $x_2^s \in X$ such that $w_s(x) \geq w_s(x_2^s)$ ν-a.s. for all $x \in X$.

For each $w \in W$ and $s \in S$, let $u_s^w \in \mathbb{R}^X$ be such that $u_s^w(x_1^s) = 1$, $u_s^w(x_2^s) = 0$ and $w_s = \alpha_s(w) u_s^w + \beta_s(w)\mathbf{1}$ for $\alpha_s(w) \geq 0$. Hence,

$$\rho_F(f) = \nu\left\{ w \in W \;\middle|\; \sum_s \alpha_s(w) u_s^w(f(s)) \geq \sum_s \alpha_s(w) u_s^w(g(s)) \;\; \forall g \in F \right\}$$

By Lemma S3, there exists some non-null $s^* \in S$. Hence, by the definition of nullity and S-determinism, $w_{s^*}\left(x_1^{s^*}\right) > w_{s^*}\left(x_2^{s^*}\right)$ ν-a.s. so $\sum_s \alpha_s(w) > 0$ ν-a.s.. Define $Q : W \to \Delta S$ such that $Q_s(w) := \frac{\alpha_s(w)}{\sum_s \alpha_s(w)}$ for every $s \in S$ so

$$\rho_F(f) = \nu\left\{ w \in W \;\middle|\; \sum_s Q_s(w) u_s^w(f(s)) \geq \sum_s Q_s(w) u_s^w(g(s)) \;\; \forall g \in F \right\}$$

Since $\rho\left(f^{x_1^s}, f^x\right) = \rho\left(f^x, f^{x_2^s}\right) = 1$, by S-determinism and continuity, for every non-null $s \in S$ and $x \in X$, there is a unique $a^x \in [0,1]$ such that f^x is tied with $f^{x_1^s} a^x f^{x_2^s}$. Hence,

$$1 = \rho\left(f^{x_1^s} a^x f^{x_2^s}, f^x\right) = \rho\left(f^x, f^{x_1^s} a^x f^{x_2^s}\right) = \nu\{ w \in W \mid u_s^w(x) = a^x \}$$

so u_s is fixed ν-a.s.. Thus

$$\rho_F(f) = \mu\left\{q \in \Delta S \;\middle|\; \sum_s q_s u_s(f(s)) \geq \sum_s q_s u_s(g(s)) \;\; \forall g \in F\right\}$$

where $\mu := \nu \circ Q^{-1}$ is the induced measure on ΔS by Q. The argument for the regularity of μ follows by the same reasoning as in Theorem S5.1. For necessity, note that Axioms 1-4 and 6 all follow as in Theorem S5.2. S-determinism follows trivially from the representation. □

Corollary (S7)**.** *ρ satisfies Axioms 1-7 iff it has an information representation.*

Proof. Since necessity follows immediately from Theorem S5.2 and Lemma S6, we prove sufficiency. By Theorem S5.1, ρ has a RSEU representation. Fix some non-null $s^* \in S$ and $y \in X$. For each $x \in X$, let $f^x \in H$ be such that $f^x(s^*) = \delta_x$ and $f^x(x) = \delta_y$ for all $s \neq s^*$. Let $F := \bigcup_{x \in X} f^x$ so by S-determinism,

$$\rho_F(f^x) = \pi\left(\Delta S \times \left\{u \in \mathbb{R}^X \mid u(x) \geq u(z) \;\; \forall z \in X\right\}\right) \in \{0, 1\}$$

Hence, we can find some $\bar{x} \in X$ such that $u(\bar{x}) \geq u(x)$ π-a.s. for all $x \in X$ and by iteration some $\underline{x} \in X$ such that $u(x) \geq u(\underline{x})$ π-a.s. for all $x \in X$. Note that $u(\bar{x}) > u(\underline{x})$ by non-degeneracy. Normalize $u(\bar{x}) = 1$ and $u(\underline{x}) = 0$ without loss of generality and let $\bar{f} := f^{\bar{x}}$ and $\underline{f} := f^{\underline{x}}$. Since $\rho\left(\bar{f}, f^x\right) = \rho\left(f^x, f^{\underline{x}}\right) = 1$, by S-determinism and continuity, there is a unique $a^x \in [0, 1]$ such that f^x is tied with $\bar{f} a^x \underline{f}$. Hence,

$$1 = \rho\left(\bar{f} a^x \underline{f}, f^x\right) = \rho\left(f^x, \bar{f} a^x \underline{f}\right) = \pi\left(\Delta S \times \left\{u \in \mathbb{R}^X \mid u(x) = a^x\right\}\right)$$

so u is fixed π-a.s.. Thus $\pi(\Delta S \times \{u\}) = 1$ and ρ has an information representation. □

Theorem (S8)**.** *ρ satisfies Axioms 1-4, 6-8 iff it has a calibrated information representation.*

Proof. Let ρ satisfy Axioms 1-4, 6-8. By Lemma S6, there is a regular μ and $u = (u_s)_{s \in S}$ with at least one u_s non-constant such that

$$\rho_F(f) = \mu\left\{q \in \Delta S \;\middle|\; \sum_s q_s u_s(f(s)) \geq \sum_s q_s u_s(g(s)) \;\; \forall g \in F\right\}$$

Let $\left\{\bar{f}, \underline{f}\right\} \subset H$ be such that $\bar{f}(s) = \delta_{x_1^s}$ and $\underline{f}(s) = \delta_{x_2^s}$ where $\{x_1^s, x_2^s\} \subset X$ are specified by Lemma S6. Since $\sum_s q_s u_s\left(\bar{f}(s)\right) = 1$ and $\sum_s q_s u_s\left(\underline{f}(s)\right) = 0$, $\bar{f}$ and $\underline{f}$ are the best and worst acts respectively.

Let $\psi_f : \Delta S \to [0,1]$ be such that $\psi_f(q) = 1 - \sum_s q_s u_s(f)$ 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]),

$$\int_{[0,1]} x \lambda^f(dx) = \int_{\Delta S} \left(1 - \sum_s q_s u_s(f(s))\right) \mu(dq)$$

For $a \in [0,1]$, let $f^a := \underline{f} a \overline{f}$. Now,

$$\begin{aligned}\lambda^f[0,a] &= \mu \circ \psi_f^{-1}[0,a] = \mu\{q \in \Delta S \mid a \geq \psi_f(q) \geq 0\} \\ &= \mu\left\{q \in \Delta S \;\middle|\; \sum_s q_s u_s(f(s)) \geq \sum_s q_s u_s(f^a(s))\right\} = f_\rho(a)\end{aligned}$$

so the cumulative distribution function of λ^f is exactly f_ρ. Lemma B1 yields

$$\int_{[0,1]} f_\rho(a)\, da = \int_{\Delta S} \sum_s q_s u_s(f(s))\, \mu(dq)$$

By Axiom 8,

$$r_s = \int_{[0,1]} \bar{f}_\rho^s(a)\, da = \int_{\Delta S} q_s \mu(dq)$$

as desired. Necessity follows immediately from Lemma S6.