ESSAYS ON BELIEFS AND BEHAVIOR

JAY LU

A DISSERTATION

PRESENTED TO THE FACULTY

OF PRINCETON UNIVERSITY

IN CANDIDACY FOR THE DEGREE

OF DOCTOR OF PHILOSOPHY

RECOMMENDED FOR ACCEPTANCE

BY THE DEPARTMENT OF ECONOMICS

ADVISER: FARUK GUL

JUNE 2014

Abstract

This dissertation is a theoretical study of the role of beliefs in both individual and group behavior. In Chapter 1, we consider a decision-maker 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 decision-maker has better information by using choice dispersion as a measure of informativeness, (3) determine if the decision-maker's beliefs about information are dynamically consistent, and (4) test to see if these beliefs are well-calibrated or rational.

In Chapter 2, we dispense with the standard assumption of expected utility maximization and introduce a theory of stochastic ambiguity aversion. In the individual interpretation of this theory, choice is random due to unobservable shocks to the individual's ambiguity aversion. In the group interpretation of this theory, choice is random due to unobservable heterogeneity in ambiguity aversion within the group. A one-parameter distribution characterizing stochastic ambiguity aversion can be fully identified from random choice. From a technical standpoint, we offer decision theoretic foundations for relaxing linearity under random utility maximization.

In Chapter 3, we study a model of competitive trading where agents have heterogeneous beliefs about the persistence of states. This model addresses a robust finding in behavioral finance known as the disposition effect where agents over-purchase stocks after prices fall and over-sell stocks after prices rise. We show that agents who believe in the least persistence exhibit the disposition effect while those who believe in the most persistence engage in momentum trading. Moreover, agents can be ordered by how much of a disposition effect they exhibit if and only if their beliefs can be ordered by a single parameter measuring persistence. This allows for identification of beliefs even when equilibrium considerations restrict the observable choice data.

Acknowledgements

I owe an enormous debt to Faruk Gul and Wolfgang Pesendorfer for their exceptional advice, generous guidance and continuous encouragement. I knew nothing of choice theory before taking their classes but can now count myself a young researcher in the field thanks to them.

I am truly grateful to Stephen Morris for some highly valuable and insightful discussions. His generosity and insight have benefited me greatly. I would also like to thank Roland Bénabou, Simone Cerreia-Vioglio, Sylvain Chassang, Mark Dean, Pietro Ortoleva and Wei Xiong for the many helpful conversations on various chapters of this dissertation.

I thank the seminar audiences at Yale University, University of California Los Angeles, Northwestern University, Cornell University, Massachusetts Institute of Technology, University of British Columbia, Boston University and the University of Pennsylvania. It was a privilege to present my first chapter at these institutions, and I appreciate all the helpful feedback I received during those visits.

To my friends and peers Nemanja Antic, Benjamin Brooks, Gonzalo Cisternas, Leandro Gorno, Diogo Guillén, Constantinos Kalfarentzsos, Rohit Lamba, Paulo Natenzon, Brian So, Kai Steverson and Doron Ravid: thank you for all those enjoyable chats in the basement of Fisher. They were great fun, and despite all the laughs, they did contribute to some serious research at the end of the day. I look forward to continuing them in the future.

Finally, I thank my parents for their constant care and for instilling in me a sense of scholarship at an early age. To Susan Chien, I owe a special thanks for her boundless support, infinite patience, and unwavering faith in my ability to survive as an academic.

Contents

1 Random Choice and Private Information

1.1 Introduction

Consider a decision-maker who at time 2, will choose an option from a set of available options. At time 1, the decision-maker receives private information that will affect this time-2 choice. An analyst (i.e. an outside observer) knows the decision-maker's set of options but does not know the decision-maker's information. Hence, to the analyst, the decision-maker's choice is probabilistic or *random*. Many problems in information economics and decision theory fit in this framework. In this chapter, we are interested in applications where the set of available options are uncertain prospects, the decision-maker's information reveals the likelihood of various outcomes, and this information is *strictly private*, that is, it is completely unobservable to the analyst.

To be concrete, suppose that the decision-maker is an individual agent choosing from a set of health insurance plans at the beginning of every year. Before choosing, the agent receives some private information (i.e. a signal) that influences her *beliefs* about her health for the rest of the year. For example, she may get a health exam that informs her about the likelihood of falling sick. As a result, her choice of an insurance plan each year depends on the realization of her signal that year. The analyst however, does not observe this signal. Hence, from the analyst's perspective, the individual's choice of health insurance every year is probabilistic or random. This is captured by a frequency distribution of choices over the years. We call this the *individual interpretation* of random choice.

Alternatively, suppose that the decision-maker is a group of agents choosing from the same set of health insurance plans in a single year. Before choosing, each agent in the group has some private information that influences her individual *beliefs* about her health. For example, she may have some personal knowledge about her lifestyle that affects her choice of insurance. This information however, is beyond what is captured by all the characteristics observable by the analyst. As a result, to the analyst, agents are observationally identical.[1]

[1] We can think of this group as the end result after applying all possible econometric (e.g. non-parametric)

Hence, from the analyst's perspective, the choice of health insurance within the group is probabilistic or random. This is captured by a frequency distribution of choices over all the agents in the group. We call this the *group interpretation* of random choice.

In both the individual interpretation (where the decision-maker is an individual agent) and the group interpretation (where the decision-maker is a group of agents), the main feature is that the decision-maker's private information is unobservable to the analyst. Other applications include consumers choosing a retail banking service, private investors deciding on an investment, buyers bidding at a Treasury auction or users clicking on an online ad. In all these examples, it is likely that an analyst is unable to directly observe the decision-maker's private information or has difficulty discerning how that information will be interpreted.

The first main contribution of this chapter is providing a theory in which private information can be fully identified from random choice.[2] Moreover, we perform the following inferences. First, by observing the decision-maker's random choice from a set of options, we can compute the set's valuation, that is, its ex-ante utility (in the individual interpretation) or its welfare (in the group interpretation). Call this *evaluating option sets*. Second, we can discern which decision-maker has better information by using choice dispersion as a measure of informativeness. Call this *assessing informativeness*. Third, if valuations of option sets are known, then we can compare them to the decision-maker's random choice to detect when beliefs about information are dynamically inconsistent. Call this *detecting biases*. Finally, from the joint distribution of choices and payoff-relevant outcomes, we can determine if beliefs are well-calibrated or rational. Call this *calibrating beliefs*.

When information is observable, the above inferences are important and well-understood exercises in information theory and information economics. Our second contribution is to provide the tools that allow us to carry out the same analysis even when information is

analysis available to differentiate the data.

[2] An alternative approach is to directly elicit private information from survey data (for example, see Finkelstein and McGarry [34] and Hendren [50]). However, respondents may not accurately report their true beliefs or the data may be subject to other complications (such as excess concentrations at focal points). In contrast, our approach follows the original spirit of Savage [73] by inferring beliefs from choice behavior.

not directly observable and can only be inferred from choice behavior. Our theorems reveal that when all relevant choice data is available, all these inferences can be performed just as effectively as in the case with observable information. The more practical question of drawing inferences when choice data is only partially available is left for future research.

Formally, let S represent an objective state space. For example, in the case of health insurance, S could represent the binary state space where an agent can either be healthy or sick. We call a set of options a *decision-problem* and each option in the set an *act*. An act specifies a payoff for each realization of the state. For example, forgoing health insurance corresponds to the act that yields a high payoff if the agent is healthy and a low payoff if she is sick. On the other hand, choosing a no-deductible (full-insurance) health plan corresponds to the act that yields the same payoff regardless of whether the agent is sick or not. At time 2, a decision-maker, either an individual agent or a group of agents, chooses some act from a decision-problem. In the individual interpretation, an agent chooses from the same decision-problem repeatedly every year which results in a frequency distribution of choices over all the years. In the group interpretation, all agents in the group choose from the same decision-problem in a single year which results in a frequency distribution of choices over all agents in the group. In either interpretation, we call the observable distribution of choices a *random choice rule (RCR).*

At time 1, an agent receives some private information that allows her to form beliefs about S. Since this information is unobservable to the analyst, beliefs are subjective. Each agent then evaluates the expected utility of every act in the decision-problem using her subjective belief. At time 2, she chooses the best act in the decision-problem. Since the private information of the decision-maker (either an individual agent or a group of agents) is unknown to the analyst, the decision-maker's time-2 choice is probabilistic and can be modeled as a RCR. We call this an *information representation* of the RCR.

An information representation is a model of random utility maximization (RUM).[3] In

[3] For more about random utility maximization, see Block and Marschak [12], Falmagne [31], McFadden and Richter [61] and Gul, Natenzon and Pesendorfer [46].

particular, the random utility is a subjective expected utility where the subjective beliefs depend on the agent's private information. In the individual interpretation, each realization of this random utility corresponds to a realization of the agent's private signal. In the group interpretation, each realization of this random utility corresponds to a realization of a random draw of an agent in the group. In both the individual and group interpretations, the probability that an act is chosen is equal to the probability that the act attains the highest subjective expected utility in the decision-problem.

In general, RUM models have difficulty dealing with indifferences in the random utility. We address this issue by drawing an analogy with standard deterministic (i.e. not random) 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 modelling approach has two advantages. First, it allows the analyst to be agnostic about any choice data that is beyond the scope of the model and provides some additional freedom to interpret data. Second, it allows for just enough flexibility so that we can include deterministic choice as a special case of random choice. In particular, the subjective expected utility model of Anscombe-Aumann [3] obtains as a degenerate case.

We first provide axioms (i.e. testable implications) that characterize information representations. The first four axioms (*monotonicity*, *linearity*, *extremeness* and *continuity*) are direct translations of the random expected utility axioms from Gul and Pesendorfer [47]. Next, we introduce three new axioms. *Non-degeneracy* ensures that the decision-maker is not universally indifferent. *C-determinism* states that the decision-maker must choose deterministically over constant acts, that is, acts that yield the same payoff in all states. This follows from the fact that private information only affects beliefs and not tastes. Since a constant act yields the same payoff regardless of what beliefs are, choice over constant acts must be deterministic. Finally, *S-monotonicity* states that if an act is the best act in a decision-problem for every state, then that act must be chosen for sure. It is the random choice version of the state-by-state monotonicity condition under deterministic choice.

Theorem 1.1 is a representation theorem. It states that a RCR has an information representation if and only if it satisfies the seven axioms above. Theorem 1.2 describes the uniqueness properties of information representations. It asserts that analyzing binary decision-problems is sufficient to completely identify the decision-maker's private information.

We then introduce a key technical tool that will feature prominently in our subsequent analysis. Given a decision-problem, consider adding a *test act* (e.g. a fixed payoff) to the original decision-problem in order to entice the decision-maker. Consider the probability that some act in the original decision-problem will be chosen over this test act. As we gradually decrease the value of the test act, this probability will increase. We call this schedule the *test function* for the decision-problem. Test functions are cumulatives that characterize the utility distributions of decision-problems. They also serve as sufficient statistics for identifying private information.

Following, we proceed to address our main questions of identification. First, we *evaluate option sets*. In the individual interpretation, the valuation of an option set is the ex-ante utility of the option set at time 0, that is, before the individual agent receives her information. In the group interpretation, the valuation of an option set is the welfare or total utility of the option set for all agents in the group.[4] Given a decision-maker's random choice, how do we compute the valuations of option sets (i.e. decision-problems)? We show that there is an intimate connection between random choice and valuations. Theorem 1.3 shows that computing integrals of test functions allow us to recover the valuations of decision-problems. Conversely, Theorem 1.4 shows that computing the marginal valuations of decision-problems with respect to test acts allow us to recover the decision-maker's random choice. This latter result is directly analogous to Hotelling's Lemma from classical producer theory. Theorems 1.3 and 1.4 describe operations that are mathematical inverses of each other; given random

[4] McFadden [60] calls this the "social surplus".

choice, we can directly compute valuations, and given valuations, we can directly compute random choice.

Next, we *assess informativeness*. In the classical approach of Blackwell [10, 11], more information is characterized by higher valuations (i.e. ex-ante utilities) of option sets. What is the random choice characterization of more information? Theorem 1.5 shows that in both the individual and group interpretations, more private information is characterized by second-order stochastic dominance of test functions. Given two decision-makers (either two individual agents or two groups of agents), one is more informed than the other if and only if test functions under the latter second-order stochastic dominate test functions under the former. This allows us to equate an unobservable multi-dimensional information ordering with observable single-dimensional stochastic dominance relations. Intuitively, more information is characterized by more dispersion or randomness in choice. For example, in the special case where information corresponds to events that partition S, more information is exactly characterized by less deterministic choice.

We then apply these results to *detect biases*. Suppose valuations of option sets (i.e. decision-problems) are observable via a preference relation (or ranking) over all decision-problems. Can we detect situations when this preference relation is inconsistent with the decision-maker's random choice? In the individual interpretation, this describes a form of dynamic inconsistency where this time-0 preference relation (reflecting ex-ante utilities) suggests a more (or less) informative signal than that implied by time-2 random choice. We call this *prospective overconfidence* (or *underconfidence*). An example of the former would be the *diversification bias* where an individual prefers large option sets at time 0 but always chooses the same option at time 2. An example of the latter would be the *confirmation bias* where an individual's beliefs after receiving her signal at time 2 are more extreme (i.e. more dispersed) than what she anticipated before receiving her signal at time 0. Both are examples of *subjective misconfidence*. These biases also apply in the group interpretation. For example, in the case of health insurance, consider a firm that chooses sets of health plans for its employees based on total welfare (i.e valuation). In this case, any inconsistency suggests that the firm has an incorrect assessment of the distribution of beliefs among its employees. By studying both valuations and random choice, we can detect these biases even when information is not directly observable.

Lastly, we *calibrate beliefs*. So far, we have adopted a completely subjective treatment of beliefs. In other words, we have been silent about whether beliefs are well-calibrated. By well-calibrated, we mean that beliefs as implied by random choice are consistent with both choice data *and* actual state realizations. In the individual interpretation, this implies that the agent has rational expectations about her signals. In the group interpretation, this implies that agents have beliefs that are predictive of actual state realizations and suggests that there is genuine private information in the group. If we consider the joint data over choices and state realizations, can we tell if beliefs are well-calibrated? We first define a *conditional test function* where we vary the payoffs of a *conditional test act* only in a given state. Theorem 1.6 shows that beliefs are well-calibrated if and only if conditional test functions and unconditional test functions have the same mean. This provides a test for rational beliefs even when the decision-maker's information is not directly observable. Finally, we use this result to determine when our notions of subjective misconfidence discussed above are measures of actual objective misconfidence.

1.2 An Informational Model of Random Choice

1.2.1 Random Choice Rules

We now describe the main primitive (i.e. choice data). A decision-maker, either an individual agent or a group of agents, faces a set of options to choose from. We call the set of options a *decision-problem* and each option in the set an *act*. If chosen, an act yields a payoff that depends on the realization of some underlying state. For example, an act could correspond to "forgoing health insurance", in which case an agent will receive a high payoff if she is healthy and a low payoff if she is sick. The object of our analysis is a *random choice rule (RCR)* that specifies the probabilities that acts are chosen in every decision-problem. In the *individual interpretation* of random choice, where the decision-maker is an individual agent, the RCR specifies the frequency distribution of choices by the agent if she chooses from the same decision-problem repeatedly. In the *group interpretation* of random choice, where the decision-maker is a group of agents, the RCR specifies the frequency distribution of choices in the group if every agent in the group chooses from the same decision-problem.

Formally, let S and X be finite sets. We interpret S as an objective state space and X as a set of possible prizes. For example, in the case of health insurance, $S = \{s_1, s_2\}$ where agents can either either be sick (s_1) or healthy (s_2). Let ΔS and ΔX be their respective probability simplexes. We interpret ΔS as the set of beliefs about the state space and ΔX as the set of lotteries over prizes. Following the setup of Anscombe and Aumann [3], an *act* is a mapping $f : S \to \Delta X$ that specifies a payoff in terms of a lottery on X for each realization of $s \in S$. Let H be the set of all acts. A *decision-problem* is a finite non-empty subset of H. Let $\mathcal{K}$ be the set of all decision-problems, which we endow with the Hausdorff metric.[5] For notational convenience, we also let f denote the singleton set $\{f\}$ whenever there is no risk of confusion.

In the classic model of rational choice, if an agent prefers one option over another, then this preference is revealed via her choice of the preferred option. If the two options are indifferent (i.e. they have the same utility), then the model is silent about which option will be chosen. We introduce an analogous innovation to address indifferences under random choice and random utility. Consider the decision-problem $F = \{f, g\}$. If the two acts f and g are "indifferent" (i.e. they have the same random utility), then we declare that the random choice rule is unable to specify choice probabilities for each act in the decision-problem. For

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

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

instance, it could be that f is chosen over g with probability a half, but any other probability would also be perfectly consistent with the model. Thus, similar to how the classic model is silent about which act will be chosen in the case of indifference, the random choice model is silent about what the choice probabilities are. In both cases, we can interpret indifferences as choice behavior that is beyond the scope of the model. This provides the analyst with some additional freedom to interpret data.

Let $\mathcal{H}$ be some σ-algebra on H. Formally, we model indifference as non-measurability with respect to $\mathcal{H}$. For example, if $\mathcal{H}$ is the Borel algebra, then this corresponds to the benchmark case where every act is measurable with respect to $\mathcal{H}$. In general though, $\mathcal{H}$ can be coarser than the Borel algebra. Note that given a decision-problem, the decision-problem itself must be measurable. This is because we know that something will be chosen from the decision-problem. For $F \in \mathcal{K}$, let $\mathcal{H}_F$ be the σ-algebra generated by $\mathcal{H} \cup \{F\}$.[6] Let Π be the set of all probability measures on any measurable space of H. We now formally define a random choice rule.

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

For $F \in \mathcal{K}$, we let ρ_F denote the measure $\rho(F)$. A RCR thus assigns a probability measure on $(H, \mathcal{H}_F)$ for each decision-problem $F \in \mathcal{K}$ such that $\rho_F(F) = 1$. Note that the definition of $\mathcal{H}_F$ ensures that $\rho_F(F)$ is well-defined. We interpret $\rho_F(G)$ as the probability that some act in $G \in \mathcal{H}_F$ will be chosen given the decision-problem $F \in \mathcal{K}$. For example, if an insurance company offers two health plans $F = \{f, g\}$, then $\rho_F(f)$ is the probability that plan f is chosen over plan g. For ease of exposition, we denote RCRs by ρ with the implicit understanding that it is associated with some $\mathcal{H}$.

If $G \subset F$ is not $\mathcal{H}_F$-measurable, then $\rho_F(G)$ is not well-defined. To address this, let

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

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

be the measure of the smallest measurable set containing G.[7] Note that ρ_F^* is exactly the outer measure of ρ_F. Both ρ_F and ρ_F^* coincide on $\mathcal{H}_F$-measurable sets. Going forward, we let ρ denote ρ^* without loss of generality.

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

Example 1.1. Let $S = \{s_1, s_2\}$ and $X = \{x, y\}$. Without loss of generality, we can let $f = (a, b) \in [0, 1]^2$ denote the act $f \in H$ where

$$f(s_1) = a\delta_x + (1 - a)\delta_y$$
$$f(s_2) = b\delta_x + (1 - b)\delta_y$$

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 $\rho_F(f) = 1$ iff $f_1 \geq g_1$ for all $g \in F$. Thus, acts are ranked based on how likely they will yield prize x if state s_1 occurs. In the health insurance example, this describes agents who prefer x to y and believe that they will fall sick (i.e. s_1 occurs) for sure. If we let $F = \{f, g\}$ be such that $f_1 = g_1$, then neither f nor g is $\mathcal{H}_F$-measurable. In other words, the RCR is unable to specify choice probabilities for f or g. This is because both acts yield prize x with the same probability in state s_1. Hence, the two acts are "indifferent". Observe that ρ corresponds exactly to classic deterministic choice where f is preferred to g iff $f_1 \geq g_1$.

1.2.2 Information Representations

We now describe the relationship between random choice and private information. At time 2, the decision-maker, either an individual agent or a group of agents, chooses from a decision-

[7] Lemma 1A.1 in the Appendix ensures that this is well-defined.

problem. At time 1, the decision-maker receives some private information about the underlying state space S. For instance, if the decision-problem is whether to purchase health insurance or not, then this private information will affect beliefs about the likelihood of falling sick. After receiving this information, each agent will choose the best act from the decision-problem conditional on her subjective belief about S. Since this private information is unobservable to the analyst, choice at time 2 is probabilistic and can be modeled as a RCR. We call this an *information representation* of random choice.

Formally, after receiving her private information, each agent forms a posterior belief $q \in \Delta S$. Thus, we model private information as an unobservable distribution μ on ΔS. In the individual interpretation, μ is the distribution of signal realizations over the canonical signal space ΔS. Note that this approach allows us to circumvent any issues that arise with updating beliefs and enables us to work directly with the posterior beliefs driving choice. In the group interpretation, μ is the distribution of beliefs in the group. As an example, consider the degenerate distribution $\mu = \delta_q$ for some $q \in \Delta S$. In the individual interpretation, this corresponds to the case where the agent receives a degenerate signal that induces a single posterior belief q. If q is also the prior, then this describes an agent who receives no information about S. In the group interpretation, this corresponds to the case where all agents in the group share the same belief about S. Thus, all agents in the group are completely indistinguishable from each other. Note that in both interpretations, the decision-maker's choice in this example is deterministic.

Let $u : \Delta X \to \mathbb{R}$ be an affine utility function. We interpret $q \cdot (u \circ f)$ as the subjective expected utility of the act $f \in H$ conditional on the posterior belief $q \in \Delta S$.[8] Since we are interested in the studying the effects of information on random choice, we keep tastes (i.e. risk preferences) the same by holding u fixed (in Section 8, we relax this assumption). In the individual interpretation, this implies that signals only affect beliefs about S and not preferences over ΔX. In the group interpretation, this implies that all agents in the group share the same tastes but differ over beliefs about S. In both interpretations, choice

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

is stochastic only as a result of varying beliefs.

We say μ is *regular* iff the subjective expected utilities of any two acts are either always or never equal. This is a relaxation of the standard restriction in traditional RUM models where utilities are never equal.

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

Let (μ, u) consist of a regular μ and a non-constant u. We are now ready to formally state the relationship between a RCR and its information structure.

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

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

This is a RUM model where the random utilities are subjective expected utilities that depend on unobservable private information. If a RCR is represented by (μ, u), then the probability that an act f is chosen in the decision-problem F is equal to the probability that the subjective expected utility of f is higher than that of every other act in F. For instance, in the health insurance example, the probability of buying health insurance corresponds to the probability of receiving a bad report from a health exam. Note that the latter probability (i.e. μ) is subjective as the decision-maker's private information is unobservable to the analyst. Thus, any inference about this information can only be gleaned by studying the RCR.

One of the classic critiques of subjective expected utility (especially in the context of health insurance) is the state independence of the (taste) utility. In an information representation, utilities are independent of both the unobservable subjective states affecting information and the objective state space S. The former is addressed in Section 8 below where we characterize a general model that allows for unobservable utility shocks. The latter can by addressed by any random choice generalization of the classic solutions to state-dependent utility (see Karni, Schmeidler and Vind [54] and Karni [53]).[9]

[9] In practice however, the empirical literature on health insurance has largely assumed state independence due to a dearth of empirical evidence (see Finkelstein, Luttmer and Notowidigdo [33]).

We follow with two examples of information representations. The first example is deterministic while the second is not.

Example 1.2. Let $S = \{s_1, s_2\}$, $X = \{x, y\}$ and $u\left(a\delta_x + (1-a)\,\delta_y\right) = a \in [0,1]$. Let $\mu = \delta_q$ for $q \in \Delta S$ such that $q_{s_1} = 1$. In this example, agents believe that s_1 will occur for sure so they only care about payoffs in state s_1. Let (μ, u) represent ρ, and let $F = \{f, g\}$. If $(u \circ f)_{s_1} \geq (u \circ g)_{s_1}$, then

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

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

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

and neither f nor g is $\mathcal{H}_F$-measurable. Note that this is exactly the RCR described in Example 1.1 above.

Example 1.3. Let $S = \{s_1, s_2, s_3\}$, $X = \{x, y\}$ and $u\left(a\delta_x + (1-a)\,\delta_y\right) = a \in [0,1]$. Let μ be the uniform measure on ΔS, and let (μ, u) represent ρ. Given two acts f and g such that $u \circ f = v \in [0,1]^3$ and $u \circ g = w \in [0,1]^3$, we have

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

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

Example 1.2 above is exactly the standard subjective utility model where agents believe that s_1 will realize for sure. It serves to demonstrate how our random choice model includes standard subjective expected utility as a special deterministic case.

We conclude this section with a technical remark about our definition of regularity. As briefly mentioned above, in traditional RUM models, indifferences in the random utility must occur with probability zero. This is because all choice options are assumed to be measurable with respect to the RCR, so these models run into trouble when dealing with indifferences in the random utility.[10] Our definition of regularity enables us to circumvent these issues by allowing for just enough flexibility so that we can model indifferences using non-measurability.[11] For instance in Example 1.2 above, if $q \cdot (u \circ f) = q \cdot (u \circ g)$ μ-a.s., then neither f nor g is $\mathcal{H}_{f \cup g}$-measurable. Acts that have the same utility μ-a.s. correspond exactly to non-measurable singletons. Note that our definition however still imposes certain restrictions on μ. For example, multiple mass points are not allowed if μ is regular.[12]

1.2.3 Axiomatic Characterization

In this section, we provide a set of axioms (i.e testable properties) on the RCR to characterize information representations. We also identify the uniqueness properties of information representations.

Given two decision-problems F and G, let $aF + (1-a)G$ denote the Minkowski mixture of the two sets for some $a \in [0,1]$.[13] Let ext$F$ denote the set of extreme acts of $F \in \mathcal{K}$.[14] We assume $f \in F \in \mathcal{K}$ throughout. The first three axioms below are standard restrictions on RCRs.

Axiom 1.1. (Monotonicity) $G \subset F$ *implies* $\rho_G(f) \geq \rho_F(f)$.

Axiom 1.2. (Linearity) $\rho_F(f) = \rho_{aF+(1-a)g}(af + (1-a)g)$ *for* $a \in (0,1)$.

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

[11] 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 [47] on the other hand, requires μ to be full-dimensional (see their Lemma 2). See Block and Marschak [12] for the case of finite alternatives.

[12] See Example 1.7 below.

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

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

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

Axiom 1.3. (Extremeness) $\rho_F(extF) = 1$.

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

To understand linearity and extremeness, note that the random utilities in our model are linear. In other words, after receiving private information, agents are standard subjective expected utility maximizers. As a result, choice behavior must satisfy the standard properties of expected utility maximization. Linearity is exactly the random choice analog of the standard independence axiom. In fact, it is *the* version of the independence axiom that is tested in many experimental settings (for example, see Kahneman and Tversky [51]). Extremeness implies that only extreme acts of the decision-problem will be chosen. This follows from the fact that linear utilities are used to evaluate acts. Thus, any act that is a mixture of other acts in the decision-problem will never be chosen (except for the borderline case of indifference). Note that both linearity and extremeness rule out situations where the decision-maker may exhibit behaviors associated with random non-linear utilities (such as ambiguity aversion for example).

We now introduce the continuity axiom for our model. Given a RCR, let $\mathcal{K}_0 \subset \mathcal{K}$ be the set of decision-problems where every act in the decision-problem is measurable with respect to the RCR. To be explicit, $F \in \mathcal{K}_0$ iff $f \in \mathcal{H}_F$ for all $f \in F$. Let Π_0 be the set of all Borel measures on H, endowed with the topology of weak convergence. Since all acts in $F \in \mathcal{K}_0$ are $\mathcal{H}_F$-measurable, $\rho_F \in \Pi_0$ for all $F \in \mathcal{K}_0$ without loss of generality.[15] We say ρ is *continuous* iff it is continuous on the restricted domain $\mathcal{K}_0$.

Axiom 1.4. (Continuity) $\rho : \mathcal{K}_0 \to \Pi_0$ *is continuous.*

If $\mathcal{H}$ is the Borel algebra, then $\mathcal{K}_0 = \mathcal{K}$. In this case, our continuity axiom condenses to standard continuity. In general though, the RCR is not continuous over all decision-problems.

[15] We can easily complete ρ_F so that it is Borel measurable.

In fact, the RCR is discontinuous at decision-problems that contain indifferences. In other words, choice data that is beyond the scope of the model exhibits discontinuities with respect to the RCR. In our model, every decision-problem is arbitrarily (Hausdorff) close to some decision-problem in $\mathcal{K}_0$.[16] Thus, continuity is preserved over almost all decision-problems.

Taken together, the first four axioms above are the necessary and sufficient conditions for a random expected utility representation (see Gul and Pesendorfer [47]). We now introduce the axioms that are particular to our model. We say $f \in H$ is *constant* iff $f(s)$ is the same for all $s \in S$. A decision-problem is *constant* iff it only contains constant acts. Given $f \in H$ and $s \in S$, define $f_s \in H$ as the constant act that yields the lottery $f(s) \in \Delta X$ in every state. For $F \in \mathcal{K}$, let $F_s := \bigcup_{f \in F} f_s$ be the constant decision-problem consisting of f_s for all $f \in F$. We now introduce the final three axioms.

Axiom 1.5. (Non-degeneracy) $\rho_F(f) < 1$ *for some* F *and* $f \in F$.

Axiom 1.6. (C-determinism) $\rho_F(f) \in \{0, 1\}$ *for constant* F.

Axiom 1.7. (S-monotonicity) $\rho_{F_s}(f_s) = 1$ *for all* $s \in S$ *implies* $\rho_F(f) = 1$.

Non-degeneracy rules out the trivial case where all acts are indifferent. C-determinism states that the RCR is deterministic over constant decision-problems. This is because choice is stochastic only as a result of varying beliefs about S. In a constant decision-problem, all acts yield the same payoff regardless of which state occurs. Hence, in the individual interpretation, the agent will choose the same act regardless of which signal she receives. In the group interpretation, all agents will choose the same act since they all share the same tastes. In both interpretations, choice over constant decision-problems is deterministic. In fact, if ρ is represented by (μ, u), then ρ induces a preference relation over constant acts that is exactly represented by u.

To understand S-monotonicity, note that in standard deterministic choice, the state-by-state monotonicity axiom says that if f_s is preferred to g_s for every state $s \in S$, then f must

[16] In other worlds, $\mathcal{K}_0$ is dense in $\mathcal{K}$ (see Lemma 1A.15 in the Appendix).

be preferred to g. Translated to the realm of random choice, this means that if f_s is chosen with certainty in F_s for every state $s \in S$, then f must be chosen with certainty in F. This is exactly S-monotonicity. Theorem 1.1 shows that Axioms 1.1-1.7 are necessary and sufficient for an information representation.

Theorem 1.1. *ρ has an information representation iff it satisfies Axioms 1.1-1.7.*

Proof. See Appendix. □

Theorem 1.2 highlights the uniqueness properties of information representations. The main highlight is that studying binary choices is enough to completely identify private information. In other words, given two decision-makers (either two individual agents or two groups of agents) with RCRs that have information representations, comparing binary choices is sufficient to completely differentiate between the two information structures.

Theorem 1.2 (Uniqueness)**.** *Suppose ρ and τ are represented by (μ, u) and (ν, v) respectively. Then the following are equivalent:*

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

(2) $\rho = \tau$

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

Proof. See Appendix. □

Note that if we allow the utility u to be constant, then non-degeneracy can be dropped in Theorem 1.1 without loss of generality. However, the uniqueness of μ in the representation would obviously fail in Theorem 1.2.

1.3 Test Functions

We now introduce the key technical tool that will play an important role in our subsequent analysis. To motivate the discussion, first recall the setup of our model. At time 2, a decision-maker (either an individual agent or a group of agents) will choose something from an option set (i.e. decision-problem) $F \in \mathcal{K}$. At time 1, the decision-maker receives private information about the underlying state. Now, imagine enticing the decision-maker with some *test act* that yields a fixed payoff in every state. For example, this could be a no-deductible (full-insurance) health plan introduced by the insurance company. What is the probability that at time 2, something in F will be chosen over the test act? If the test act is very valuable (i.e. the fixed payoff is high), then this probability will be low. On the other hand, if the test act is not very valuable (i.e. the fixed payoff is low), then this probability will be high. Thus, as we decrease the value of the test act, the probability that something in F will be chosen increases. We call this schedule the *test function* of F.

For a concrete example, suppose $F = g$ is the act that corresponds to choosing some high-deductible health insurance plan. Our test act corresponds to choosing a no-deductible (full-insurance) plan. In this case, the test function of g is the probability of choosing plan g over the full-insurance test plan as a function of its premium. As we increase the test plan's premium, it becomes less valuable and the probability of choosing g increases.

Formally, we define test functions as follows. First, define the *best* and *worst* acts for a RCR. An act is the *best* (*worst*) act under ρ iff in any binary choice comparison, the act (other act) is chosen with certainty.

Definition. $\overline{f}$ and $\underline{f}$ are the *best* and *worst* acts under ρ respectively iff $\rho_{f \cup \overline{f}}\left(\overline{f}\right) = \rho_{f \cup \underline{f}}\left(f\right) = 1$ for all $f \in H$.

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

Formally, a *test act* is the mixture act $a\underline{f} + (1-a)\overline{f}$ for some $a \in [0,1]$. Thus, test acts are mixtures between the best and worst acts. Since both $\overline{f}$ and $\underline{f}$ are constant acts, test acts are also constant acts. Thus, under an information representation, their utilities are fixed and independent of any beliefs about S. Test acts serve as the enticing option in the opening discussion above. We now define *test functions* as follows.

Definition. Given ρ, let $F_\rho : [0,1] \to [0,1]$ be the *test function* of $F \in \mathcal{K}$ where for $f^a := a\underline{f} + (1-a)\overline{f}$,

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

Given a RCR ρ and a decision-problem $F \in \mathcal{K}$, we let F_ρ denote the test function of F. Suppose ρ has an information representation. As we increase a, the test act f^a progresses from the best to worst act and becomes increasingly unattractive. Thus, the probability of choosing something in F increases so F_ρ is an increasing function. In fact, it characterizes the utility distribution of F. Lemma 1.1 states that test functions are cumulative distribution functions under information representations.

Lemma 1.1. *If ρ has an information representation, then F_ρ is a cumulative for all $F \in \mathcal{K}$.*

Proof. See Appendix. □

If $F = f$ is a singleton act, then we denote $F_\rho = f_\rho$. An immediate corollary of the uniqueness properties of information representations is that test functions for singleton acts are sufficient for identifying information structures.

Corollary 1.1. *Let ρ and τ have information representations. Then $\rho = \tau$ iff $f_\rho = f_\tau$ for all $f \in H$.*

Proof. Follows from Theorem 1.2. □

Corollary 1.1 implies that we can treat test functions as sufficient statistics for identifying private information. In the following sections, we demonstrate how test functions can be used to perform various exercises of inference. We end this section with a couple examples of test functions.

Example 1.4. Recall Example 1.3 where $S = \{s_1, s_2, s_3\}$, $X = \{x, y\}$, $u\left(a\delta_x + (1-a)\,\delta_y\right) = a \in [0,1]$ and μ is the uniform measure on ΔS. Let (μ, u) represent ρ. Thus, $u \circ \overline{f} = (1,1,1)$ and $u \circ \underline{f} = (0,0,0)$. For $a \in [0,1]$, the test act f^a satisfies

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

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

$$f_\rho\,(a) = \rho_{f \cup f^a}\,(f) = \mu\left\{q \in \Delta S \mid q_{s_1} \geq 1-a\right\} = a$$

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

Note that since g is a constant act, its utility is fixed regardless of what beliefs are. Thus, its test function increases abruptly at the critical value $1-b$. On the other hand, the utility of f depends on the agent's belief, so its test function increases more gradually as a increases.

1.4 Evaluating Option Sets

We now address our first identification exercise. Given an option set (i.e. decision-problem), we will compute its *valuation*. In the individual interpretation, the valuation of an option set is the ex-ante utility of the option set at time 0, that is, before the individual agent receives her signal. In the group interpretation, the valuation of an option set is the welfare or total utility of the option set for all agents in the group. What is the relationship between the valuations *of* option sets and random choice *from* option sets? We call this exercise *evaluating option sets.*

Formally, we model valuations via a preference relation $\succeq$ over all decision-problems $\mathcal{K}$. We call $\succeq$ the *valuation preference relation.* In the individual interpretation, $\succeq$ reflects the agent's time-0 utility of each decision-problem. Thus, at time 0, the agent prefers decision-problem F to decision-problem G (i.e. $F \succeq G$) iff the ex-ante utility of F is greater than the ex-ante utility of G. Since information is unknown at time 0, $\succeq$ is based on the agent's beliefs about signals that she expects to receive at time 1. For example, if the agent expects to receive results from her health exam next week that will inform her of her health status, then she may prefer to postpone the decision of which health insurance plan to purchase until then. In other words, at time 0, she prefers the option set that contains both health plans to committing to a single health plan.

In the group interpretation, $\succeq$ reflects the group's total utility or welfare for each decision-problem. Thus, the group prefers decision-problem F to decision-problem G (i.e. $F \succeq G$) iff the total utility of F is greater than the total utility of G. Note that $\succeq$ is based on the distribution of heterogeneous beliefs in the group. For example, if a firm thinks that its employees have very disperse beliefs about their health, then it may prefer an insurance company that offers a larger set of health plans to one that does not. In other words, the firm may prefer a more flexible set of health plans (even if it is more expensive) in order to increase welfare for its employees.

In this section, we show that there is an intimate connection between this valuation preference relation and random choice. Given a decision-maker's random choice, we can directly recover this valuation preference relation. Vice-versa, given the valuation preference relation, we can directly recover the decision-maker's random choice. In other words, knowing either valuations or random choice allows us to completely and uniquely identify the other.

We now address the formal model. We say that a preference relation $\succeq$ is *represented by* (μ, u) iff it satisfies the following.

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

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

If we interpret $\succeq$ as the individual valuation preference relation, then $V(F)$ is exactly the time-0 expected utility of $F \in \mathcal{K}$ for an agent with information μ and utility u. Note that this is the subjective learning representation axiomatized by Dillenberger, Lleras, Sadowski and Takeoka [23] (henceforth DLST). If we interpret $\succeq$ as the group valuation preference relation, then $V(F)$ is exactly the total utility or welfare of $F \in \mathcal{K}$ for a group of agents with beliefs distributed according to μ and utility u.

Suppose that we know the decision-maker's random choice (i.e. ρ). What can we infer about the valuation preference relation (i.e. $\succeq$)? First, in this section, we assume that ρ admits test functions that are cumulatives, that is, ρ has a best and worst act and F_ρ is a well-defined cumulative for all $F \in \mathcal{K}$.[17] We say that ρ is *standard* iff it satisfies monotonicity, linearity and continuity (Axioms 1.1, 1.2 and 1.4).

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

If ρ has an information representation, then it is necessarily standard. Note that these conditions on ρ are relatively mild. For example, they are insufficient to ensure that a random utility representation even exists for ρ.

We now demonstrate how we can completely recover $\succeq$ from ρ. To gain some intuition for how, note that if a decision-problem F is very valuable, then acts in F will be chosen with high probabilities. Hence, the test function F_ρ will take on high values. As a result, consider evaluating decision-problems as follows.

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

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

Theorem 1.3 below confirms that $\succeq_\rho$ is the valuation preference relation corresponding to the RCR ρ. It shows that we can simply use V_ρ to evaluate decision-problems.

Theorem 1.3. *The following are equivalent:*

(1) *ρ is represented by (μ, u)*

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

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

Proof. See Appendix. □

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

Theorem 1.3 also demonstrates that if a standard RCR induces a preference relation that has a subjective learning representation, then that RCR must have an information representation. In fact, both the RCR and the preference relation are represented by the same (μ, u). We can thus interpret Theorem 1.3 as an alternate characterization of information representations using properties of its induced preference relation.

The discussion above suggests that perhaps there is a more direct method of obtaining the RCR from the valuation preference relation. We now demonstrate how this can be accomplished. First, we say that a preference relation $\succeq$ is *dominant* iff it satisfies the following.

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

Dominance is one of the axioms of a subjective learning representation in DLST. It captures the intuition that adding acts that are dominated in every state does not affect the valuation of the decision-problem.

We now define a RCR induced by a preference relation as follows.

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

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

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

Note that given any generic preference relation $\succeq$, the RCR $\rho_\succeq$ may not even exist. This is because we may not be able to find any V and ρ that satisfy this definition. On the other hand, there could be a multiplicity of RCRs that satisfy this definition. Our result below shows that for our purposes, these issues need not be of concern. Theorem 1.4 asserts that if $\succeq$ has a subjective learning representation, then $\rho_{\succeq}$ exists and is unique. It confirms that $\rho_{\succeq}$ is the RCR corresponding to $\succeq$.

Theorem 1.4. *The following are equivalent:*

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

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

Proof. See Appendix. □

Thus, if $\succeq$ has a subjective learning representation, then $\rho_{\succeq}$ is the unique RCR corresponding to $\succeq$. The probability that the act f_a is chosen in a decision-problem is exactly its marginal contribution to the valuation of the decision-problem. In other words, if increasing the value of f_a does not affect the valuation of the decision-problem, then f_a is never chosen. For example, consider a set of health insurance plans that includes a no-deductible (full-insurance) plan. If lowering the premium of the full-insurance plan does not affect the valuation of the set of plans, then the full-insurance plan will never be chosen from the set. Any violation of this would indicate some form of inconsistency (which we will explore in Section 6).

Theorem 1.4 is actually the random choice version of Hotelling's Lemma from classical producer theory.[18] The analogy follows if we interpret choice probabilities as "outputs" and conditional utilities as "prices". In this case, we can interpret the valuation of a decision-problem as a firm's maximizing "profit". Thus, similar to how Hotelling's Lemma allows us to recover a firm's output choices from its profit function, Theorem 1.4 allows us to recover a decision-maker's random choice from valuations.[19]

[18] In the econometrics literature, Theorem 1.4 is related to the Williams-Daly-Zachary Theorem (McFadden [60]). The presence of constant acts in the Anscombe-Aumann setup however allows us to formulate Theorem 1.3 which has no counterpart in that literature.

[19] For a formal exposition, consider the following. 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 $p_f := \int_{Q_f} q \cdot (u \circ f) \frac{1}{\mu(Q_f)} \mu(dq)$

The other implication of Theorem 1.4 is that if a dominant preference relation induces a RCR that has an information representation, then that preference relation must have a subjective learning representation. As in Theorem 1.3, we interpret this as an alternate characterization of subjective learning representations using properties of its induced RCR.

Given any $\succeq$ represented by (μ, u), we can easily construct the corresponding RCR $\rho = \rho_{\succeq}$ as follows. First define ρ so that it coincides with u over all constant acts. This allows us to set the best and worst acts of ρ so that they are also the best and worst acts of u. We can now define $\rho_{F \cup f_a}(f_a)$ for all $a \in [0,1]$ using the definition of $\rho_{\succeq}$. Linearity then allows us to extend ρ to the space of all acts. By Theorem 1.4, the ρ so constructed is represented by (μ, u). Hence, ρ is exactly the RCR corresponding to $\succeq$.

Test functions play a distinguished role in the analysis above. Integrating them allows us to recover valuations, while differentiating valuations allows us to recover random choice. Thus, $\succeq_\rho$ and $\rho_{\succeq}$ are invertible operations that correspond to integration and differentiation respectively. In the individual interpretation, this means that observing choice behavior in one time period allows us to identify and directly compute choice behavior in the other. This eliminates the need of identifying the signal distribution and utility. We summarize these insights in Corollary 1.2 below.

Corollary 1.2. *Let $\succeq$ and ρ be represented by (μ, u). Then $\succeq_\rho = \succeq$ and $\rho_{\succeq} = \rho$.*

Proof. Follows immediately from Theorems 1.3 and 1.4. □

The following example demonstrate how these operations can be used to recover valuations from random choice and vice-versa.

Example 1.5. Let $S = \{s_1, s_2\}$, $X = \{x, y\}$ and $u(a\delta_x + (1-a)\delta_y) = a \in [0,1]$. We associate each $q \in \Delta S$ with $t \in [0,1]$ such that $t = q_{s_1}$. Let μ have density $6t(1-t)$. Let $\succeq$ and ρ be represented by (μ, u) and $V : \mathcal{K} \to [0,1]$ represents $\succeq$. An insurance company

denote the conditional utility of f. If we interpret y as "output" and p as "price", then $V(F) = \sup_{y, \mathcal{Q}_y} p \cdot y$ is the maximizing "profit". Note that $a = p_{f_a}$ is exactly the "price" of f_a. Of course, in Hotelling's Lemma, "prices" are fixed while in our case, p_f depends on the partition $\mathcal{Q}_y$.

offers two health plans: a no-deductible (full-insurance) plan f and a high-deductible plan g. Let $u \circ f = \left(\frac{2}{5}, \frac{2}{5}\right)$, $u \circ g = \left(\frac{1}{4}, \frac{3}{4}\right)$ and $F = \{f, g\}$.

Valuations from random choice: What is the valuation of F if we only observe ρ? The test function of F is given by

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

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

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

for $a \in \left(\frac{1}{4}, \frac{3}{5}\right)$. Integrating F_ρ yields

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

Random choice from valuations: What is the probability of choosing f over g if we only observe $\succeq$? Let $f_a := a\overline{f} + (1-a)\underline{f}$ where $a \in [0,1]$ and note that $f_{\frac{2}{5}} = f$. It is straightforward to check that for $a \in \left(\frac{1}{4}, \frac{3}{4}\right)$

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

Differentiating $V(g \cup f_a)$ at $a = \frac{2}{5}$ yields

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

1.5 Assessing Informativeness

1.5.1 Random Choice Characterization of More Information

We now provide a random choice characterization of more information. Given two decision-makers (either two individuals or two groups of agents), we can determine which decision-maker is more informed even when information is not directly observable and can only be inferred from random choice. We call this *assessing informativeness*.

First, consider the classic methodology of assessing informativeness when information is observable. Let μ and ν be two measures on ΔS. We say that a transition kernel[20] on ΔS is *mean-preserving* iff it preserves average beliefs about S.

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

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

We say that 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 corresponds exactly to Blackwell's [10, 11] ranking of informativeness based on signal sufficiency. In other words, μ is a sufficient signal for generating ν. Note that in the case where K is the identity kernel, no information is lost and $\nu = \mu$.

In the classical approach, Blackwell [10, 11] showed that more information is characterized by higher valuations of option sets. What is the random choice characterization of more

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

information? To gain some intuition for our result below, consider the following individual interpretation of random choice. An agent receives a completely uninformative (i.e. degenerate) signal about the underlying state. As a result, her choice behavior is deterministic. Consider her test function of some act f. As we gradually lower the test act, there will be a critical value where her choice changes abruptly from never choosing f to always choosing it. Since test functions are cumulatives under information representations, her test function of f corresponds to a single mass point at this critical value. A second agent on the other hand, receives a very informative signal. Depending on her signal, f could either be very valuable or not. Thus, her test function of f will increase more gradually. If both test functions have the same mean under both agents, then the test function under the more informed agent will be a mean-preserving spread of the test function under the less informed agent. The same reasoning equally applies in the group interpretation of random choice. In general, this property is captured by second-order stochastic dominance.

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

The result below demonstrates that we can assess informativeness simply by comparing test functions via second-order stochastic dominance.

Theorem 1.5. *Let ρ and τ be represented by (μ, u) and (ν, u) respectively. Then μ is more informative than ν iff $F_\tau \geq_{SOSD} F_\rho$ for all $F \in \mathcal{K}$.*

Proof. See Appendix. □

Theorem 1.5 equates an unobservable multi-dimensional information ordering with an observable single-dimensional stochastic dominance relation. It is the random choice characterization of more information. The general intuition is that less information results in choice behavior that is more concentrated (i.e. deterministic) whereas more information results in choice behavior that is more dispersed (i.e. random). By studying test functions of two decision-makers (either two individual agents or two groups of agents), we can assess which decision-maker is more informed. For example, in the individual interpretation, the agent who is more informed about her health status every year will exhibit greater dispersion in her choice of yearly health insurance. In the group interpretation, the group with more private information will exhibit greater variation in the distribution of health insurance choices.

What is the relationship between this random choice characterization of more information and valuation preference relations? In DLST, more information is characterized by a greater preference for flexibility in the valuation preference relation. This is a translation of Blackwell's [10, 11] results to preference relations. Given two decision-makers, we say that one exhibits *more preference for flexibility* than the other iff whenever the other prefers a set of options to a single option, the first must do so as well.

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

Corollary 1.3 relates our random choice characterization of more information with more preference for flexibility.

Corollary 1.3. *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 1.5, (1) and (3) are equivalent. By Corollary 1.2, $\succeq_\rho$ and $\succeq_\tau$ are represented by (μ, u) and (ν, u) respectively. Hence, by Theorem 1.2 of DLST, (2) is equivalent to (3). □

Thus, in the individual interpretation, more information manifests itself as more ex-ante (before time 1) preference for flexibility and more ex-post (after time 1) variability in random choice. On the other hand, in the group interpretation, more information manifests itself as more preference for flexibility by the group and more variability in individual random choice. Note that by Corollary 1.2, we could have formulated Corollary 1.3 entirely in terms of valuation preference relations.

Corollary 1.3 also summarizes the prominent role of test functions in much of our analysis. Computing their integrals allow us to evaluate options sets while comparing them via second-order stochastic dominance allow us to assess informativeness.

If μ is more informative than ν, then the two measures must have the same average beliefs about S. In this case, we say that they *share average beliefs*.

Definition. μ and ν *share average beliefs* iff

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

In the the individual interpretation, two agents with information μ and ν share average beliefs iff they have the same prior about S before the arrival of information (i.e. at time 0). Note the distinction between this prior over S and the more general "prior" over the universal space $\Delta S \times S$.[21] In the the group interpretation, two groups with beliefs distributed according to μ and ν share average beliefs iff the average belief about S in both groups are the same. Lemma 1.2 shows that in order to determine if two decision-makers share average beliefs or not, it is necessary and sufficient to just compare means of test functions for singleton acts.

Lemma 1.2. *Let ρ and τ be represented by (μ, u) and (ν, u) respectively. Then μ and ν share average beliefs iff f_ρ and f_τ share the same mean for all $f \in H$.*

Proof. See Appendix. □

Combined with Theorem 1.5, Lemma 1.2 implies that a necessary condition for μ being more informative than ν is that every f_ρ is a mean-preserving spread of f_τ. This condition however is insufficient for assessing informativeness. It corresponds to a strictly weaker stochastic dominance relation known as the linear concave order.[22] Note that if f_ρ and f_τ

[21] Agreeing on the latter prior necessitates that both agents must have identical information structures.
[22] See Section 3.5 of Muller and Stoyan [63] for more about the linear concave order.

have the same mean, then f has the same valuation under both decision-makers. Thus, in a sense, the random choice characterization of more information is richer than the preference relation characterization of more information.

Finally, we end this section with an illustrative example.

Example 1.6. Let $S = \{s_1, s_2\}$, $X = \{x, y\}$ and $u(a\delta_x + (1-a)\delta_y) = a \in [0,1]$. We associate each $q \in \Delta S$ with $t \in [0,1]$ such that $t = q_{s_1}$. Let μ have density $6t(1-t)$ and ν be the uniform distribution. Since beliefs are more dispersed under ν, we say that ν is more informative than μ. Let ρ and τ be represented by (μ, u) and (ν, u) respectively. As in Example 1.5, consider the set of plans $F = \{f, g\}$ where $u \circ f = \left(\frac{2}{5}, \frac{2}{5}\right)$ and $u \circ g = \left(\frac{1}{4}, \frac{3}{4}\right)$. Recall that $F_\rho(a) = 0$ for $a \leq \frac{1}{4}$, $F_\rho(a) = 1$ for $a \geq \frac{3}{5}$ and

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

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

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

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

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

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

1.5.2 Special Case: Partitional Information

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

Formally, let $\left(S, 2^S, r\right)$ be a probability space for some $r \in \Delta S$. We assume that r has full support without loss of generality.[23] Given an algebra $\mathcal{F} \subset 2^S$, let $Q_{\mathcal{F}}$ be the conditional probability given $\mathcal{F}$, that is, for $s \in S$ and the event $E \subset S$,

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

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

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

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

We would like to consider RCRs generated by information structures that have beliefs distributed according to $\mu_{\mathcal{F}}$. However, except for the case where $\mathcal{F}$ is the trivial algebra, $\mu_{\mathcal{F}}$ is in general not regular. The following example demonstrates how violations of regularity can create issues with our method of modeling indifferences using non-measurability.

Example 1.7. Let $S = \{s_1, s_2, s_3\}$, $X = \{x, y\}$ and $u\left(a\delta_x + (1-a)\delta_y\right) = a \in [0,1]$. Let $r = \left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)$ and $\mathcal{F}$ be generated by the partition $\{s_1, s_2 \cup s_3\}$. Let $q_1 := \delta_{s_1}$ and $q_2 :=$

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

$\frac{1}{2}\delta_{s_2} + \frac{1}{2}\delta_{s_3}$ so

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

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

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

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

Let $F := \{f, g, h\}$ and note that the model cannot distinguish between f and g one third of the time. By similar reasoning, the model cannot distinguish between g and h two-thirds of the time. If we use non-measurability to model indifferences, then no singleton act in F is measurable. This approach will not be able to capture all the choice data implied by the model. For example, it omits the fact that h will definitely not be chosen one third of the time.

Example 1.7 illustrates that since $\mu_{\mathcal{F}}$ violates regularity, complications arise whenever a decision-problem contains acts that are neither always nor never equal (i.e. they have the same random utility on some $\mu_{\mathcal{F}}$-measure that is strictly between zero and one). We circumvent this problem by only considering decision-problems that do not contain such acts. These decision-problems are called *generic*.

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

Regularity is equivalent to requiring that all decision-problems are generic. Note that generic decision-problems are dense in the set of all decision-problems. Moreover, any $\mu_{\mathcal{F}}$ can always be approximated as the limit of a sequence of regular μ's. We now formally present the partitional information representation of an RCR.

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

is generic,

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

A partitional information representation is thus an information representation over generic decision-problems where the distribution of beliefs is $\mu_{\mathcal{F}}$. We say a decision-problem $F \in \mathcal{K}$ is *deterministic* under ρ iff $\rho_F(f) \in \{0,1\}$ for all $f \in F$. Let $\mathcal{D}_\rho$ denote the set of all generic decision-problems that are deterministic under ρ. Proposition 1.1 shows that in a partitional information model, we can assess informativeness simply by comparing deterministic decision-problems.

Proposition 1.1. *Let ρ and τ be represented by $(\mathcal{F}, u)$ and $(\mathcal{G}, u)$ respectively. Then $\mathcal{F} \subset \mathcal{G}$ iff $\mathcal{D}_\tau \subset \mathcal{D}_\rho$.*

Proof. See Appendix. □

Thus, in the special case where information correspond to events that partition the state space, more information is equivalent to less deterministic (i.e. more random) choice. This captures the intuition shared by Theorem 1.5 in this special case with partitional information. Note that Theorem 1.5 still holds in this setting. The only complication is dealing with test functions when there are non-generic decision-problems.[24]

1.6 Detecting Biases

In this section, we study situations where the valuation preference relation ($\succeq$) and the random choice rule (ρ) are inconsistent. By inconsistent, we mean that the decision-maker's information as revealed through $\succeq$ is misaligned with the decision-maker's information as revealed through ρ. We call this *detecting biases*. In the individual interpretation, this misalignment describes an agent whose *prospective* (i.e. before time 1) beliefs about information

[24] One way to resolve this issue is to define the test function of $F \in \mathcal{K}$ at $a \in [0,1]$ as $\lim_{b \downarrow a} F_\rho(b)$. Since generic decision-problems are dense, this is a well-defined cumulative distribution function. Theorem 1.5 then follows naturally.

are misaligned with her *retrospective* (i.e. after time 1) beliefs about information. This is an informational version of the naive Strotz [78] model. We call this form of informational dynamic inconsistency *subjective misconfidence.*

In the group interpretation, this misalignment describes a situation where valuations of option sets indicate a more (or less) dispersed distribution of beliefs in the group than that implied by random choice. For example, a firm that evaluates health plans based on total welfare for its employees may overestimate (underestimate) the dispersion of employee beliefs and choose a more (less) flexible set of health plans than necessary. As both the individual and group interpretations of this inconsistency are similar, for ease of exposition, we focus exclusively on the individual interpretation of random choice in this section.

To elaborate, consider an agent who expects to receive a very informative signal at time 1. Hence, at time 0, she prefers large option sets and may be willing to pay a cost in order to postpone choice and "keep her options open". When we analyze time-2 random choice however, we observe that she consistently chooses the same option. For example, in the diversification bias, although an agent may initially prefer a large option set containing a variety of food options, she ultimately always end up choosing the same food.[25] If her choice is driven by informational reasons, then we can infer from her behavior that at time 0, she anticipated a more informative signal than what she ended up receiving. This could be due to some "false hope" of better information. We call this *prospective overconfidence.*

On the flip side, there may be situations where time-2 choice behavior reflects greater confidence than that implied by time-0 preferences. To elaborate, consider an agent who expects to receive a very uninformative signal at time 1. Hence, at time 0, 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 at time 0. For example,

[25] See Read and Loewenstein [71]. Note that in our case, the uncertainty is over future beliefs and not future tastes. Nevertheless, there could be informational reasons for why one would prefer one food over another. A food recall scandal for a certain candy would be such an example.

this could be induced by some confirmatory bias where consecutive good and consecutive bad signals generate posterior beliefs that are more dispersed.[26] We call this *prospective underconfidence.*

Since beliefs in our model are purely subjective, we are silent as to whether time-0 or time-2 choice behavior is more "correct". Thus, both prospective overconfidence and underconfidence are relative comparisons involving subjective misconfidence. The theory is completely agnostic as to what beliefs *should* be. We are thus detecting a form of belief misalignment that is completely independent of the true information structure. However, as we will show in Section 7, if we had a richer data set (such as the joint data over choices and state realizations), then we could discern which period's choice behavior is correct.

Formally, we define subjective misconfidence as follows. Let the pair $(\succeq, \rho)$ denote both the valuation preference relation $\succeq$ and the RCR ρ. Recall that in the individual interpretation, $\succeq$ and ρ are the agent's choice behavior from time 0 and time 2 respectively. Motivated by Theorem 1.5, we define prospective overconfidence and underconfidence as follows.

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

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

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

We now characterize subjective misconfidence as follows.

Corollary 1.4. *Let $\succeq$ and ρ be represented by (μ, u) and (ν, u) respectively. Then the following are equivalent:*

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

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

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

[26] See Rabin and Schrag [69] for a model and literature review of the confirmatory bias.

Proof. By Corollary 1.2, $\rho_{\succeq}$ and $\succeq_{\rho}$ are represented by (μ, u) and (ν, u) respectively. The rest follows from Corollary 1.3. □

Both (1) and (2) in Corollary 1.4 are restrictions on observable behavior while (3) is an unobservable condition on the underlying information structures. Note that by Corollary 1.3, we could have equivalently defined prospective overconfidence (underconfidence) via more (less) preference for flexibility.

Corollary 1.4 allows us to order levels of prospective overconfidence and underconfidence via Blackwell's partial ordering of information structures. In other words, by studying the choice behavior of two decision-makers, we can distinguish when one is more prospectively overconfident (or underconfident) than the other. This provides a unifying methodology to measure the severity of various behavioral biases, such as the diversification and confirmatory biases.

1.7 Calibrating Beliefs

Our analysis of information so far has adopted a purely subjective treatment of beliefs. We follow in the footsteps of the traditional models of Savage [73] and Anscombe and Aumann [3] in remaining silent as to what beliefs *should* be. Thus, although our theory identifies when observed choice behavior is consistent with some information structure, it is unable to recognize when beliefs are incorrect. For example, our notions of prospective overconfidence and underconfidence in the previous section are descriptions of *subjective* belief misalignment, and we are restrained from making any statements about *objective* overconfidence and underconfidence.

In this section, we incorporate additional data about the underlying state to achieve this distinction. By studying the joint distribution over choices and state realizations, we can test whether 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. In either interpretation, we call this exercise *calibrating beliefs*.

If information is observable, then calibrating beliefs is a well-understood exercise in statistics.[27] The main result in this section allows us to calibrate beliefs even when information is not observable. For example, in the case of health insurance, an analyst may observe a correlation between choosing health insurance and ultimately falling sick. Even though information is not observable, we can analyze data on both choices (whether an agent chooses health insurance or not) and state realizations (whether an agent gets sick or not) to infer if beliefs are well-calibrated.

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

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

Unless otherwise stated, ρ in this section refers to a cRCR. For $s \in S$ and $f \in F \in \mathcal{K}$, we interpret $\rho_{s,F}(f)$ as the probability of choosing the act f in the decision-problem F conditional on the state s realizing. For example, let f and g be acts corresponding to "forgoing health insurance" and "buying health insurance" respectively. Let $s \in S$ represent the state of falling sick. In the individual interpretation, $\rho_{s,f \cup g}(f)$ is the frequency that the individual agent chooses to forgo health insurance in all the years in which she ultimately falls sick. In the group interpretation, $\rho_{s,f \cup g}(f)$ is the frequency of agents who chooses to forgo health insurance among the subgroup of agents who ultimately fall sick. The probability that $f \in F$ is chosen *and* $s \in S$ occurs is given by the product $r_s \rho_{s,F}(f)$. Since each $\rho_{s,F}$

[27] For example, see Dawid [20].

is a measure on $(H, \mathcal{H}_F)$, the measurable sets of $\rho_{s,F}$ and $\rho_{s',F}$ coincide for all s and s'. We can thus define the unconditional RCR as

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

Note that $\bar{\rho}_F(f)$ gives the unconditional probability that $f \in F$ is chosen.

The probability r in conjunction with the cRCR ρ completely specify the joint distribution over choices and state realizations. Note that the marginal distributions of this joint distribution on choices and state realizations are $\bar{\rho}$ and r respectively. In both the individual and group interpretations, this form of state-dependent choice data is easily obtainable.[28]

We now address the private information of the decision-maker. As before, we assume that this is unobservable to the analyst. Each information structure corresponds to a unique joint distribution over beliefs about S and actual state realizations. Given $s \in S$, let μ_s be the distribution of beliefs conditional on state s realizing. For example, let $s \in S$ represent the state of falling sick. In the individual interpretation, μ_s is the agent's distribution of posteriors in the years in which she falls sick. In the group interpretation, μ_s is the distribution of beliefs in the subgroup of agents who fall sick.

Let $\mu := (\mu_s)_{s \in S}$ be the mapping from states to distributions over beliefs about S. Unless otherwise stated, μ in this section refers to this mapping. We say that the cRCR ρ has an information representation iff each RCR ρ_s has an information representation for all $s \in S$. Recall that $u : \Delta X \to \mathbb{R}$ is an affine utility function. Let (μ, u) denote the mapping μ and a non-constant u.

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

Note that by Theorem 1.1, a cRCR ρ has an information representation iff for every $s \in S$, the RCR ρ_s satisfies Axioms 1.1 to 1.7. Even if a cRCR ρ has an information representation, beliefs may still not be well-calibrated. Well-calibrated beliefs require μ to be consistent with

[28] In the individual interpretation, this data can be easily obtained in experimental work (for example, see Caplin and Dean [15]). In the group interpretation, this data is also readily available (for example, see Chiappori and Salanié [17]).

the observed frequency of states r. First, define the unconditional distribution of beliefs as

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

We now formally define well-calibrated beliefs as follows.

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 asserts that μ must satisfy Bayes' rule. In other words, for each $s \in S$, μ_s is exactly the conditional distribution of posteriors implied by μ. The following is a simple example of a well-calibrated μ.

Example 1.8. Let $S = \{s_1, s_2\}$ and again, we associate each $q \in \Delta S$ with $t \in [0,1]$ such that $t = q_{s_1}$. Let $r = \left(\frac{1}{2}, \frac{1}{2}\right)$ and $\bar{\mu}$ have density $6t(1-t)$. Let μ_{s_1} and μ_{s_2} have densities $12t^2(1-t)$ and $12t(1-t)^2$ respectively. For example, in the case of health insurance, we interpret μ_{s_1} as the distribution of beliefs conditional on ultimately falling sick (i.e. s_1 occurs). Now, for $b \in [0,1]$,

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

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

Suppose that a decision-maker's cRCR ρ is represented by (μ, u) and μ is well-calibrated. In this case, choice behavior is not only consistent with an information representation but it also implies beliefs that agree with the observed joint distribution over choices and state realizations. In order to see this, let $f \in F \in \mathcal{K}$ and Q be the set of subjective beliefs that rank f higher than all other acts in F according to subjective expected utility. Since ρ_s is represented by (μ_s, u), we have $\rho_{s,F}(f) = \mu_s(Q)$ for all $s \in S$. By the definitions of $\bar{\rho}$ and $\bar{\mu}$, we also have $\bar{\rho}_F(f) = \bar{\mu}(Q)$. Now, the conditional probability that $s \in S$ occurs given that $f \in F$ is chosen is

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

Hence, the observed probabilities of state realizations conditional on f being chosen in F exactly match the implied probabilities that correspond to μ. In the health insurance example, this corresponds to the situation where the probability of getting sick conditional on buying health insurance agrees exactly with that implied by choice behavior. In the individual interpretation, this implies that the agent has rational (i.e. correct) expectations about her signals. In the group interpretation, this implies that all agents in the group have rational (i.e. correct) beliefs about their future health and so there is genuine private information in the group.

We now demonstrate how we can test for well-calibrated beliefs. Let ρ be represented by (μ, u). Since the utility u is fixed under ρ_s for all $s \in S$, both the best act $\overline{f}$ and the worst act $\underline{f}$ are well-defined for the cRCR ρ. Consider a *conditional worst act* which yields the worst act $\underline{f}$ only if a particular state occurs and otherwise yields the best act $\overline{f}$.

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

Given a state $s \in S$, the conditional worst act yields the worst payoff if the state s realizes and the best payoff otherwise. Recall that we defined test functions by using test acts $f^a = a\underline{f} + (1-a)\overline{f}$ that are mixtures between the best and worst acts. We now define *conditional test functions* by using *conditional test acts* $f_s^a = a\underline{f}^s + (1-a)\overline{f}$ that are mixtures between the best and conditional worst acts.

Definition. Given ρ, let $F_\rho^s : [0, r_s] \to [0,1]$ be the *conditional test function* of $F \in \mathcal{K}$ where for $f_s^a := a\underline{f}^s + (1-a)\overline{f}$ and $a \in [0,1]$,

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

Conditional test functions specify conditional choice probabilities as we vary the conditional test act from the best act to the conditional worst act. As in unconditional test functions, as a increases, the conditional test act becomes less attractive so F_{ρ}^{s} increases. Note that the domain of the conditional test function is scaled by a factor r_s to $[0, r_s]$. We say that F_{ρ}^{s} is *well-defined* iff $F_{\rho}^{s}(r_s) = 1$. Thus, well-defined conditional test functions are cumulative distribution functions on the interval $[0, r_s]$. Let $\mathcal{K}_s$ denote the set of decision-problems with well-defined conditional test functions. The following result allows us to determine whether beliefs are well-calibrated.

Theorem 1.6. *Let ρ be represented by (μ, u). Then μ is well-calibrated iff F_{ρ}^{s} and $F_{\bar{\rho}}$ share the same mean for all $F \in \mathcal{K}_s$ and $s \in S$.*

Proof. See Appendix. □

Theorem 1.6 equates well-calibrated beliefs with the requirement that both conditional and unconditional test functions have the same mean. It is a random choice characterization of rational beliefs. The following example illustrates.

Example 1.9. Let $S = \{s_1, s_2\}$ and again, we associate each $q \in \Delta S$ with $t \in [0,1]$ such that $t = q_{s_1}$. Following Example 1.8, let $r = \left(\frac{1}{2}, \frac{1}{2}\right)$, $\bar{\mu}$ have density $6t(1-t)$ and μ_{s_1} have density $12t^2(1-t)$. Agents can either choose a health plan g with $u \circ g = \left(\frac{1}{4}, \frac{3}{4}\right)$ or choose to forgo health insurance represented by h with $u \circ h = (0, 1)$. Let $G := \{g, h\}$. Conditional on being sick, the probability of choosing something in G over a conditional test plan h^a for $a \in [0,1]$ is

$$\mu_1 \left\{ t \in [0,1] \;\middle|\; \max\left\{1 - t,\ t\frac{1}{4} + (1-t)\frac{3}{4}\right\} \geq 1 - at \right\}$$

If we let G_{ρ_1} be this conditional probability scaled by the probability of being sick $\frac{1}{2}$, then $G_{\rho_1}(a) = 0$ for $a \leq \frac{3}{8}$, $G_{\rho_1}(a) = 1$ for $a \geq \frac{1}{2}$ and

$$G_{\rho_1}(a) = 1 + \frac{1}{2(1-4a)^3} + \frac{3}{16(1-4a)^4}$$

for $a \in \left(\frac{3}{8}, \frac{1}{2}\right)$. Now, the unconditional test function $G_{\bar{\rho}}$ satisfies $G_{\bar{\rho}}(a) = (3-2a)\,a^2$ for $a \leq \frac{1}{2}$, $G_{\bar{\rho}}(a) = (1-a)(1-4a)^2$ for $a \in \left(\frac{1}{2}, \frac{3}{4}\right)$ and $G_{\bar{\rho}}(a) = 1$ for $a \geq \frac{3}{4}$. Note that

$$\int_{[0,1]} a \; dG_{\rho_1} = \frac{29}{64} = \int_{[0,1]} a \; dG_{\bar{\rho}}$$

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

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

Finally, for completion, we summarize our results on assessing informativeness in this richer setting with cRCRs. In his original work, Blackwell [11] mentions another methodology of assessing informativeness attributed to Bohnenblust, Shapley and Sherman. This ordering concentrates on a ranking of information based on the set of state-contingent utilities admissible by any choice rule. We translate this result in our setting as follows. For $F \in \mathcal{K}$, let $\mathcal{C}_F$ denote the set of all measurable functions $c : \Delta S \to F$. In other words, each $c \in \mathcal{C}_F$ represents a specific choice rule that maps beliefs to actual choices from the decision-problem F. Given a belief $q \in \Delta S$, the act $c(q) \in F$ will be chosen under choice rule $c \in \mathcal{C}_F$. We interpret $\mathcal{C}_F$ as the set of all possible choice rules. Let $u : \Delta X \to \mathbb{R}$ be an affine utility. If the state is $s \in S$, then the state-contingent utility is

$$u \circ c_s(q) := u(c(q)(s))$$

Given a cRCR ρ that has an information representation and a decision-problem $F \in \mathcal{K}$, we let $W(\rho, F)$ denote the set of all possible conditional utilities under any choice rule.

Definition. For ρ represented by (μ, u) and $F \in \mathcal{K}$, define

$$W(\rho, F) := \bigcup_{c \in \mathcal{C}_F} \left(\int_{\Delta S} u \circ c_s(q) \, \mu_s(dq) \right)_{s \in S}$$

Blackwell shows that more information is equivalent to a larger set of admissible conditional utilities. In other words, any conditional utility attainable with less information must also be attainable with more information. We summarize these results as follows.

Corollary 1.5. *Let ρ and τ be represented by (μ, u) and (ν, u) respectively where μ and ν are well-calibrated. Then the following are equivalent:*

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

(2) $\bar{\mu}$ *is more informative than* $\bar{\nu}$

(3) $W(\tau, F) \subset W(\rho, F)$ *for all* $F \in \mathcal{K}$

Proof. See Appendix. □

In Corollary 1.5, only (1) relates to actual observable data. Both (2) and (3) relate to unobservables of the model but are useful in demonstrating that Blackwell's results translate well in our setting. We interpret (3) to imply that if a decision-maker were to follow any choice rule (which may not necessarily be the optimal choice rule as revealed through ρ) and obtains a certain conditional utility, then a decision-maker with more information can also achieve the same conditional utility. Note that Corollary 1.5 also highlights the observation that provided beliefs are rational, analyzing unconditional RCRs is sufficient for assessing informativeness via (1). On the other hand, if there is reason to suspect that beliefs may not be fully rational, then studying cRCRs is necessary for calibrating beliefs.

1.8 Random Subjective Expected Utility

In this section, we consider the general case where choice is driven by both belief and taste (i.e. risk preference) shocks. Let $\mathbb{R}^X$ be the space of all affine utility functions on ΔX and π be a measure on $\Delta S \times \mathbb{R}^X$. We interpret π as the joint distribution over beliefs and tastes. Note that the marginal distribution of π on ΔS corresponds exactly to the signal distribution μ. 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 and u is non-constant π-a.s..

A RCR ρ has a random subjective expected utility (RSEU) representation iff the following holds.

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 the choice of a decision-maker who receives unobservable shocks to both beliefs and tastes. In the group interpretation, this describes a group with heterogeneity in both beliefs and risk aversion. Note that in the special case where $\pi(\Delta S \times \{u\}) = 1$ for some non-constant $u \in \mathbb{R}^X$, this corresponds exactly to an information representation.

To characterize a RSEU representation, C-determinism must be relaxed. In particular, we also need to replace S-monotonicity with a state-by-state independence axiom below. For $f \in H$ and $\{s_1, s_2\} \subset S$, define $f_{s_1,s_2} \in H$ as the act such that $f_{s_1,s_2}(s_1) = f_{s_1,s_2}(s_2) = f(s_1)$ and $f_{s_1,s_2}(s) = f(s)$ for all $s \notin \{s_1, s_s\}$.

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

S-independence ensures that two acts that are constant over two states will be chosen for sure over an act that is non-constant over those states. This follows from the fact that non-constant acts are only chosen over constant acts when the decision-maker has state-dependent utility. This is the random choice version of the state-by-state independence axiom.[29]

Theorem 1.7 below shows that we obtain a RSEU representation if we replace C-determinism and S-monotonicity with S-independence.

Theorem 1.7. *ρ has a RSEU representation iff it satisfies Axioms 1.1-1.5 and S-independence.*

Proof. See Appendix. □

1.9 Related Literature

1.9.1 Relation to Existing Literature

This chapter is related to a long literature on stochastic choice. In particular, the information representation we introduce is a special case of RUM.[30] Testable implications of RUM were first studied by Block and Marschak [12]. The model was fully characterized by McFadden and Richter [61], Falmagne [31] and Cohen [19], who extended the exercise to arbitrary infinite sets of alternatives. Gul and Pesendorfer [47] obtain a more intuitive characterization of RUM by enriching the choice space with lotteries. More recently, Gul, Natenzon and Pesendorfer [46] characterize a special class of RUM models called attribute rules that can approximate any RUM model.

In relation to this literature, our work shows that a characterization of random expected utility can be comfortably extended to the realm of Anscombe-Aumann acts. Thus, the axioms of subjective expected utility yield nice analogs in random choice. Moreover, by

[29] 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 1.7 implies that this is equivalent to state-by-state independence axiom in the presence of the other standard axioms. Note that the definition of null states becomes unnecessary in this characterization.

[30] RUM is used extensively in discrete choice estimation. Most models in this literature assume specific parametrizations such as the logit, the probit, the nested logit, etc. (see Train [80]).

allowing our RCR to be silent on acts that are indifferent, we are able to include deterministic choice as a special case of random choice. This is an issue that most RUM models have difficulty with. Finally, one could interpret Theorem 1.3 as presenting an alternative characterization of RUM via properties of its induced valuation preference relation.

Some recent papers have also investigated the relationship between stochastic choice and information. Natenzon [64] studies a model where the decision-maker (in his model, an individual agent) gradually learns her own tastes. This is in contrast to our model where tastes are fixed and utilities vary only as a result of learning about the underlying state. Caplin and Dean [15] and Matejka and McKay [59] study cRCRs where the decision-maker (individual agents in both models) exhibits rational inattention. Ellis [26] studies a similar model with partitional information so the resulting cRCR is deterministic. In contrast, the information structure in our model is fixed. This is closer to the standard model of information processing and choice. Note that since the information structure in these other models is allowed to vary with the decision-problem, the resulting random choice model is not necessarily a RUM model. Caplin and Martin [16] do characterize and test a model where the information structure is fixed. We can recast their model in our richer Anscombe-Aumann setup, in which case our conditions for a well-calibrated cRCR imply their conditions. Note that by working with a richer setup, our representation can be uniquely identified from choice behavior.

This chapter is also related to the large literature on choice over menus (i.e. option sets). This line of research commenced with Kreps' [56] seminal paper on preference for flexibility over finite alternatives. The model was extended to the lottery space by Dekel, Lipman and Rustichini [21] (henceforth DLR) and more recently to the Anscombe-Aumann space by DLST [23]. Our main contribution to this literature is showing that there is an intimate link between ex-ante choice *over* option sets (i.e. our valuation preference relation) and ex-post random choice *from* option sets. In fact, Theorem 1.4 can be interpreted as characterizing the ex-ante valuation preference relation via properties of its ex-post random choice. Ahn and Sarver [1] also study this relationship although in the lottery space. Their work connecting DLR preferences with Gul and Pesendorfer [47] random expected utility is analogous to our results connecting DLST preferences with our random choice 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 (we elaborate on this further in the next subsection of this chapter). Fudenberg and Strzalecki [36] also analyze the relationship between preference for flexibility and random choice but in a dynamic setting with recursive random utilities. In contrast, in both Ahn and Sarver [1] and our model, the ex-ante choice over option sets is static.

Nehring [65] also studies preference for flexibility in a Savage setting where he employs Möbius inversion to characterize Kreps' representation. Although he does not address random choice, the same technical tool can be used to obtain a random utility from the ex-ante valuation preference relation in this ordinal setting. Thus, in a sense, the operations in Corollary 1.2 can be interpreted as the cardinal analogs of these tools in the richer Anscombe-Aumann space. Grant, Kajii and Polak [42, 43] also study decision-theoretic models involving information. However, they consider generalizations of the Kreps and Porteus [57] model where the decision-maker (an individual agent) has an intrinsic preference for information even when she is unable to or unwilling to act on that information. In contrast, in our model, the decision-maker prefers information only as a result of its instrumental value as in the classical sense of Blackwell.

Another strand of related literature studies the various biases in regards to information processing. This includes the confirmatory bias (Rabin and Schrag [69]), the hot-hand fallacy (Gilovich, Vallone and Tversky [41]) and the gambler's fallacy (Rabin [68]). Our model can be applied to study all these behaviors. To see this, consider the individual interpretation of our model. First assume that at time 0, the agent is immune to these biases and has rational beliefs about the signal that she will be receiving at time 1. However, after receiving her signal, she becomes afflicted with the bias and exhibits time-2 random choice that deviates from rationality. With this setup, the confirmatory bias and the hot-hand fallacy will yield prospective underconfidence while the gambler's fallacy will yield prospective overconfidence. Note that Corollary 1.4 allows us to rank the severity of these biases via the Blackwell ordering of information structures. Finally, although not necessarily about biased information processing, the diversification bias (Read and Loewenstein [71]) can also be studied in this setup.

In the strategic setting, Bergemann and Morris [8] study information structures in Bayes' correlated equilibria. In the special case where there is a single bidder, our results translate directly to their setup for a single-person game. Thus, we could interpret our model as describing the actions of a bidder assuming that the bids of everyone else are held fixed. Kamenica and Gentzkow [52] and Rayo and Segal [70] characterize optimal information structures where a sender can control the information that a receiver gets. In these models, the sender's ex-ante utility is a function of the receiver's random choice rule. Our results relating random choice with valuations thus provide a technique for expressing the sender's utility in terms of the receiver's utility and vice-versa.

This chapter is also related to the recent literature on testing for private information in insurance markets. Hendren [50] uses elicited subjective beliefs from survey data to test whether there is more private information in one group of agents (insurance rejectees) than another group (non-rejectees). Under the group interpretation, Theorem 1.5 allows us to perform this same test by inferring beliefs directly from choice data. Also, we can interpret Theorem 1.6 as providing a sufficient condition for the presence of private information that is similar to tests for private information in the empirical literature (e.g. Chiappori and Salanié [17], Finkelstein and McGarry [34] and also Hendren [50]).

Finally, there is the literature on comparisons of information structures that traces its lineage back to Blackwell [10, 11]. Lehmann [58], Persico [67], and Athey and Levin [4] all study decision-problems in restricted domains that allow for information comparisons that are strictly weaker than that of Blackwell. Although we work in a finite state space, we could extend our model and obtain restrictions on random choice that would correspond to these weaker orderings.

1.9.2 Relation to Lotteries

In this section, we relate our model with that of Ahn and Sarver [1]. For both ease of comparison and exposition, we focus on the individual interpretation of random choice in this section. In their paper, Ahn and Sarver consider choices over lotteries and introduce a condition called consequentialism to link choice behavior from the two time periods.[31] In the setup with Anscombe-Aumann acts, consequentialism translates into the following.

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.

Example 1.10. Let $S = \{s_1, s_2\}$, $X = \{x, y\}$ and $u\left(a\delta_x + (1-a)\,\delta_y\right) = a$. We associate each $q \in \Delta S$ with $t \in [0,1]$ such that $t = q_{s_1}$. Let μ have the uniform distribution and ν have density $6t\,(1-t)$. Thus, μ is more informative than ν. Let $\succeq$ be represented by (μ, u) and ρ be represented by (ν, u). We show that $(\succeq, \rho)$ satisfies consequentialism. Let $F^+ \subset F \cap G$ denote the support of $\rho_F = \rho_G$. Since $f \in F \backslash F^+$ implies it is dominated by F^+ μ-a.s., it is also dominated by F^+ ν-a.s. so $F \sim F^+$. A symmetric analysis for G yields $F \sim F^+ \sim G$. Thus, consequentialism is satisfied, but $\mu \neq \nu$.

The reason for why consequentialism fails in the Anscombe-Aumann setup is that the representation of DLR is more permissive than that of DLST. In the lottery setup, if consequentialism is satisfied, then this extra freedom allows us to construct an ex-ante representation that is completely consistent with that of ex-post random choice. On the other hand, information is uniquely identified in the representation of DLST, so this lack of flexibility

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

prevents us from performing this construction even when consequentialism is satisfied. In other words, a stronger condition is needed to perfectly equate choice behavior from the two time periods. Essentially, if the test functions of two decision-problems share the same mean, then they must be ex-ante indifferent.

Axiom (Strong Consequentialism)**.** *If F_ρ and G_ρ share the same mean, then $F \sim G$.*

The following lemma demonstrates why this is a strengthening of consequentialism.

Lemma 1.3. *For ρ monotonic, $\rho_F = \rho_G$ implies $F_\rho = G_\rho$.*

Proof. See Appendix. □

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.

Finally, we show that in the Anscombe-Aumann setup, strong consequentialism delivers the corresponding connection between ex-ante and ex-post choice behaviors that consequentialism delivered in the lottery setup.

Proposition 1.2. *Let $\succeq$ and ρ be represented by (μ, u) and (ν, v) respectively. Then $(\succeq, \rho)$ is strongly consequential iff $(\mu, u) = (\nu, \alpha v + \beta)$ for $\alpha > 0$.*

Proof. See Appendix. □

Note that in light of Corollary 1.2, we could have equivalently defined strong consequentialism using the induced preference relation $\succeq_\rho$. In other words, if the integrals of F_ρ and G_ρ are the same, then $F \sim G$. This follows immediately from the fact that the integral of F_ρ is just one minus the mean of F_ρ.

Appendix 1A

1A.1. Representation Theorem

In this section of Appendix 1A, we prove the main representation theorem. Given a non-empty collection $\mathcal{G}$ of subsets of H and some $F \in \mathcal{K}$, define

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

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

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

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

Lemma (1A.1)**.** *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 a RCR. By Lemma 1A.1, we can now define

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

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

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

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

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

Lemma (1A.2)**.** *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 1A.1, $f \cup g = f_{f \cup g} = f_F \cap (f \cup g)$ 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 1A.1, $g_F \cap f_F \in \mathcal{H}_F$ 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 1A.1 again,

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

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

Lemma (1A.3)**.** *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 1A.2, we can find unique $h^i \in F$ for $i \in \{1,\dots,k\}$ such that $\left\{h_F^1, \dots h_F^k\right\}$ forms a partition on F. Without loss of generality, assume g is tied with some $g' \in h_F^1$. By Lemma 1A.2 again, $h_{F\cup g}^1 = h_F^1 \cup g$ and $h_{F\cup g}^i = h_F^i$ for $i > 1$. By monotonicity, for all i

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

Now, for any $f \in h_F^j$, $f \in h_{F\cup g}^j$ and

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

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 $FaF' := aF + (1-a)F'$.

Lemma (1A.4)**.** *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 1A.2 and $f_F ah \subset f'_{F'}$. Now, let $g' \in f'_{F'}$ so $g' = gah$ is tied with fah. By linearity again, f and g are tied so $g' \in f_F ah$. Thus, $f'_{F'} = f_F ah$. □

We now associate each act $f \in H$ with the vector $f \in [0,1]^{S \times X}$ without loss of generality. Find $\{f_1, g_1, \ldots, f_k, g_k\} \subset H$ such that $f_i \neq g_i$ are tied and $z_i \cdot z_j = 0$ for all $i \neq j$, where $z_i := \frac{f_i - g_i}{\|f_i - g_i\|}$. Let $Z := \text{lin}\{z_1, \ldots, z_k\}$ 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 1A.3 and 1A.4 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 1A.5 below shows that φ projects H onto a space without ties.

Lemma (1A.5)**.** *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

$$f - \sum_{1 \le i \le k} (f \cdot z_i) z_i = g - \sum_{1 \le i \le k} (g \cdot z_i) z_i$$

$$f - g = \sum_{1 \le i \le k} ((f - g) \cdot z_i) z_i \in Z$$

and f and g are tied.

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

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

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

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

for all $w \in W$.

□

Lemma (1A.6)**.** *If ρ satisfies Axioms 1.1-1.4, then there exists a measure ν on W such that*

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

Proof. Let $m := \dim(W)$. Note that if $m = 0$, then $W = \varphi(H)$ is a singleton so everything is tied by Lemma 1A.5 and the result follows trivially. Thus, assume $m \ge 1$ and let $\Delta^m \subset \mathbb{R}^{S \times X}$ be the m-dimensional probability simplex. Now, there exists an affine transformation $T = \lambda A$ where $\lambda > 0$, A is an orthogonal matrix and $T \circ \varphi(H) \subset \Delta^m$. Let $V := \text{lin}(\Delta^m)$ so $T(W) = V$. Now, for each finite set $D \subset \Delta^m$, we can find a $p^* \in \Delta^m$ and $a \in (0, 1)$ such that $Dap^* \subset T \circ \varphi(H)$. Thus, we can define a RCR τ on Δ^m such that

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

where $T \circ \varphi(F) = Dap^*$ and $T \circ \varphi(f) = pap^*$. Linearity and Lemma 1A.5 ensure that τ is well-defined.

Since the projection mapping φ is linear, Axioms 1.1-1.4 correspond exactly to the axioms of Gul and Pesendorfer [47] on Δ^m. Thus, by their Theorem 3, there exists a measure ν_T on V such that for $F \in \mathcal{K}_0$

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

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

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

Thus,

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

where $\nu := \nu_T \circ T$ is the measure on W induced by T. Note that the last equality follows from Lemma 1A.5.

Finally, consider any generic $F \in \mathcal{K}$, and let $F_0 \subset F$ be such that $f \in F_0 \in \mathcal{K}_0$. By Lemma 1A.3,

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

By Lemma 1A.5, if h and g are tied, then

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

for all $w \in W$. Thus,

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

□

Henceforth, assume ρ satisfies Axioms 1.1-1.4 and let ν be the measure on W as specified by Lemma 1A.6. We let $w_s \in \mathbb{R}^X$ denote the vector corresponding to $w \in W$ and $s \in S$. For $u \in \mathbb{R}^X$, define $R(u) \subset \mathbb{R}^X$ as the set of all $\alpha u + \beta \mathbf{1}$ for some $\alpha > 0$ and $\beta \in \mathbb{R}$. Let $U := \left\{ u \in \mathbb{R}^X \middle| u \cdot \mathbf{1} = 0 \right\}$ and note that $R(u) \cap U$ is the set of all αu for some $\alpha > 0$.

A state $s^* \in S$ is null iff it satisfies the following.

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

Lemma (1A.7)**.** *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 1A.2, f and g are tied. This implies $\rho(f, g) = 1$ for all $\{f, g\} \subset H$ contradicting non-degeneracy so there must exist at least one non-null state. □

Lemma (1A.8)**.** *Let ρ satisfy Axioms 1.1-1.4 and S-independence. Suppose $\{s_1, s_2\} \subset S$ are non-null. Define $\phi : W \to U \times U$ such that*

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

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

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

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

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

Proof. We prove the lemma in order.

(1) Since s_1 is non-null, we can find $\{f, g\} \subset H$ such that $f_s = g_s$ for all $s \neq s_1$ and f and

g are not tied. Let $f_{s_1} = p$ and $g_{s_1} = q$ so

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

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

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

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

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

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

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

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

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

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

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

and let $\mathcal{E}$ be the set of all finite intersection of such halfspaces. Consider a partition $\mathcal{P} = \{0\} \cup \bigcup_i A_i$ of U where for each A_i, we can find two sequences $A_{ij} \in \mathcal{E}$ and $\bar{A}_{ij} \in \mathcal{E}$ such that $A_{ij} \nearrow A_i \cup \{0\}$, $A_{ij} \subset \text{int}\left(\bar{A}_{ij}\right) \cup \{0\}$ and $\bar{A}_{ij} \cap \bar{A}_{i'j} = \{0\}$ for all $i' \neq i$. Note that since sets in $\mathcal{E}$ are η-measurable, every $A_{ij} \times A_{i'j'}$ is η-measurable. By (1)

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

By a standard separating hyperplane argument (Theorem 1.3.8 of Schneider [75]), we can find some $r \in U$ such that $u_1 \cdot r \geq 0 \geq u_2 \cdot r$ for all $(u_1, u_2) \in \bar{A}_{ij} \times \bar{A}_{i'j}$. Since $A_{ij} \backslash \{0\} \subset \text{int}\left(\bar{A}_{ij}\right)$, we must have $u_1 \cdot r > 0 > u_2 \cdot r$ for all $(u_1, u_2) \in (A_{ij} \backslash \{0\}) \times (A_{i'j} \backslash \{0\})$. By (1) and (2),

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

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

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

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

$C_0 = \bigcap_k C_k$ so

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

□

Theorem (1A.9)**.** *If ρ satisfies Axioms 1.1-1.5 and S-independence, then it has a RSEU representation.*

Proof. We begin with the first statement. Let ρ satisfy Axioms 1.1-1.5 and S-independence, and ν be the measure on W as specified by Lemma 1A.6. Let $S^* \subset S$ be the set of non-null states with some $s^* \in S^*$ as guaranteed by Lemma 1A.7. Define

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

and note that by Lemma 1A.8,

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

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

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

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

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

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

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

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

Thus

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

Finally, we show that π is regular. Suppose $\exists \{f,g\} \subset H$ such that

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

If f and g are tied, then $q \cdot (u \circ f) = q \cdot (u \circ g)$ π-a.s. yielding a contradiction. Since f and g are not tied, then

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

a contradiction. Lemma 1A.8 implies that u is non-constant π-a.s. so π is regular. Thus, ρ is represented by π. □

Theorem (1A.10)**.** *If ρ has a RSEU representation, then it satisfies Axioms 1.1-1.5 and S-independence.*

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,

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

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 1A.2 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 [18]).

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 $B \in \mathcal{B}(H)$,

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

Hence, ξ_k and ξ are random variables. Moreover,

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

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 [18]), $\rho_{F_k} \to \rho_F$ and continuity is satisfied. □

Corollary (1A.11)**.** *ρ satisfies Axioms 1.1-1.7 iff it has an information representation.*

Proof. We first prove sufficiency. Note that if ρ satisfies Axioms 1.1-1.6 and S-independence, then by Theorem 1A.9, ρ has an information representation. We show that Axioms 1.1-1.7 imply 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. Note that if h_{s_1} and h_{s_2} are tied, then S-monotonicity implies h is tied to both, so assume $\rho(h_{s_1}, h_{s_2}) = 1$ without loss of generality. By S-monotonicity again, $\rho(f,h) = 1$ implying $\rho(h,f) = 0$. Thus, $\rho_{\{f,g,h\}}(h) = 0$ so S-independence is satisfied.

For necessity, note that Axioms 1.1-1.5 all follow from Theorem 1A.10. C-determinism follows trivially from the representation. To show S-monotonicity, suppose $\rho_{F_s}(f_s) = 1$ for all $s \in S$. Thus, $u(f_s) \geq u(g_s)$ for all $g \in F$ and $s \in S$ which implies $q \cdot (u \circ f) \geq q \cdot (u \circ g)$ for all $g \in F$. Hence, $\rho_F(f) = 1$ from the representation yielding S-monotonicity. □

1A.2. Uniqueness

In this section of Appendix 1A, we use test functions to prove the uniqueness properties of information representations. Let $H_c \subset H$ denote the set of all constant acts.

Lemma (1A.12)**.** *Let ρ be represented by (μ, u). Then for any measurable $\phi : \mathbb{R} \to \mathbb{R}$,*

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

Proof. For $F \in \mathcal{K}$, let $\psi_F : \Delta S \to [0,1]$ be such that $\psi_F(q) = \frac{u(\overline{f}) - \sup_{f \in F} q \cdot (u \circ f)}{u(\overline{f}) - u(\underline{f})}$ which is measurable. Let $\lambda^F := \mu \circ \psi_F^{-1}$ be the image measure on $[0,1]$. By a standard change of variables (Theorem I.5.2 of Çinlar [18]),

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

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,

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

First, assume f^a is tied with nothing in F. Since μ is regular,

$$\mu\left\{q \in \Delta S | \, u(f_a) = q \cdot (u \circ f)\right\} = 0$$

for all $f \in F$. Thus,

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

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

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

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

For convenience, we define the following.

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

Lemma (1A.13)**.** *Let ρ and τ be represented by (μ, u) and (ν, v) respectively. Then the following are equivalent:*

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

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

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

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

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

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

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

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

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

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

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

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

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

Theorem (1A.14)**.** *Let ρ and τ be represented by (μ, u) and (ν, v) respectively. Then the following are equivalent:*

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

(2) $\rho = \tau$

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

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

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

Hence, all that remains is to prove that (4) implies (1). Assume (4) is true so $f_\rho = f_\tau$ for all $f \in H$. By Lemma 1A.13, this implies $u = \alpha v + \beta$ for $\alpha > 0$. Thus, without loss of generality, we can assume $1 = u(\overline{f}) = v(\overline{f})$ and $0 = u(\underline{f}) = v(\underline{f})$ so $u = v$. Now,

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

where $\psi_f : \Delta S \to [0, 1]$. Let $\lambda_\rho^f = \mu \circ \psi_f^{-1}$ and $\lambda_\tau^f = \nu \circ \psi_f^{-1}$, so by the lemma above, they correspond to the cumulatives f_ρ and f_τ. Now, by Ionescu-Tulcea's extension (Theorem IV.4.7 of Çinlar [18]), 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 1A.12,

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

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 [18]), X and Y have the same distribution, so $\mu = \nu$. Thus, $(\mu, u) = (\nu, \alpha v + \beta)$ for $\alpha > 0$ and (1) is true. Hence, (1) to (4) are all equivalent. □

Lemma 1A.15 below shows that (1) every decision-problem is arbitrarily (Hausdorff) close to some decision-problem in $\mathcal{K}_0$, and (2) ρ is discontinuous at precisely those decision-problems that contain ties (indifferences).

Lemma (1A.15)**.** *Let ρ have an information representation.*

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

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

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

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

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

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

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

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

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

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

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

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

and $\bar{Z} := \text{lin}\left(\bigcup_j z_j\right) \cap [0,1]^S$. Let $z := u \circ f - u \circ g$ and since f and g are not tied, $z \notin \bar{Z}$ by linearity. Thus, z and $\bar{Z}$ can be strongly separated (see Theorem 1.3.7 of Schneider [75]), but $z_j \to z$ yielding a contradiction. Hence, there is some $m \in \mathbb{N}$ such that f_k and g_k are not tied for all $k > m$. Continuity yields $\rho(f_k, g_k) \to \rho(f, g)$.

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

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

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

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

violating continuity.

$\square$

Appendix 1B

In Appendix 1B, we prove our results relating valuations with random choice. In this section, consider RCRs ρ such that there are $\{\underline{f}, \bar{f}\} \subset H_c$ where $\rho(\overline{f}, f) = \rho(f, \underline{f}) = 1$ 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 $f^a := \underline{f} a \overline{f}$.

Lemma (1B.1)**.** *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 [9], we have

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

The result then follows immediately. □

Lemma (1B.2)**.** *For cumulatives F and G on $[0,1]$, $F = G$ iff $F = G$ a.e..*

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

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

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

Lemma (1B.3)**.** *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 $\rho(\underline{f}, \overline{f}) > 0$, then $\underline{f}$ and $\overline{f}$ are tied so by Lemma 1A.3, $\rho(\underline{f}, f) = \rho(f, \underline{f}) = 1$ 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 1A.2, $f^a_G=f^a_{F\cup f^a}=f^a$ 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 1A.3,

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

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 (1B.4)**.** *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 1A.14, we can assume u is normalized without loss of generality. Let

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

so V represents $\succeq_\rho$. Since test functions are well-defined under ρ, let $\overline{f}$ and $\underline{f}$ be the best and worst acts respectively. We first show that $\rho\left(\underline{f}, \overline{f}\right) = 0$. Suppose otherwise so $\underline{f}$ and $\overline{f}$ must be tied. By Lemma 1A.4, f^b and f^a are tied for all $\{a, b\} \subset [0,1]$. Thus, $f^b(a) = 1$ for all $\{a, b\} \subset [0,1]$. Hence $V_\rho\left(f^b\right) = V_\rho\left(f^a\right)$ 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

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

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

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

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

so $\phi(a) = a$ for all $a \in [0,1]$. Now, by Lemmas 1A.12 and 1B.1,

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

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

By Lemma 1B.3, for all $b \in [0,1]$,

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

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

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

Let λ be the measure corresponding to G so $\lambda[0,b] = G(b)$. Thus, by the Radon-Nikodym Theorem (see Theorem I.5.11 of Çinlar [18]), we have a.e.

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

Lemma 1B.2 then establishes that $F_\rho = F_\tau$ for all $F \in \mathcal{K}$. □

Lemma (1B.5)**.** *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 1A.3, 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

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

Suppose f^b is not ρ-tied with g. Thus, $\rho\left(f^b, f\right) = 0$. Now, for $a < b$, $\rho(f, f^a) = 0$ implying $\rho(f^a, f) = 1$. 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 1A.4. 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 1A.2.

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

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

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

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

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 1A.3, 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,

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

Now, if f^a is tied with some act in G, then let $F' := F \backslash f_G^a$. By Lemma 1A.3, $\rho_G(f) = \rho(f, F')$ and $\tau_G(f) = \tau(f, F')$ where f is tied with nothing in F'. Applying the above on

F' yields $\rho_G(f) = \tau_G(f)$ for all $f \in G \in \mathcal{K}$. Hence, $\rho = \tau$. $\square$

Theorem (1B.6)**.** *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

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

so from Lemmas 1A.12 and 1B.1,

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

so (2) is true. Now, suppose (2) is true and let τ be represented by (μ, u) with u normalized. By Lemma 1B.4, $F_\rho = F_\tau$ for all $F \in \mathcal{K}$. By Lemma 1B.5, $\rho = \tau$ so (1) is true. $\square$

Lemma (1B.7)**.** *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 $V\left(\overline{f}\right)=1\geq V\left(F\right)\geq 0=V\left(\underline{f}\right)$ 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 $F\sim F\cup\underline{f}$.

□

Lemma (1B.8)**.** *Let ρ be monotone, linear and $\rho\left(\underline{f},\overline{f}\right)=0$. Then a.e.*

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

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 1A.2, 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 1B.3,

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

Since $V_\rho(F \cup f_0) = \int_{[0,1]} F_\rho(1-a)\, da$, we have

$$V_\rho(F \cup f_b) - V_\rho(F \cup f_0) = b - \int_{[0,b]} F_\rho(1-a)\, da$$
$$= \int_{[0,b]} (1 - F_\rho(1-a))\, da$$

Thus, we have a.e.

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

□

Theorem (1B.9)**.** *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

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

First, suppose (1) 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 $f_a := a\overline{f} + (1-a)\underline{f}$. Since V also represents $\succeq$, $W = \phi \circ V$ for some monotonic $\phi : \mathbb{R} \to \mathbb{R}$. By Lemma 1B.7, $\overline{f} \succeq F \succeq \underline{f}$ so $u(\overline{f}) \geq u(f) \geq u(\underline{f})$ 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 1A.12 and 1B.1,

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

so by Lemma 1B.8, $\tau(f_a, F) = \frac{dV(F \cup f_a)}{da}$.

Suppose $\rho(\underline{f}, \overline{f}) > 0$ so $\underline{f}$ and $\overline{f}$ are ρ-tied. Thus, by Lemma 1A.2, $\rho(\underline{f}, f) = \rho(f, \underline{f}) = 1$ 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 1B.7 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 $b \geq 0$,

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

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

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

so $W = V$. By Lemma 1B.8, we have a.e.

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

so $F_\tau = F_\rho$ a.e.. By Lemma 1B.2, $F_\tau = F_\rho$ so by Lemma 1B.5, $\rho_{\succeq} = \rho = \tau$ and (2) holds.

Now, suppose (2) is true and let $\rho = \rho_{\succeq}$ where $W : \mathcal{K} \to [0,1]$ represents $\succeq$ and $\rho\left(f_a, F\right) = \frac{dW(F \cup f_a)}{da}$ for $f_a := a\overline{f} + (1-a)\underline{f}$. Suppose ρ is represented by (μ, u) and since $V_\rho = V$, we have $\rho\left(f_a, F\right) = \frac{dV(F \cup f_a)}{da}$ by Lemma 1B.8. Now, by dominance,

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

so $W = V$ proving (1). ☐

Lemma 1B.10 shows that test functions have at most a single point of discontinuity.

Lemma (1B.10)**.** *Let ρ be represented by (μ, u). Then for $a \in [0,1]$, the following are equivalent:*

(1) *f^a is tied with some $f \in F$ where $\rho_F\left(f\right) > 0$*

(2) *F_ρ is discontinuous at $a \in [0,1]$*

(3) $V_\rho (F \cup f^a)$ *is not differentiable at* a

Proof. We first show that (1) and (2) are equivalent. First, fix $F \in \mathcal{K}$ and let $U := u \circ F$. Let λ^F be the distribution corresponding to F_ρ. Thus, from Lemma 1A.12, $\lambda^F = \mu \circ \psi_F^{-1}$ where

$$\psi_F (q) = 1 - h (U, q)$$

and $h (U, q) := \sup_{f \in F} q \cdot (u \circ f)$ denotes the support function of U at $q \in \Delta S$. Let

$$F^+ := \{ f \in F | \, \rho_F (f) > 0\}$$

Now, for any $h \in H_c$, let $a = u (h)$ and

$$Q_a := \psi_F^{-1} (1 - a) = \{q \in \Delta S| \, h (U, q) = a\}$$

For $f \in F^+$, let

$$Q_a (f) : = \{ q \in \Delta S| \, q \cdot (u \circ f) = h (U, q) = a\}$$

Clearly, $\bigcup_{f \in F_+} Q_a (f) \subset Q_a$. Note that

$$\begin{aligned} \mu \left(Q_a \backslash \bigcup_{f \in F_+} Q_a (f) \right) &\leq \mu \left\{ q \in \Delta S| \, q \cdot (u \circ f) < h (U, q) = a \;\forall f \in F^+ \right\} \\ &\leq \mu \left\{ q \in \Delta S| \, q \cdot (u \circ g) \geq h (U, q) \text{ for some } g \in F \backslash F^+ \right\} \\ &\leq \sum_{g \in F \backslash F^+} \rho_F (g) = 0 \end{aligned}$$

Hence $\mu (Q_a) = \mu \left(\bigcup_{f \in F_+} Q_a (f) \right)$. Thus, for $1 - a \in [0, 1]$,

$$\lambda^F \{1 - a\} = \mu \left(\psi_F^{-1} (1 - a) \right) = \mu (Q_a) = \mu \left(\bigcup_{f \in F^+} Q_a (f) \right)$$

First, suppose (2) is true, so $\lambda^F \{a\} > 0$. Thus, there must be some $f \in F^+$ such that

$$0 < \mu (Q_{1-a} (f)) = \mu \{ q \in \Delta S| \, q \cdot (u \circ f) = h (U, q) = 1 - a\}$$

However, since $u(f^a) = u\left(\underline{f}a\overline{f}\right) = 1 - a$, we have

$$\mu\{q \in \Delta S|\, q \cdot (u \circ f) = u(f^a)\} > 0$$

which implies f^a is tied with $f \in F^+$ as μ is regular. Thus, (1) holds.

Now suppose (1) is true. Thus,

$$\begin{aligned}\mu(Q_{1-a}(f)) &= \mu\{q \in \Delta S|\, q \cdot (u \circ f) = h(U, q) = 1 - a\} \\ &= \mu\{q \in \Delta S|\, q \cdot (u \circ f) = h(U, q)\} = \rho_F(f) > 0\end{aligned}$$

Hence,

$$\lambda^F\{a\} = \mu\left(\bigcup_{f \in F^+} Q_{1-a}(f)\right) \geq \mu(Q_{1-a}(f)) > 0$$

so F_ρ is discontinuous at $a \in [0,1]$ and (2) holds.

Note that the equivalence of (1) and (2) imply that F_ρ cannot be discontinuous at more than one point. Assume otherwise so f^a is tied with some $f \in F$ and f^b is tied with some $g \in F$ for some $\{a, b\} \subset [0,1]$ where $\rho_F(f) > 0$ and $\rho_F(g) > 0$. By Lemma 1A.3 and monotonicity, $\rho\left(f^a, f^b\right) > 0$ and $\rho\left(f^b, f^a\right) > 0$ so f^a and f^b are tied. Thus, $a = b$. Now, by Lemma 1B.3,

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

so $\left(F \cup f^b\right)_\rho$ is continuous at every point other than $b \in [0,1]$. Hence, by Theorem 7.11 of Rudin [72]

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

is differentiable at every point other than b. Thus, (1), (2) and (3) are all equivalent. $\square$

Appendix 1C

1C.1. Assessing Informativeness

In this section of Appendix 1C, we prove our result on assessing informativeness.

Theorem (1C.1)**.** *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 $q \in \Delta S$.[32] Let $\phi : \mathbb{R} \to \mathbb{R}$ be increasing concave, and note that by Lemma 1A.12,

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

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

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

[32] See Theorem 1.7.5 of Schneider [75] for elementary properties of support functions.

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

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

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

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

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

Thus,

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

for all $F \in \mathcal{K}$. Hence, by Blackwell [10, 11], μ is more informative than ν . □

Lemma (1C.2)**.** *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 1A.12,

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

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

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

so $f_\rho =_m f_\tau$ for all $f \in H$. Now assume $f_\rho =_m f_\tau$ for all $f \in H$ so by Lemma 1A.13,

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

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

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

$$\begin{aligned}1 - r_\mu \cdot (u \circ f) &= 1 - r_\nu \cdot (u \circ f) \\ 0 &= (r_\mu - r_\nu) \cdot (u \circ f)\end{aligned}$$

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

Lemma 1C.3 below shows that our definition of "more preference for flexibility than" coincides with that of DLST.

Lemma (1C.3)**.** *Let $\succeq_1$ and $\succeq_2$ have subjective learning representations. Then $\succeq_1$ has more preference for flexibility than $\succeq_2$ iff $F \succ_2 f$ implies $F \succ_1 f$.*

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

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

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

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

1C.2. Partitional Information Representations

In this section of Appendix 1C, we consider partitional information representations. Given an algebra $\mathcal{F}$ on S, let $Q_{\mathcal{F}}(S) := \bigcup_{s \in S} Q_{\mathcal{F}}(s)$. For each $q \in Q_{\mathcal{F}}(S)$, let

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

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

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

Lemma (1C.4)**.** *$\sigma(\mathcal{P}_{\mathcal{F}}) = \mathcal{F}$.*

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

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

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

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

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

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

Suppose $\exists s \in E \backslash A$ for some $E \in \mathcal{P}_A$. Thus, we can find an $s' \in A \cap E$ so $Q_{\mathcal{F}}(s) = Q_{\mathcal{F}}(s')$. However, $Q_{\mathcal{F}}(s', A) = 1 > 0 = Q_{\mathcal{F}}(s, A)$ a contradiction. Thus, $A = \overline{E} \in \sigma(\mathcal{P}_{\mathcal{F}})$ so $\mathcal{P}_{\mathcal{F}} \subset \mathcal{F} \subset \sigma(\mathcal{P}_{\mathcal{F}})$. This proves that $\sigma(\mathcal{P}_{\mathcal{F}}) = \mathcal{F}$ (see Exercise I.1.10 of Çinlar [18]). $\square$

Lemma (1C.5)**.** *$E_q^{\mathcal{F}} = \{ s \in S | \, q_s > 0 \}$ for $q \in Q_{\mathcal{F}}(S)$.*

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

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

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

Lemma (1C.6)**.** *ext*$(C_{\mathcal{F}}) \subset Q_{\mathcal{F}}(S)$.

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

Proposition (1C.7)**.** *Let ρ and τ be represented by $(\mathcal{F}, u)$ and $(\mathcal{G}, u)$ respectively. Then the following are equivalent:*

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

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

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

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

First, suppose (1) is true but $C_{\mathcal{F}} \not\subset C_{\mathcal{G}}$ and let $p \in C_{\mathcal{F}} \backslash C_{\mathcal{G}}$. Note that $C_{\mathcal{G}}$ is compact (see Theorem 1.1.10 of Schneider [75]). Thus, by a separating hyperplane argument (Theorem 1.3.4 of Schneider [75]), there is a $a \in \mathbb{R}$, $\varepsilon > 0$ and $v \in \mathbb{R}^S$ such that for all $q \in C_{\mathcal{G}}$,

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

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

and $Q_{\mathcal{G}}(S)$ are finite, we can find

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

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

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

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

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

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

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

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

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

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

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

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

Now, assume (2) is true. Let $q \in Q_{\mathcal{F}}(S)$ so $q \in C_{\mathcal{F}} \subset C_{\mathcal{G}}$. Since $\operatorname{ext}(C_{\mathcal{G}}) \subset Q_{\mathcal{G}}(S)$ from Lemma 1C.6, Minkowski's Theorem (Corollary 1.4.5 of Schneider [75]) yields that $q = \sum_i \alpha_i p^i$ for $\alpha_i > 0$, $\sum_i \alpha_i = 1$ and $p^i = Q_{\mathcal{G}}(s_i)$. Note that by Lemma 1C.5, $\sum_{s \in E_q^{\mathcal{F}}} q_s = 1$.

If $p^i_s > 0$ for some $s \notin E^{\mathcal{F}}_q$, then $q_s > 0$ a contradiction. Thus, $\sum_{s \in E^{\mathcal{F}}_q} p^i_s = 1$ for all p^i. Now, by Lemma 1C.5 again, for each p^i,

$$E^{\mathcal{G}}_{p^i} = \left\{ s \in S | \, p^i_s > 0 \right\} \subset E^{\mathcal{F}}_q$$

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

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

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

Thus, $\mathbb{E}_{\mathcal{G}} \left[u \circ f - u \circ g \right] = 0$ or $\mathbb{E}_{\mathcal{G}} \left[u \circ f - u \circ g \right] \neq 0$. Since $\mathcal{F} \subset \mathcal{G}$, by repeated conditioning (see Theorem IV.1.10 of Çinlar [18]),

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

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

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

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

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

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

Appendix 1D

In Appendix 1D, we prove our results for calibrating beliefs.

Lemma (1D.1)**.** *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* $F \in \mathcal{K}_s$,

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

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

(1) Note that

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

Thus, $q_s > 0$ μ_s-a.s..

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

$$\int_{\mathbb{R}} x \lambda_s^F(dx) = \int_{\Delta S} \psi_F^s(q) \, \mu_s(dq)$$

Note that by (1), the right integral is well-defined. We now show that the cumulative distribution function of λ_s^F is exactly F_ρ^s. For $a \in [0,1]$, let $f_s^a := \underline{f}^s a \overline{f}$ and first assume f_s^a is tied with nothing in F. Thus,

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

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

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

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

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

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

□

Lemma (1D.2)**.** *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.

□

Theorem (1D.3)**.** *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 1D.1. 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 1D.1 and 1D.2, 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 $F_a := \left(G a \overline{f} \right) \cup \underline{f}^s$ for $a \in (0,1)$. Since $\underline{f}^s \in F$, $\rho_s \left(F_a, \underline{f}^s \right) = 1$ so $F_a \in \mathcal{K}_s$. Let

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

and note that

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

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

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

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

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

Note that $q_s > 0$ $\bar{\mu}$-a.s. by Lemma 1D.2, so by dominated convergence

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

For $q\in Q_a^c$,

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

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 $Q \in \mathcal{B}(\Delta S)$ and $s \in S_+$.

Finally, for $s \notin S_+$, $\rho_s\left(\underline{f}^s, \overline{f}\right) = 1$ so $q_s = 0$ μ_s-a.s.. By Lemma 1D.2, $q_s = 0$ μ-a.s.. Let

$$Q_0 := \left\{ q \in \Delta S \;\middle|\; \sum_{s \notin S_+} q_s = 0 \right\}$$

and note that $\mu(Q_0) = 1$. Now,

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

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

Theorem (1D.4)**.** *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 $S_+ := \left\{ s \in S \middle| \rho_s\left(\underline{f}^s, \overline{f}\right) = 0 \right\} \subset S$. Thus, $s \notin S_+$ implies $\underline{f}^s$ and $\overline{f}$ are tied and $q_s = 0$ a.s. under all measures. By the same argument as the sufficiency proof above, letting $Q_0 := \left\{ q \in \Delta S \middle| \sum_{s \notin S_+} q_s = 0 \right\}$ yields

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

a contradiction. Thus, $S_+ = S$.

Let $F \in \mathcal{K}_s$ and $s \in S$. Since $\rho_s\left(\underline{f}^s, \overline{f}\right) = 0$, by Lemmas 1A.12 and 1D.1 and the fact that μ is well-calibrated,

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

so $F^s_\rho =_m F_{\bar{\rho}}$. □

Corollary (1D.5)**.** *Let ρ and τ be represented by (μ, u) and (ν, u) respectively where μ and ν are well-calibrated. Then the following are equivalent:*

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

(2) $\bar{\mu}$ *is more informative than* $\bar{\nu}$

(3) $W(\tau, F) \subset W(\rho, F)$ *for all* $F \in \mathcal{K}$

Proof. Note that the equivalence of (1) and (2) follows directly from Theorem 1C.1. We now show the equivalence of (2) and (3). For $F \in \mathcal{K}$, let $c^* : \Delta S \to F$ such that for $q \in \Delta S$,

$$q \cdot (u \circ c^*)(q) \geq q \cdot v$$

for all $v \in u \circ F$. Note that $c^* \in \mathcal{C}_F$. Note that since μ is well-calibrated, for $c \in \mathcal{C}_F$,

$$\begin{aligned} \sum_s r_s \int_{\Delta S} u \circ c_s(q)\, \mu_s(dq) &= \int_{\Delta S} \sum_s r_s u(c_s(q)) \frac{q_s}{r_s} \bar{\mu}(dq) \\ &= \int_{\Delta S} q \cdot (u \circ c)(q)\, \bar{\mu}(dq) \end{aligned}$$

Now

$$\begin{aligned} \sum_s r_s \int_{\Delta S} u \circ c_s^*(q)\, \mu_s(dq) &= \int_{\Delta S} q \cdot (u \circ c^*)(q)\, \bar{\mu}(dq) \\ &\geq \int_{\Delta S} q \cdot (u \circ c)(q)\, \bar{\mu}(dq) = \sum_s r_s \int_{\Delta S} u \circ c_s(q)\, \mu_s(dq) \end{aligned}$$

for all $c \in \mathcal{C}_F$. Since $W(\rho, F)$ is closed and convex,

$$\begin{aligned} \sup_{w \in W(\rho, F)} r \cdot w &= \sup_{c \in \mathcal{C}_F} \sum_s r_s \int_{\Delta S} u \circ c_s(q)\, \mu_s(dq) \\ &= \int_{\Delta S} q \cdot (u \circ c^*)(q)\, \bar{\mu}(dq) = \int_{\Delta S} \sup_{f \in F} q \cdot (u \circ f)\, \bar{\mu}(dq) \end{aligned}$$

First, suppose (3) is true so $W(\tau, F) \subset W(\rho, F)$ for all $F \in \mathcal{K}$. Thus,

$$\int_{\Delta S} \sup_{f \in F} q \cdot (u \circ f)\, \bar{\mu}(dq) = \sup_{w \in W(\rho, F)} r \cdot w \geq \sup_{w \in W(\tau, F)} r \cdot w = \int_{\Delta S} \sup_{f \in F} q \cdot (u \circ f)\, \bar{\nu}(dq)$$

which implies (2) via Theorem 1C.1. Now, suppose (2) is true. Let $T : [0,1]^S \to \mathbb{R}^S$ be an affine transformation and for $F \in \mathcal{K}$, let $G \in \mathcal{K}$ and $\{a, b\} \subset (0,1)$ such that $u \circ G = aT(u \circ F) + (1-a)\{b\mathbf{1}\}$. Now,

$$\begin{aligned} W(\rho, G) &= \bigcup_{c \in \mathcal{C}_G} \left(\int_{\Delta S} u \circ c_s(q)\, \mu_s(dq) \right)_{s \in S} \\ &= \bigcup_{c \in \mathcal{C}_F} \left(\int_{\Delta S} \left[aT_s(u \circ c(q)) + (1-a) b \right] \mu_s(dq) \right)_{s \in S} \\ &= aT(W(\rho, F)) + (1-a)\{b\mathbf{1}\} \end{aligned}$$

Thus,

$$\begin{aligned} \sup_{w' \in W(\rho, G)} r \cdot w' &= \sup_{w \in W(\rho, F)} r \cdot (aT(w) + (1-a) b\mathbf{1}) \\ &= a \left(\sup_{w \in W(\rho, F)} T(r) \cdot w \right) + (1-a) b \end{aligned}$$

By Theorem 1C.1 again,

$$\begin{aligned} a \left(\sup_{w \in W(\rho, F)} T(r) \cdot w \right) + (1-a) b &= \sup_{w' \in W(\rho, G)} r \cdot w' \\ &\geq \sup_{w' \in W(\tau, G)} r \cdot w' = a \left(\sup_{w \in W(\tau, F)} T(r) \cdot w \right) + (1-a) b \end{aligned}$$

so $\sup_{w \in W(\rho, F)} T(r) \cdot w \geq \sup_{w \in W(\tau, F)} T(r) \cdot w$ for all such T. Hence $W(\tau, F) \subset W(\rho, F)$ for all $F \in \mathcal{K}$ so (2) and (3) are equivalent. □

Appendix 1E

In Appendix 1E, we relate our results to those of Ahn and Sarver [1].

Lemma (1E.1)**.** *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 1A.2, $\{F^0, F^+\} \in \mathcal{H}_F$. 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 1A.3 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

$$F_\rho = F_\rho^+ = G_\rho^+ = G_\rho$$

□

Proposition (1E.2)**.** *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 1B.6 and the uniqueness properties of the subjective learning representation (see Theorem 1 of DLST). That (2) implies (1) is immediate, so we only need to prove that (1) implies (2).

Assume (1) is true. Since $\succeq_\rho$ is represented by (ν, v), we have $F \sim_\rho G$ implies $F \sim G$. Without loss of generality, we assume both u and v are normalized. First, consider only constant acts and let $\underline{f}$ and $\overline{f}$ be the worst and best acts under v. Now, for any $f \in H_c$, we can find $a \in [0,1]$ such that $\underline{f} a \overline{f} \sim_\rho f$ which implies $\underline{f} a \overline{f} \sim f$. Thus

$$v(f) = v\left(\underline{f} a \overline{f}\right) = 1 - a$$

and

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

for all $f \in H_c$. Thus, $u = \alpha v + \beta$ where $\alpha := u\left(\overline{f}\right) - u\left(\underline{f}\right)$ and $\beta := u\left(\underline{f}\right)$. Since $\underline{f} \cup \overline{f} \sim_\rho \overline{f}$ implies $\underline{f} \cup \overline{f} \sim \overline{f}$, we have $u\left(\overline{f}\right) \geq u\left(\underline{f}\right)$ so $\alpha \geq 0$. If $\alpha = 0$, then $u = \beta$ contradicting the fact that u is non-constant. Thus, $\alpha > 0$.

We can now assume without loss of generality that $\succeq_\rho$ is represented by (ν, u). Now, given any $F \in \mathcal{K}$, we can find $f \in H_c$ such that $F \sim_\rho f$ which implies $F \sim g$. Thus,

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

so $\succeq_\rho$ and $\succeq$ represent the same preference which implies (2). Thus, (1), (2) and (3) are all equivalent. $\square$

2 Random Ambiguity

2.1 Introduction

Ambiguity aversion has been useful in various contexts to explain phenomena that cannot be easily addressed using risk aversion alone. Consider first the case where rare shocks trigger massive uncertainty (i.e. ambiguity) about the environment, resulting in extreme behavior that eludes simple explanations involving only risk. In Caballero and Krishnamurthy [14] for example, during surprise defaults or bank runs, investors are faced with an enlarged set of potential priors, compelling them to scramble toward safer assets in a phenomenon termed "flight to quality". A second case where ambiguity aversion has been informative is in explaining the observed non-participation in financial markets, a behavior that is inconsistent with standard models of optimal portfolio choice. In these models, standard investors remain in the market while investors with high levels of uncertainty aversion choose not to participate (see Easley and O'Hara [25] and Dow and Werlang [24]).

In both cases discussed above, the crucial modelling device driving the narrative is varying ambiguity aversion. In the first case, ambiguity attitudes vary across payoff-relevant states, while in the second, they vary across different individuals in the market. Both examples suggest a probabilistic model of ambiguity aversion where the magnitude of ambiguity aversion depends on the relevant state or individual. In particular, if data on individual or state-specific behavior are scarce or difficult to obtain, then a model of random choice becomes especially useful. Motivated by the above, this chapter characterizes a random utility model where the only source of probabilistic or *random* choice is varying attitudes toward ambiguity aversion. Owing to the central role played by ambiguity aversion, peripheral factors which do not have much to contribute to the story, such as taste or risk preferences, are held constant.

Consider a decision-maker who exhibits traditional ambiguity aversion. One of the most popular and widely applied models of ambiguity is the multiple priors model of Gilboa and Schmeidler [40]. In this model, a decision-maker ranks each Anscombe-Aumann act f according to its maxim expected utility

$$U_K(f) := \min_{p \in K} \; p \cdot (u \circ f)$$

where u is a von Neumann-Morgenstern utility index and K is some non-empty closed convex set of priors on the objective state space.[33] Thus, a decision-maker with maxmin preferences evaluates each act assuming the worst possible prior in the set of priors K. In applications, this is consistent with the widespread use of worst-case scenario analysis by many financial firms.

Within the context of maxmin preferences, a natural and common way of modelling different degrees of uncertainty aversion is by enlarging or shrinking the set of priors K. We consider the simplest parametrization of such priors. For $t \in [0,1]$, let

$$K_t := t\overline{K} + (1-t)\underline{K}$$

be a mixture of two sets of priors $\underline{K} \subset \overline{K}$.[34] Note that higher values of t correspond to larger priors and result in behavior more skewed by ambiguity aversion. Since $\underline{K} \subset K_t$ for all $t \in [0,1]$, $\underline{K}$ can be interpreted as a common set of priors. This agrees with our original focus on modelling varying ambiguity attitudes and not the more general issue of updating beliefs.

Given a finite set of acts F and a specific act $f \in F$, let $\rho_F(f)$ be the probability that f is chosen in F. A *random ambiguity* model specifies that

$$\rho_F(f) = \mu \{ K_t \mid U_{K_t}(f) \geq U_{K_t}(g) \;\; \forall g \in F \}$$

where μ is some distribution on the parametrized priors K_t. In other words, the probability that f is chosen in F is exactly the measure of the set of priors that rank f higher than

[33] We let $u \circ f \in \mathbb{R}^S$ denote the utility vector for the act f.

[34] The Minkowski mixture $t\overline{K} + (1-t)\underline{K}$ is defined as the set of priors of the form $tp + (1-t)q$ for $p \in \overline{K}$ and $q \in \underline{K}$.

everything else in F. Random ambiguity thus belongs to the class of random utility models where choice frequencies are determined by a distribution over utility functions. Here, our random utilities are maxmin expected utilities and we can interpret μ as the distribution of ambiguity aversion. Note that the standard deterministic maxmin model obtains if $\underline{K} = \overline{K}$ is a singleton.

The individual interpretation is that the decision-maker is hit by stochastic shocks that affect her ambiguity aversion. Each $t \in [0, 1]$ thus represents a particular state of the environment. Sometimes the set of subjective priors may be small while other times, Murphy's law seems to prevail. When faced with a set containing only objective risk however, the decision-maker's choice behavior remains deterministic. Thus, random choice arises only as a result of varying attitudes toward ambiguity aversion. Note that the individual in this interpretation is either naively unaware of or powerless against (as in the case of Caballero and Krishnamurthy [14]) these shocks to ambiguity aversion. This approach contrasts with that of Epstein and Kopylov [28], where the decision-maker exhibits a preference for commitment in anticipation of these "cold feet" contingencies.

In the group interpretation, random choice is reflective of heterogeneity in a population. Here, each $t \in [0, 1]$ represents an individual, and μ is interpreted as the distribution of ambiguity aversion in the population. In the case where individual choice behavior is unobservable, characterizing aggregate random choice behavior from a group is useful for identifying μ. There are several papers that highlight the significance of heterogeneity in this context. Easley and O'Hara [25] study the role of regulation in a heterogeneous population with different levels of uncertainty aversion. Bose, Ozdenoren and Pape [13] investigate varying ambiguity attitudes in an auction setting. Epstein and Schneider [29] review various limited participation problems that are addressed using heterogeneous ambiguity attitudes.

The parametrization of the sets of priors above also includes ε-contamination models as a special case (if K_0 is a singleton). These have been widely used in robust statistics and were even suggested by Ellsberg [27] as a simple functional form to address his namesake paradox. This chapter thus provides an axiomatic foundation for ε-contamination models using random choice. Deterministic characterizations include Kopylov [55] and Gajdos et. al. [37]. The former uses observable deferred choice while the latter uses objective sets of priors.

In the benchmark random expected utility model of Gul and Pesendorfer [47], the probability of choosing a lottery p from the set D is given by

$$\rho_D(p) = \mu\{u \in \mathcal{U} \,| u \cdot p \geq u \cdot q \;\forall q \in D\}$$

where μ is some measure on $\mathcal{U}$, the space of all normalized von Neumann-Morgenstern utility indices. Note that in addition to being on the domain of Anscombe-Aumann acts, random ambiguity crucially differs from random expected utility in that the random utilities are non-linear. In general, any random linear utility on a mixture space must satisfy two axioms: linearity and extremeness. The former is the stochastic equivalent of the standard independence axiom and states that choice frequencies remain unchanged when mixed with singletons. The second axiom is unique to random linear utility and asserts that extreme options are chosen with certainty.

That linearity and extremeness are not particularly desirable descriptive properties is supported by the experimental literature. For example, Kahneman and Tversky [51] demonstrate classic violations of linearity in the lottery space while violations of extremeness have been noted very early by Becker, DeGroot and Marschak [7]. Since the maxmin expected utilities in random ambiguity are non-linear, the resulting random choice need not be linear or extreme. In the absence of any ambiguity aversion, a random utility model over Anscombe-Aumann acts would necessarily satisfy extremeness. In random ambiguity, mixtures are chosen with the probability at which ambiguity aversion is powerful enough so that they become more attractive than other more extreme acts. Thus, by introducing uncertainty aversion into a random utility model, we obtain a relaxation of both linearity and extremeness.

2.2 Random Choice Rules

We now describe the main primitive (i.e. choice data) of our model. The setup is identical to that of Chapter 1, but for completeness, we reproduce it here as well. Formally, let S and X be finite sets. We interpret S as an objective state space and X as a set of possible prizes. Let ΔS and ΔX be their respective probability simplexes. We interpret ΔS as the set of beliefs about the state space and ΔX as the set of lotteries over prizes. Following the setup of Anscombe and Aumann [3], an *act* is a mapping $f : S \to \Delta X$ that specifies a payoff in terms of a lottery on X for each realization of $s \in S$. Let H be the set of all acts. A *decision-problem* is a finite non-empty subset of H. Let $\mathcal{D}$ be the set of all decision-problems, which we endow with the Hausdorff metric.[35] The object of our analysis is a *random choice rule (RCR)* that specifies the probabilities that acts are chosen in every decision-problem. For notional convenience, we also let f denote the singleton set $\{f\}$ whenever there is no risk of confusion.

In the classic model of rational choice, if a decision-maker prefers one option over another, then this preference is revealed via her choice of the preferred option. If the two options are indifferent (i.e. they have the same utility), then the model is silent about which option the decision-maker will choose. We introduce an analogous innovation to address indifferences under random choice and random utility. Consider the decision-problem $F = f \cup g$. If the two acts f and g are "indifferent" (i.e. they have the same random utility), then we declare that the random choice rule is unable to specify choice probabilities for each act in the decision-problem. For instance, it could be that the decision-maker ultimately chooses

[35] 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 over g half the time, but ex-ante, it impossible to say what that probability will be. Any probability would be perfectly consistent with the model. Thus, similar to how the classic model is silent about which act the decision-maker will choose, the random choice model is silent about what the individual choice probabilities are. In both cases, we can interpret indifferences as choice behavior that is beyond the scope of the model.[36]

Let $\mathcal{H}$ be some algebra on H. Formally, we model indifference as non-measurability with respect to $\mathcal{H}$. For example, if $\mathcal{H}$ is the Borel algebra, then this corresponds to the benchmark case where every act is measurable. In general though, $\mathcal{H}$ can be coarser than the Borel algebra. Note that given a decision-problem, the decision-problem itself must measurable. This is because we know the decision-maker will choose something in the decision-problem. For $F \in \mathcal{D}$, let $\mathcal{H}_F$ be the algebra generated by $\mathcal{H} \cup \{F\}$.[37] Let Π be the set of all measures on any measurable space of H. We now formally define a random choice rule.

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

We use the notation ρ_F to denote the measure $\rho(F)$. A RCR thus assigns a probability measure on $(H, \mathcal{H}_F)$ for each decision-problem $F \in \mathcal{D}$ such that $\rho_F(F) = 1$. Note that the definition of $\mathcal{H}_F$ ensures that $\rho_F(F)$ is well-defined. We interpret $\rho_F(G)$ as the probability that the decision-maker will choose some act in $G \in \mathcal{H}_F$ given the decision-problem $F \in \mathcal{D}$. For ease of exposition, we denote RCRs by ρ with the implicit understanding that it is associated with some $\mathcal{H}$. We also use the notation $\rho(f, F) := \rho_{F \cup f}(f)$ for any $f \in H$ and $F \in \mathcal{D}$.

If $G \subset F$ is not $\mathcal{H}_F$-measurable, then $\rho_F(G)$ is not well-defined. To address this, let

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

[36] In fact, we could easily distinguish choice data that is beyond the scope of our model by discontinuities in the RCR. See our definition of continuity below.

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

be the measure of the smallest measurable set containing G.[38] Note that ρ_F^* is exactly the outer measure of ρ_F. Both ρ_F and ρ_F^* coincide on $\mathcal{H}_F$-measurable sets. Going forward, we let ρ denote ρ^* without loss of generality.

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

Example 2.1. Let $S = \{s_1, s_2\}$ and $X = \{x, y\}$. Without loss of generality, we can let $f = (a, b) \in [0,1]^2$ denote the act $f \in H$ where

$$f(s_1) = a\delta_x + (1-a)\delta_y$$
$$f(s_2) = b\delta_x + (1-b)\delta_y$$

Let $\mathcal{H}$ be the 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 $\rho_F(f) = 1$ iff $f_1 \geq g_1$ for all $g \in F$. This describes a decision-maker who ranks acts solely based on how likely she will receive prize x if state s_1 realizes. If we let $F = f \cup g$ be such that $f_1 = g_1$, then neither f nor g is $\mathcal{H}_F$-measurable. In other words, the decision-maker is unable to specify individual choice probabilities for f or g. This is because regardless of which act she chooses, the decision-maker will receive x with the same probability in state s_1. The two acts 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 address continuity for RCRs. Given a RCR, let $\mathcal{D}_0 \subset \mathcal{D}$ be the set of decision-problems where every act in the decision-problem is measurable. To be explicit, $F \in \mathcal{D}_0$ iff $f \in \mathcal{H}_F$ for all $f \in F$. Let Π_0 be the set of all Borel measures on H, endowed with the topology of weak convergence. Since all acts in $F \in \mathcal{D}_0$ are $\mathcal{H}_F$-measurable, we can set $\rho_F \in \Pi_0$ for $F \in \mathcal{D}_0$ without loss of generality.[39] We say ρ is *continuous* iff it is continuous on the restricted domain $\mathcal{D}_0$.

[38] Lemma 2A.1 in the Appendix ensures that this is well-defined.
[39] We can easily complete ρ_F so that it is Borel measurable.

Definition. ρ is *continuous* iff $\rho : \mathcal{D}_0 \to \Pi_0$ is continuous

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

2.3 Random Ambiguity Utility

We now describe random ambiguity utility. Let $\mathcal{K}$ be the set of compact, convex and non-empty subsets of ΔS. We interpret $K \in \mathcal{K}$ as a set of priors reflecting uncertainty about the underlying state. Let $\underline{K} \subset \overline{K}$ be two sets of priors, and for $t \in T := [0,1]$, consider the following parametrization of beliefs

$$K_t := t\overline{K} + (1-t)\,\underline{K}$$

Let $\mathcal{K}_T \subset \mathcal{K}$ be the set of all K_t and let μ be a probability measure on $\mathcal{K}_T$. Here, $\underline{K}$ represents a baseline set of priors with uncertainty growing affinely at rate t. We adopt this affine parametrization as it offers a simple one-dimensional model of random uncertainty. It is highly tractable as the entire model is completely specified by two sets of beliefs and a scalar distribution on T. Note that we can also interpret K_t as the set of priors that are "close" to $\underline{K}$ according to some divergence measurement.[40] In this case, t serves as a measurement of the dispersion of beliefs from the set of common priors $\underline{K}$.

[40] In particular, K_t has an affine parametrization if the divergence is linear; that is, if r is equidistant from p and q, then it is also equidistant from $ap+(1-a)\,r$ and $aq+(1-a)\,r$ for $a \in [0,1]$. The ε-Gini-contamination for example satisfies this property (see Grant and Kajii [44]).

Let $u : \Delta X \to \mathbb{R}$ be an affine utility function. For $K_t \in \mathcal{K}_T$, let

$$U_{K_t}(f) := \min_{p \in K_t} p \cdot (u \circ f)$$

We interpret $U_{K_t}(f)$ as the maxmin expected utility of the act $f \in H$ conditional on perceiving uncertainty as reflected by the set of priors K_t. By the additivity of support functions[41], we can rewrite the random utility as

$$U_{K_t} = tU_{\overline{K}} + (1-t)\, U_{\underline{K}}$$

In this interpretation, the decision-maker has two sets of priors in mind but chooses randomly depending on how attractive either premise is. We can also rearrange and interpret t as an affine parametrization of the magnitude of uncertainty aversion. In the special case where $\underline{K}$ is a singleton, the random utility decomposes to expected utility plus a random cost of uncertainty aversion with distribution corresponding to μ.

Given μ, we say it is *regular* iff the utilities of any two acts are either always are never equal.

Definition. μ is *regular* iff $U_{K_t}(f) = U_{K_t}(g)$ with μ-measure zero or one.

Let (μ, u) consist of a regular μ and a non-constant u. We are now ready to introduce the relationship between stochastic sets of priors and the observable random choice rule.

Definition (Random Ambiguity)**.** ρ is *represented by* (μ, u) iff for $f \in F \in \mathcal{D}$

$$\rho_F(f) = \mu \{ K_t \in \mathcal{K}_T \mid U_{K_t}(f) \geq U_{K_t}(g) \;\; \forall g \in F \}$$

Thus, we can view ρ_F as an induced distribution on H by μ. In particular, the probability that an act is chosen is precisely the measure of the set of posteriors that maximize the utility of that act in the decision problem.

In traditional random utility models, indifferences in the random utility must necessarily occur with probability zero. For example, in Gul and Pesendorfer [47], their definition of

[41] For elementary properties of support functions, see Theorem 1.7.5 of Schneider [75]

regularity requires that random utilities are never equal.[42] Here, since not all singletons are necessarily measurable, we can relax this by allowing acts to have the same utility with probability one. In fact, acts that have the same utility μ-a.s. correspond exactly to non-measurable acts. The following example demonstrates.

Example 2.2. Let $S = \{s_1, s_2, s_3\}$, $\underline{K} = r = \left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)$ and $\overline{K} = B_\varepsilon(r) \cap \Delta S$ where $B_\varepsilon(r)$ is the ball with radius $\varepsilon > 0$ centered at r. Let $X = \{x_1, x_2\}$ and $u(a\delta_{x_1} + (1-a)\delta_{x_2}) = a \in [0,1]$. Let $u \circ f_1 = (1,0,0)$ and $u \circ f_2 = (0,1,0)$. If we let ε be small enough, then $U_{\underline{K}}(f_i) = \frac{1}{3}$ and $U_{\overline{K}}(f_i) = \frac{1-\varepsilon\sqrt{6}}{3}$ for $i \in \{1,2\}$. Hence,

$$U_{K_t}(f_1) = \frac{1}{3} - \frac{\varepsilon\sqrt{6}}{3}t = U_{K_t}(f_2)$$

for all $t \in [0,1]$. Thus, f_1 and f_2 are indifferent.

Our definition of regularity enables us to circumvent these issues by allowing for just enough flexibility so that we can model indifferences using non-measurability. Note that the standard subjective expected utility representation obtains as a special case for $\mu\left\{\underline{K} = \overline{K}\right\} = 1$, if we generalize regularity in this fashion. Note that regularity still imposes certain restrictions on μ. For example, multiple mass points are not allowed if μ is regular.

Theorem 2.1 highlights the uniqueness properties of random ambiguity. The main highlight is that studying binary choices is enough to completely identify the distribution of ambiguity attitudes.

Theorem 2.1. *Let ρ and τ be represented by (μ, u) and (ν, v) respectively. Then the following are equivalent*

(1) $\rho(f, g) = \tau(f, g)$ *for all* f *and* g

(2) $\rho = \tau$

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

Proof. See Appendix. □

[42] Also see Block and Marschalk [12] for the case of finite alternatives.

We end this section with a simple example of random ambiguity.

Example 2.3. Let $S = \{s_1, s_2\}$, $X = \{x_1, x_2\}$ and $u(a\delta_{x_1} + (1-a)\delta_{x_2}) = a \in [0,1]$. Let $\underline{K} = \left(\frac{1}{3}, \frac{2}{3}\right)$, $\overline{K} = \left\{\left(\frac{1}{3}+k, \frac{2}{3}-k\right) \middle| k \in \left[0, \frac{1}{2}\right]\right\}$ and μ be uniform. Note that each act corresponds to a specific point in the unit square and for $u \circ f = (a,b) \in [0,1]^2$,

$$U_{K_t}(f) = \min_{p \in \left[\frac{1}{3}, \frac{1}{3}+t\frac{1}{2}\right]} (ap + (1-p)b) = \begin{cases} \frac{a+2b}{3} & \text{if } a \geq b \\ \frac{a+2b}{3} + \frac{a-b}{3}t & \text{if } a < b \end{cases}$$

Let ρ be represented by (μ, u). In the region $a \geq b$, ρ is deterministic with indifference lines $b = \kappa - \frac{1}{2}a$. For $a < b$, the indifference lines are $b = \kappa - \frac{1+t}{2-t}a$ with t distributed uniformly on T. This RCR describes a decision-maker whose set of priors on s_1 range uniformly from $\left\{\frac{1}{3}\right\}$ to $\left[\frac{1}{3}, \frac{5}{6}\right]$. In this case, since the s_1 prior is bounded below by $\frac{1}{3}$, the only factor driving random choice is the stochastic upper bound on the prior.

2.4 Random Non-Linearity

Before we proceed to the characterization of random ambiguity, we first explore some of the non-linear properties of random ambiguity in its stochastic context. In random expected utility [47], two necessary conditions are linearity and extremeness. We present these two conditions in our setting. First, given two decision-problems F and G, let $aF + (1-a)G$ denote the Minkowski mixture of the two sets for $a \in [0,1]$.[43]

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

An act $f \in F$ is *extreme* in F iff it is not in the interior of the convex hull of F. Let $\text{ext}F \subset F$ denote the extreme points of the decision problem F.

Definition. ρ is *extreme* iff $\rho_F(\text{ext}F) = 1$.

[43] The Minkowski mixture for $\{F, G\} \subset \mathcal{D}$ and $a \in [0,1]$ is defined as $aF + (1-a)G := \{af + (1-a)g | (f,g) \in F \times G\}$.

Linearity is the stochastic equivalent of the standard independence axiom. In particular, it is the version of the standard independence axiom that is actually tested in many experimental settings (such as in Kahneman and Tversky [51]). If we impose linearity on the random utility, then linearity of the induced RCR follows as a natural consequence. On the other hand, extremeness is a condition that is unique to random choice. It requires that extreme acts of a decision problem are always chosen, or vice-versa, interior acts are never chosen. Since random linear utilities admit indifference sets that are hyperplanes, the resulting RCR must be extreme (ignoring indifferences).

Given that random ambiguity utility is non-linear, one would expect the induced RCR to permit violations of both linearity and extremeness. The following example provides confirmation.

Example 2.4. We return to the setup in Example 2.3 above. Let $u \circ f = (0, 1)$, $u \circ g = (1, 0)$ and $h = \frac{1}{2} f + \frac{1}{2} g$. Letting λ be the Lebesgue measure on T, we have

$$\rho(f, g) = \lambda \{t \in T \mid 2 - t \geq 1\} = 1$$
$$\rho(f, h) = \lambda \left\{ t \in T \mid 2 - t \geq \frac{3}{2} \right\} = \frac{1}{2}$$

Since $\frac{1}{2}\{f, g\} + \frac{1}{2} f = \{f, h\}$, we have $\rho(f, g) > \rho(f, h)$ violating linearity. If we let $F := \{f, g, h\}$, then

$$\rho_F(h) = \lambda \left\{ t \in T \mid \frac{3}{2} \geq \max(1, 2 - t) \right\} = \frac{1}{2} > 0$$

violating extremeness. Note that since g exhibits less uncertainty than either f or h, ambiguity aversion suggests that g should be chosen with some probability.

As the example above illustrates, violations of extremeness is precisely the behavior that characterizes uncertainty aversion. We now introduce a property that exactly captures this fact. First note that although maxmin expected utility is non-linear, it does satisfy quasiconcavity. Under static choice, a utility is quasiconcave iff the preference relation it represents is convex.[44] In the realm of random choice, quasiconcavity of the random utility

[44] A preference relation $\succeq$ is convex iff $f \succeq h$ implies $af + (1 - a) h \succeq h$ for $a \in (0, 1)$.

is directly related to a property of its induced RCR which we will aptly call convexity.

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

Let f and g be two acts and $h = af + (1-a)g$ be any mixture of the two for some $a \in (0,1)$. Convexity requires that for any decision problem that already includes h, adding g will not affect the probability that f is chosen. In other words, acts are immune to the addition of new acts provided that some mixture act already exists in the original decision problem. The following example illustrates.

Example 2.5. We again return to the setup of Example 2.3. Let $u \circ f = (0,1)$, $u \circ g = \left(1, \frac{1}{4}\right)$ and $h = \frac{1}{2}f + \frac{1}{2}g$. Now,

$$\rho(f,h) = \lambda\left\{t \in T \mid 2 - t \geq \frac{14-t}{8}\right\} = \frac{2}{7}$$

If we let $F = \{f, g, h\}$, then

$$\begin{aligned}\rho_F(f) &= \lambda\left\{t \in T \mid 2 - t \geq \max\left(\frac{14-t}{8}, \frac{3}{2}\right)\right\} \\ &= \frac{2}{7} = \rho(f,h)\end{aligned}$$

so convexity is satisfied in this case.

Finally, we conclude this section by providing some justification for our choice of convexity. Let $\mathcal{V}$ be the set of all measurable $v : H \to \mathbb{R}$ and let μ be a probability measure on $\mathcal{V}$. For ease of exposition, we assume that μ has no indifferences. Thus, for all $\{f,g\} \subset H$, $v(f) = v(g)$ with μ-measure zero. We now define *quasiconcavity* and *quasiconvexity* for random utilities.

Definition. μ is *quasiconcave* iff for all $\{f,h\} \subset H$ and $a \in (0,1)$,

$$v(fah) \geq \min(v(f), v(h))$$

μ-a.s.. It is *quasiconvex* iff

$$v(fah) \leq \max(v(f), v(h))$$

μ-a.s..

Definition. ρ is represented by μ iff for all $f \in F \in \mathcal{D}$,

$$\rho_F(f) = \mu\{v \in \mathcal{V} | v(f) \geq v(g) \ \ \forall g \in F\}$$

Proposition 2.1. *Let ρ be represented by μ. Then μ is quasiconcave (quasiconvex) iff ρ is convex (extreme).*

Proof. See Appendix. □

Thus, convexity and extremeness are the relevant stochastic properties of random choice for characterizing quasiconcavity and quasiconvexity respectively. Note that since random linear utility is both quasiconcave and quasiconvex, it follows that its induced RCR must be both convex and extreme.

2.5 Characterization

We now provide a complete characterization of random ambiguity. We say $f \in H$ is *constant* iff $f(s)$ is the same for all $s \in S$. A decision-problem is *constant* iff it contains constant acts only. Given an act f and a state $s \in S$, let $f_s \in H$ denote the constant act that yields the lottery $f(s) \in \Delta X$ in every state. For $F \in \mathcal{D}$, let $F_s := \bigcup_{f \in F} f_s$ be the constant decision problem consisting of f_s for all $f \in F$.

The first five conditions below are the stochastic counterparts to the original axioms of maxmin expected utility under static choice. In what follows, assume $f \in F \in \mathcal{D}$.

Axiom 2.1. (C-linearity) $\rho_F(f) = \rho_{aF+(1-a)ag}(af + (1-a)g)$ *for constant g and $a \in (0,1)$.*

Axiom 2.2. (Continuity) *ρ is continuous.*

Axiom 2.3. (S-monotonicity) $\rho_{F_s}(f_s) = 1$ *for all $s \in S$ implies $\rho_F(f) = 1$.*

Axiom 2.4. (Convexity) *ρ is convex.*

Axiom 2.5. (Non-degeneracy) *$\rho(f, g) = 0$ for some constant f and g.*

Certainty-linearity (or C-linearity) imposes linearity only for mixtures with constant acts. It is a weakening of linearity that directly corresponds to C-independence, the weakening of the standard independence axiom under deterministic maxmin. The continuity condition is standard albeit adjusted for the measurability issues discussed above. State-monotonicity (or S-monotonicity) is the stochastic analog of the monotonicity axiom in maxmin preferences. It is both necessary and sufficient for any random utility that satisfies state-wise monotonicity. Convexity serves the same purpose as the uncertainty aversion condition in static maxmin by enforcing quasiconcavity. Thus, in the context of random ambiguity, convexity can be interpreted as the stochastic version of uncertainty aversion. Finally, non-degeneracy ensures that the RCR is non-trivial and plays the same role as its counterpart in the deterministic model.

The last two conditions are particular to random choice. First, note that in any random utility model, the probability of an act being chosen decreases as the decision problem is enlarged. This property called monotonicity is necessary (but not sufficient) for any RCR induced by a random utility. Since random ambiguity is a random utility model, we naturally include monotonicity as a condition.

Axiom 2.6. (Monotonicity) *$F \subset G$ implies $\rho_F(f) \geq \rho_G(f)$.*

One interesting implication of monotonicity in collaboration with convexity is that an act is chosen less frequently as other acts converge linearly toward it. To illustrate this, let $f \in F$ and $h = af + (1 - a) g$ for some $a \in (0, 1)$. Thus, h is closer to f than g is to f. Now, by monotonicity and convexity,

$$\rho_{F \cup g}(f) \geq \rho_{F \cup g \cup h}(f) = \rho_{F \cup h}(f)$$

Thus, f become less prominent in choice as other acts move toward it. This is an observable characteristic of random quasiconcave utility that is completely unique to random choice.

Before we present the final condition, we first consider a special class of random utility models known as Luce rules. One feature of Luce rules is that they satisfy an appealing form of stochastic independence which we present here in our context.

Definition. ρ is *independent* iff $\rho(f,G) \geq \rho(g,G)$ implies $\rho(f,F) \geq \rho(g,F)$

Gul, Natenzon and Pesendorfer [46] show that under a richness condition, independence completely characterizes the Luce rule. Random ambiguity however, is in general not a Luce rule as the following example illustrates.

Example 2.6. Return to the setup of Example 2.3 above. Let $u \circ f = (0,1)$, $u \circ h = \left(\frac{1}{2}, \frac{1}{2}\right)$, $g = \frac{1}{2}f + \frac{1}{2}h$ and $u \circ h' = \left(\frac{2}{5}, \frac{2}{5}\right)$. Now,

$$\rho(f,h) = \lambda \left\{ t \in T \,\middle|\, 2 - t \geq \frac{3}{2} \right\} = \frac{1}{2} = \lambda \left\{ t \in T \,\middle|\, \frac{7}{4} - \frac{1}{2}t \geq \frac{3}{2} \right\} = \rho(g,h)$$

However,

$$\rho(f,h') = \lambda \left\{ t \in T \,\middle|\, 2 - t \geq \frac{6}{5} \right\} = \frac{4}{5} < 1 = \lambda \left\{ t \in T \,\middle|\, \frac{7}{4} - \frac{1}{2}t \geq \frac{6}{5} \right\} = \rho(g,h')$$

violating independence.

Nevertheless, random ambiguity does satisfy a weakened form of independence called certainty-dominance (or C-dominance) which we present as our last condition.

Axiom 2.7. (C-dominance) *If* $\rho(f,h) > \rho(g,h) = 0$ *and* $1 = \rho(f,h') > \rho(g,h')$ *for constants* h *and* h'*, then* $\rho(f,F) \geq \rho(g,F)$ *for all* $F \in \mathcal{D}$*.*

C-dominance specifies sufficient conditions for when an act f *stochastically dominates* another act g i.e. for any set F, f is chosen more often in $F \cup f$ than g is in $F \cup g$. Let h and h' be two constant acts. Suppose f is chosen sometimes over h while g is never chosen, while f is always chosen over h' while g is not always chosen. Then it must be that f stochastically dominates` g. The restrictions in the premise that g is never chosen over h and that f is always chosen in over h' are important. They ensure that the comparisons between f and g are stark enough for f to stochastically dominate g. Put in another way, there are constant acts h and h' such that $\rho(f,h) > \rho(g,h)$ and $\rho(f,h') > \rho(g,h')$ but $\rho(f,F) < \rho(g,F)$ for some $F \in \mathcal{K}$. The following illustrates an example of C-dominance at work.

Example 2.7. Returning once again to the setup of Example 2.3, let $u \circ f = \left(\frac{1}{2}, 1\right)$, $u \circ g = \left(\frac{1}{4}, \frac{3}{4}\right)$, $u \circ h = \left(\frac{3}{4}, \frac{3}{4}\right)$, and $u \circ h' = \left(\frac{1}{2}, \frac{1}{2}\right)$. Now,

$$\rho(f,h) = \lambda\left\{t \in T \,\middle|\, \frac{5}{2} - \frac{1}{2}t \geq \frac{9}{4}\right\} = \frac{1}{2} > 0 = \lambda\left\{t \in T \,\middle|\, \frac{7}{4} - \frac{1}{2}t \geq \frac{9}{4}\right\} = \rho(g,h)$$

and

$$\rho(f,h') = \lambda\left\{t \in T \,\middle|\, \frac{5}{2} - \frac{1}{2}t \geq \frac{3}{2}\right\} = 1 > \frac{1}{2} = \lambda\left\{t \in T \,\middle|\, \frac{7}{4} - \frac{1}{2}t \geq \frac{3}{2}\right\} = \rho(g,h')$$

so by C-dominance, $\rho(f,F) \geq \rho(g,F)$ for any $F \in \mathcal{D}$. Note that this must hold since $\forall t \in T$,

$$U_{K_t}(f) = \frac{10-2t}{4} > \frac{7-2t}{4} = U_{K_t}(g)$$

We now present the representation result. Note that this representation is unique as per Theorem 2.1 above.

Theorem 2.2. *ρ satisfies Axioms 2.1-2.7 iff it is has a random ambiguity representation.*

Proof. See Appendix. □

Necessity of Axioms 2.1-2.7 follow easily from the representation. To understand the necessity of C-dominance, note that the affine parametrization of K_t restricts the random utilities to be affine functions on T. The premise of C-dominance then implies that the utility of f must be greater than that of g so the desired implication follows.

We now offer a brief outline of the sufficiency argument. First, note that the lower contour sets in maxmin expected utility are translated convex cones. If we consider the affine subspace containing some act f and two distinct constant acts, then the random utilities are linear for all decision problems in this subspace. The crux of the argument rests on showing that convexity and C-linearity in conjunction with the other conditions are sufficient for establishing a random expected utility representation in this subspace. C-dominance then allows us to map all decision problems into this particular subspace and thus admit a random utility representation. Finally, C-dominance also restricts the class of random maxmin expected utilities to precisely those with a regular μ as desired in random ambiguity.

In this final section, we introduce a notion of comparative uncertainty aversion in the context of random choice.

Definition. ρ is *more uncertainty averse* than τ iff $\rho_F(f) \geq \tau_F(f)$ for all constant f

Thus, if ρ exhibits more uncertainty aversion than τ, then constant acts (which are free of uncertainty) are chosen more often under ρ than τ. The following definition extends first-order stochastic dominance to distributions over sets.

Definition. $\nu \gg \mu$ iff $\nu\{K \subset L\} \geq \mu\{K \subset L\}$ for all $L \in \mathcal{K}$.

Hence, if $\nu \gg \mu$, then μ puts less weight on smaller sets of priors than ν. In other words, μ is more uncertain than ν. The following proposition characterizes greater uncertainty under random ambiguity.

Proposition 2.2. *Let ρ and τ be represented by (μ, u) and (ν, v) respectively. Then ρ is more uncertainty averse than τ iff $\nu \gg \mu$ and $u = \alpha + \beta v$ for $\beta > 0$.*

Proof. See Appendix. □

This result thus allows for comparisons between distributions of ambiguity aversion solely based on choice frequencies. It is the random equivalent of corresponding results under deterministic choice.[45]

2.6 Summary

We introduce a model of random ambiguity where stochastic levels of uncertainty aversion drive probabilistic choice. We provide a full characterization of the representation and show

[45] See Theorem 17(ii) of Ghirardato and Marinacci [39].

that the distribution of the random set of priors is uniquely determined by the observable choice probabilities. This scalar parametrization of uncertainty aversion is tractable and generalizes models which employ idiosyncratic shocks to uncertainty aversion or exhibit heterogeneous ambiguity aversion. Finally, by relaxing linearity and extremeness, we have extended non-expected utility to the realm of random choice.

Appendix 2A

The first part of this appendix follows closely to that of Appendix 1A. We include it for completion. Given a collection of sets $\mathcal{G}$ and $F \in \mathcal{D}$, let

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

Note that if $\mathcal{G}$ is an algebra, then $\mathcal{G} \cap F$ is the trace algebra of $\mathcal{G}$ on $F \in \mathcal{D}$. For $G \subset F \in \mathcal{D}$, let

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

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

Lemma (2A.1)**.** *Let $G \subset F \in \mathcal{D}$.*

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

Proof. See proof of Lemma 1A.1. □

Let ρ be a RCR. By Lemma 2A.1, we can now define

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

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

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

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

Lemma (2A.2)**.** *For $\{f, g\} \subset F \in \mathcal{D}$, the following are equivalent.*

(1) f *and* g *are tied*

(2) $g \in f_F$

(3) $f_F = g_F$

Proof. See proof of Lemma 1A.2. □

Lemma (2A.3)**.** *Let ρ be monotonic.*

(1) *For $f \in F \in \mathcal{D}$, $\rho_F(f) = \rho_{F \cup g}(f)$ 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 $i \in \{1, \ldots, n\}$. Then $\rho_F(f_i) = \rho_G(g_i)$ for all $i \in \{1, \ldots, n\}$.*

Proof. See proof of Lemma 1A.3. □

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

Lemma (2A.4)**.** *Let ρ be monotonic and linear. For $f \in F \in \mathcal{D}$, 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. See proof of Lemma 1A.4. □

Proposition (2A.5)**.** *Let ρ be represented by μ. Then μ is quasiconcave (quasiconvex) iff ρ is convex (extreme).*

Proof. First assume ρ is convex. Thus, for $a \in (0,1)$

$$
\begin{aligned}
0 &= \rho(f, fah) - \rho(f, fah \cup h) \\
&= \mu\{v \in \mathcal{V} | \, v(f) \geq v(fah)\} - \mu\{v \in \mathcal{V} | \, v(f) \geq \max(v(fah), v(h))\} \\
&= \mu\{v \in \mathcal{V} | \, v(h) > v(f) \geq v(fah)\}
\end{aligned}
$$

and by symmetric argument, $\mu\{v \in \mathcal{V} | \, v(h) > v(f) \geq v(fah)\} = 0$. Since μ has no ties, it must be quasiconcave. For the converse, suppose μ is quasiconcave and let $\{f, fah\} \subset F$ so

$$
\begin{aligned}
\rho_F(f) - \rho_{F \cup h}(f) =& \mu\{v \in \mathcal{V} | \, v(f) \geq v(g) \;\; \forall g \in F\} \\
& - \mu\{v \in \mathcal{V} | \, v(f) \geq \max\{v(g), v(h)\} \;\; \forall g \in F\} \\
=& \mu\{v \in \mathcal{V} | \, v(h) > v(f) \geq v(g) \;\; \forall g \in F\}
\end{aligned}
$$

Since $fah \in F$ and there are no ties, $\rho_F(f) = \rho_{F\cup h}(f)$.

Now, assume ρ is extreme. Note that if μ is not quasiconvex, then there are $\{f, h\} \subset H$ such that

$$0 < \mu\{v \in \mathcal{V} | \, v(fah) \geq \max(v(f), v(h))\} = \rho(fah, f \cup h)$$

contradicting extremeness. Finally, assume μ is quasiconvex. Let $f \in F \in \mathcal{D}$ where $f \notin E := \text{ext}F$. Thus, by Minkowski's theorem (Corollary 1.4.5 of Schneider [75]), we can find $a_i \in (0,1)$ and $f_i \in E$ such that $h_0 = f_0$, $h_i = h_{i-1}a_i f_i$ and $h_k = f$. By quasiconvexity,

$$v(h_i) \leq \max(v(h_{i-1}), v(f_i))$$

μ-a.s. so by iteration, $v(f) \leq \max_i v(f_i)$ μ-a.s.. Thus,

$$\begin{aligned} \rho_F(f) &= \mu\{v \in \mathcal{V} | \, v(f) \geq v(g) \;\; \forall g \in F\} \\ &\leq \mu\{v \in \mathcal{V} | \, v(f) \geq v(f_i) \;\; \forall i\} = 0 \end{aligned}$$

as μ has no ties. As a result, ρ is extreme. □

Appendix 2B

In what follows, assume ρ satisfies Axioms 2.1-2.7. Let $H_c \subset H$ denote the set of all constant acts

Lemma (2B.1)**.** $\rho_F(f) \in \{0,1\}$ *for constant* F

Proof. Suppose $F \in \mathcal{D}$ is constant but $\rho_F(f) \in (0,1)$. Thus, we can find a $g \notin f_F$ such that $\rho_F(g) \in (0,1)$. By Lemma 2A.2, f and g are not tied, so by monotonicity, $\rho(f,g) \in (0,1)$. By non-degeneracy and C-linearity, let $\{f', g'\} \subset H_c$ be s.t. $\rho(f', f) = \rho(g, g') = 0$ without loss of generality. Note that since as $\rho(f,g) \in (0,1)$, by Lemma 2A.3, f cannot by tied with g' and g cannot be tied with f'. Thus, by continuity, we can find f' and g' close enough to f and g respectively such that $\rho(f, g') > 0 = \rho(g, g')$ and $\rho(f, f') = 1 > \rho(g, f')$. By

C-dominance, $\rho_F(f) \geq \rho_F(g)$ for any $F \in \mathcal{D}$. However, we can find h close to f such that by continuity, $\rho(f,h) = 0 < \rho(g,h)$ a contradiction. Hence $\rho_F(f) \in \{0,1\}$ for constant $F \in \mathcal{D}$. $\square$

Lemma (2B.2)**.** *There exists a non-constant and affine $u : \Delta X \to \mathbb{R}$ such that for any constant $f \cup g$, $\rho(f,g) = 1$ iff $u(f) \geq u(g)$.*

Proof. Define the binary relation $\succeq$ on H_c such that $f \succeq g$ iff $\rho(f,g) = 1$. Note that if $f \not\succeq g$, then $\rho(f,g) = 0$ so $\rho(g,f) = 1$ by Lemma 2B.1. Thus, $\succeq$ is complete. Note that $f \sim g$ iff f and g are tied. Now, assume $f \succeq g \succeq h$ but $h \succ f$ so $\rho(f,h) = 0$. If $f \sim g$, then by Lemma 2A.3, $0 = \rho(f,h) = \rho(g,h)$ a contradiction. If $g \sim h$, then $0 = \rho(f,h) = \rho(f,g)$ again a contradiction. Thus, $f \succ g \succ h \succ f$. If we let $F := \{f,g,h\}$, then by monotonicity,

$$\rho_F(g) = \rho_F(f) = \rho_F(h) = 0$$

Thus, $\rho_F(F) = 0$ a contradiction. Therefore, $\succeq$ is both complete and transitive.

We now show that $\succeq$ has an expected utility representation. Suppose $f \succ g$ so $\rho(f,g) = 1$ and $\rho(g,f) = 0$. Linearity implies $\rho(fah, gah) = 1$ and $\rho(gah, fah) = 0$ so $fah \succ gah$ for $a \in (0,1)$. Hence $\succeq$ satisfies the standard independence axiom. If $f \succ g \succ h$, then $\rho(f,g) = \rho(g,h) = 1$ and $\rho(g,f) = \rho(h,g) = 0$. Since $\succeq$ is transitive, $f \succ h$ so $\rho(f,h) = 1$ and $\rho(h,f) = 0$. Now, suppose $\exists a \in (0,1)$ such that fah and g are tied. If $\exists b \in (0,1) \setminus a$ such that fbh and g are tied, then by Lemma 2A.2, fah and fbh are tied, contradicting linearity. Thus, we can apply continuity and find $\{a,b\} \subset (0,1)$ such that $\rho(fah, g) = \rho(g, fbh) = 1$ and $\rho(g, fah) = \rho(fbh, g) = 0$. By the Mixture Space theorem (see Theorem 8.4 of Fishburn [35]), there is an affine $u : \Delta X \to \mathbb{R}$ that represents $\succeq$. Finally, by non-degeneracy, $\rho(f,g) = 0$ for some $\{f,g\} \subset H_c$. Thus, $u(f) < u(g)$ so u is non-constant. $\square$

Since u is non-constant, we can choose $\{\underline{f}, \overline{f}\} \subset H_c$ such that $u(\overline{f}) \geq u(f) \geq u(\underline{f})$ for all $f \in H_c$ with at least one inequality strict. Moreover, without loss of generality, let

$u\left(\overline{f}\right)=1$ and $u\left(\underline{f}\right)=0$. Let $\left[\underline{f},\overline{f}\right]:=\left\{\underline{f}a\overline{f}\middle|\, a\in[0,1]\right\}$ and

$$H_0:=\left\{h\in H\middle|\, h(s)\in\left[\underline{f},\overline{f}\right]\ \ \forall s\in S\right\}$$

Note that there is an isomorphism between H_0 and $u\circ H_0=[0,1]^S$ so we can associate each $f\in H_0$ with its utility vector $u\circ f\in[0,1]^S$ without loss of generality.

Suppose there exists some $h^*\in H_0\backslash\left[\underline{f},\overline{f}\right]$ such that h^* is not tied with any act in $\left[\underline{f},\overline{f}\right]$. Let $W^*:=\operatorname{lin}\{u\circ h^*,\ 0,\mathbf{1}\}$ and

$$W_+^*:=\{w\in W^*|\, w^*\cdot w\geq 0\}$$

for some $w^*\in W^*$ such that $w^*\cdot\mathbf{1}=0$ and $w^*\cdot(u\circ h^*)\geq 0$. Thus, W_+^* is the halfspace of W^* containing $u\circ h^*$. If no such h^* exists, then let $W_+^*:=W^*:=\operatorname{lin}\{0,\mathbf{1}\}$. Let $H^*\subset H_0$ be such that

$$u\circ H^*=W_+^*\cap[0,1]^S$$

Note that $\dim(u\circ H^*)\in\{1,2\}$. We say H^* is *degenerate* iff $\dim(u\circ H^*)=1$. Let $\mathcal{D}^*\subset\mathcal{D}$ be the set of decision problems consisting only of acts in H^*.

Lemma (2B.3)**.** *The following holds*

(1) $\mathcal{D}^*\subset\mathcal{D}_0$

(2) ρ *is linear on* $\mathcal{D}^*$

(3) ρ *is extreme on* $\mathcal{D}^*$

Proof. Note that all results follow trivially if H^* is degenerate, so assume $\dim(u\circ H^*)=2$. We prove the lemma in order.

(1) Let $f\in F\in\mathcal{D}^*$ and suppose $g\in f_F$ for some $g\neq f$. Thus, by Lemma 2A.2, f and g are tied. By C-linearity, we can assume $f=h^*$ and $g\in\left[\underline{f},\overline{f}\right]$ without loss of generality, contradicting the definition of h^*. Thus, $\mathcal{D}^*\subset\mathcal{D}_0$

(2) Let $f\in F\in\mathcal{D}^*$, $h\in H^*$ and $a\in(0,1)$. Let $f':=fah$ and $F':=Fah$. Let $\{f_1,f_2\}\subset\left[\underline{f},\overline{f}\right]$ and $b\in(0,1)$ such that $faf_1=f'bf_2$ without loss of generality by

C-linearity. Thus,

$$
\begin{aligned}
faf_1 &= b\left(af + (1-a)h\right) + (1-b)f_2 \\
&= abf + (1-ab)\left(\frac{b-ab}{1-ab}h + \frac{1-b}{1-ab}f_2\right) \\
&= f\,(ab)\left(h\left(\frac{b-ab}{1-ab}\right)f_2\right) = f\,(ab)\,h'
\end{aligned}
$$

for $h' := h\left(\frac{b-ab}{1-ab}\right)f_2 \in H^*$. Now,

$$
\begin{aligned}
f\,(ab)\,h' = af + (1-a)f_1 &= abf + (1-ab)\left(\frac{a-ab}{1-ab}f + \left(1-\frac{a-ab}{1-ab}\right)f_1\right) \\
&= f\,(ab)\left(f\left(\frac{a-ab}{1-ab}\right)f_1\right)
\end{aligned}
$$

so $h' = f\left(\frac{a-ab}{1-ab}\right)f_1$. Let $G := Faf_1$ and $G' := F'bf_2$. Now, for $g \in F\backslash f$

$$
\begin{aligned}
(gah)\,bf_2 &= b\left(ag + (1-a)h\right) + (1-b)f_2 \\
&= abg + (1-ab)\,h' = abg + (1-ab)\left(\frac{a-ab}{1-ab}f + \left(1-\frac{a-ab}{1-ab}\right)f_1\right) \\
&= abg + (1-b)\,af + (1-a)f_1 = (gaf_1)\,b\,(faf_1)
\end{aligned}
$$

Thus, by convexity,

$$
\rho_{G'}\left(f'bf_2\right) = \rho_{G'}\left(faf_1\right) = \rho_{G'\cup(gaf_1)}\left(faf_1\right)
$$

Since this is true for all $g \in F\backslash f$, by C-linearity and monotonicity, we have

$$
\rho_{F'}\left(f'\right) = \rho_{G'}\left(f'bf_2\right) = \rho_{G'\cup G}\left(faf_1\right) \leq \rho_G\left(faf_1\right) = \rho_F\left(f\right)
$$

for all $f \in F$. Thus, by (1), $\rho_{F'}(f') = \rho_F(f)$ so (2) is true.

(3) Let $F \in \mathcal{D}^*$, $f \in F\backslash \text{ext}F$ and suppose $\rho_F(f) > 0$. Now, by Minkowski's Theorem, we can find $a_i \in (0,1)$ and $f_i \in \text{ext}F$ such that $h_0 = f_0$, $h_i = h_{i-1}a_i f_i$ and $h_k = f$. If we let $F' := F \cup \left(\bigcup_i h_i\right)$, then by monotonicity and (1),

$$
1 > \rho_F\left(F\backslash f\right) \geq \rho_{F'}\left(F\backslash f\right)
$$

Thus $\rho_{F'}\left(\bigcup_{i\leq k} h_i\right) > 0$ so there is some $h_i = h_{i-1}a_i f_i$ such that $\rho_{F'}(h_i) > 0$. By monotonicity, $\rho(h_i, f_i \cup h_{i-1}) > 0$. By (2) and convexity, $\rho(h_{i-1}, f_i) = \rho(h_{i-1}, h_i) = \rho(h_{i-1}, f_i \cup h_i)$ and $\rho(f_i, h_{i-1}) = \rho(f_i, h_i) = \rho_G(f_i, h_i \cup h_{i-1})$ so

$$\rho(h_{i-1}, f_i \cup h_i) + \rho(f_i, h_i \cup h_{i-1}) = \rho(h_{i-1}, f_i) + \rho(f_i, h_{i-1}) = 1$$

However, this implies $\rho(h_i, f_i \cup h_{i-1}) = 0$ a contradiction. Thus, ρ is extreme on $\mathcal{D}^*$.

□

Define $Z^* := \{w \in W^* | \, w \cdot \mathbf{1} = 1\}$. Note that $\dim(Z^*) \in \{0,1\}$, and $\dim(Z^*) = 0$ iff H^* is degenerate.

Proposition (2B.4)**.** *There exists a measure μ^* on Z^* such for any $F \in \mathcal{D}^*$,*

$$\rho_F(f) = \mu^* \{z \in Z^* | \, z \cdot (u \circ f) \geq z \cdot (u \circ g) \;\; \forall g \in F\}$$

Proof. Note that if H^* is degenerate, then the result follows trivially so assume otherwise. Let Δ be the two-dimensional probability simplex. Now, there exists an affine transformation $L = \lambda A$ where $\lambda > 0$, A is an orthogonal matrix and $L(u \circ H^*) \subset \Delta$. Note that $L(W^*) = V := \text{lin}(\Delta)$. For each finite set $D \subset \Delta$, we can find a $r \in \Delta$ and $a \in (0,1)$ such that $Dar \subset L(u \circ H^*)$. Thus, we can define a RCR τ on Δ such that

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

where $L(u \circ F) = Dar$ and $L(u \circ f) = par$. Linearity ensures this is well-defined. By Lemma 2B.3, $\mathcal{D}^* \subset \mathcal{D}_0$ and τ is monotone, linear, extreme and continuous. Hence, by Theorem 3 of Gul and Pesendorfer [47], there exists a measure μ_0 on V such that

$$\begin{aligned}\rho_F(f) &= \tau_{L(u\circ F)}(L(u \circ f)) \\ &= \mu_0 \{v \in V | \, v \cdot (L(u \circ f)) \geq v \cdot (L(u \circ g)) \;\; \forall g \in F\}\end{aligned}$$

for all $F \in \mathcal{D}^*$. Since $A^{-1} = A'$,

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

so

$$\begin{aligned}\rho_F(f) &= \mu_0 \left\{ v \in V \middle| L^{-1}(v) \cdot (u \circ f) \geq L^{-1}(v) \cdot (u \circ g) \ \forall g \in F \right\} \\ &= \mu_1 \left\{ w \in W^* \middle| w \cdot (u \circ f) \geq w \cdot (u \circ g) \ \forall g \in F \right\}\end{aligned}$$

where $\mu_1 := \mu_0 \circ L$ is the measure on W^* induced by L. Finally, note that

$$0 = \rho\left(\underline{f}.\overline{f}\right) = \nu_1 \left\{ w \in W^* \middle| 0 \geq w \cdot \mathbf{1} \right\}$$

so $w \cdot \mathbf{1} > 0$ μ_1-a.s.. Since $w \in W^*$ implies $\frac{w}{w \cdot \mathbf{1}} \in Z^*$, we have

$$\begin{aligned}\rho_F(f) &= \mu_1 \left\{ w \in W^* \middle| \frac{w}{w \cdot \mathbf{1}} \cdot (u \circ f) \geq \frac{w}{w \cdot \mathbf{1}} \cdot (u \circ g) \ \forall g \in F \right\} \\ &= \mu^* \left\{ z \in Z^* \middle| z \cdot (u \circ f) \geq z \cdot (u \circ g) \ \forall g \in F \right\}\end{aligned}$$

where μ^* is the induced measure on Z^*. □

For each $f \in H$, define

$$\begin{aligned}f^- &:= \arg\max \left\{ u(h) \middle| h \in \left[\underline{f}, \overline{f}\right], \rho(f, h) = 1 \right\} \\ f^+ &:= \arg\min \left\{ u(h) \middle| h \in \left[\underline{f}, \overline{f}\right], \rho(h, f) = 1 \right\}\end{aligned}$$

Lemma (2B.5)**.** f^- and f^+ are well-defined for all $f \in H$

Proof. Let $f \in H$. First note that if f is tied with any $g \in \left[\underline{f}, \overline{f}\right]$, then by Lemma 2A.3, $f^- = f^+ = g$. Hence, assume f is not tied with any act in $\left[\underline{f}, \overline{f}\right]$. We first show that f^- is well-defined. By S-monotonicity, if $\rho\left(f, \underline{f}\right) < 1$, then $\rho\left(f_s, \underline{f}\right) < 1$ for some $s \in S$ contradicting the definition of $\underline{f}$. Thus, $\rho\left(f, \underline{f}\right) = 1$. Let $u^- := \sup \left\{ u(h) \middle| h \in \left[\underline{f}, \overline{f}\right], \rho(f, h) = 1 \right\}$ and $g := \overline{f}(u^-)\,\underline{f}$ so $u(g) = u^-$. Now, there are $h \in \left[\underline{f}, \overline{f}\right]$ arbitrarily close to g with $u(h) < u(g)$ and $\rho(f, h) = 1$. By continuity, $\rho(f, g) = 1$ so $g = f^-$. To show that f^+ is well-defined, first note that $\rho\left(\overline{f},f\right)=1$ by S-monotonicity. Let $u^{+}:=\inf\left\{u\left(h\right)\middle|\,h\in\left[\underline{f},\overline{f}\right],\rho\left(h,f\right)=1\right\}$ and $g:=\overline{f}\left(u^{+}\right)\underline{f}$ so $u\left(g\right)=u^{+}$. By symmetric argument, $g=f^{+}$. Thus, both f^{-} and f^{+} are well-defined. □

Lemma (2B.6)**.** *$(f^{-},f^{+})=(g^{-},g^{+})$ iff f and g are tied.*

Proof. Note that necessity is trivial so we prove sufficiency. First, suppose f is tied with some $h\in\left[\underline{f},\overline{f}\right]$ so $f^{-}=g^{-}=f^{+}=g^{+}=h$. Since g can at most be tied with a single act in $\left[\underline{f},\overline{f}\right]$, by continuity, g is tied with h and hence f as well. Thus, we assume that both f and g are tied with any act in $\left[\underline{f},\overline{f}\right]$. Define H^{f} and H^{g} as the corresponding H^{*} for $h^{*}=f$ and $h^{*}=g$ respectively. Moreover, define $\mathcal{D}^{f}$ and $\mathcal{D}^{g}$ as the corresponding $\mathcal{D}^{*}$ so by Lemma 2B.3, ρ is linear on both $\mathcal{D}^{f}\subset\mathcal{D}_{0}$ and $\mathcal{D}^{g}\subset\mathcal{D}_{0}$.

Let $\{f_{1},f_{2}\}\subset H^{f}$ and $\{g_{1},g_{2}\}\subset H^{g}$ be such that for some $\varepsilon>0$,

$$\begin{aligned}\varepsilon\mathbf{1}&=u\circ f_{1}-u\circ f=u\circ f-u\circ f_{2}\\&=u\circ g_{1}-u\circ g=u\circ g-u\circ g_{2}\end{aligned}$$

Thus, by linearity, $\rho\left(f_{1},f\right)=\rho\left(f,f_{2}\right)=\rho\left(g_{1},g\right)=\rho\left(g,g_{2}\right)=1$. For ε small enough, linearity yields $\rho\left(f_{1},g^{+}\right)>0=\rho\left(g_{2},g^{+}\right)$ and $1=\rho\left(f_{1},g^{-}\right)>\rho\left(g_{2},g^{-}\right)$. By C-dominance, $\rho\left(f_{1},g_{2}\right)\geq\rho\left(g_{2},g_{2}\right)=1$. By symmetric argument, $\rho\left(g_{1},f_{2}\right)\geq\rho\left(f_{2},f_{2}\right)=1$. Note that if f_{1} and g_{2} are tied, then

$$1=\rho\left(f_{1},g^{-}\right)=\rho\left(g_{2},g^{-}\right)<1$$

a contradiction. Symmetrically, g_{1} and f_{2} cannot be tied so by continuity, $\rho\left(f,g\right)=\rho\left(g,f\right)=1$. Hence, f and g are tied. □

Define $U:Z^{*}\times H\rightarrow\mathbb{R}$ such that

$$U_{z}\left(f\right):=\frac{z\cdot\left(u\circ f^{*}\right)-\left(1-a\right)u\left(h_{0}\right)}{a}$$

where $f^{*}\in H^{*}$ is tied with fah_{0} for some $a\in(0,1]$ and $h_{0}:=\underline{f}\frac{1}{2}\overline{f}$.

Lemma (2B.7)**.** *U_z is well-defined, continuous and*

$$\rho_F(f) = \mu^* \{z \in Z^* | U_z(f) \geq U_z(g) \ \ \forall g \in F\}$$

Proof. We first show that U_z is well defined. Note that by C-linearity, for each $f \in H$, we can find a $a \in (0,1]$ such that there is a $f^* \in H^*$ where $(fah_0)^- = (f^*)^-$ and $(fah_0)^+ = (f^*)^+$. By Lemma 2B.6, fah_0 and f^* are tied. Suppose there is some other $b \in (0,1]$ where $g^* \in H^*$ is tied with fbh_0. Without loss of generality, assume $b < a$ so

$$fbh_0 = (fah_0)\frac{b}{a}h_0$$

Thus, $f^* \frac{b}{a} h_0$ is tied with g^* so $g^* = f^* \frac{b}{a} h_0$ by Lemma 2B.3. Hence,

$$\begin{aligned}\frac{z \cdot (u \circ g^*) - (1-b)\, u(h_0)}{b} &= \frac{\frac{b}{a} z \cdot (u \circ f^*) + \left(1 - \frac{b}{a}\right) u(h_0) - (1-b)\, u(h_0)}{b} \\ &= \frac{z \cdot (u \circ f^*) - (1-a)\, u(h_0)}{a}\end{aligned}$$

so U_z is well-defined.

Now, let $f^* \in H^*$ be tied with fah_0 for all $f \in F$ for some $a \in (0,1)$. By Lemma 2A.3 and Proposition 2B.4,

$$\begin{aligned}\rho_F(f) = \rho_{Fah_0}(fah_0) &= \rho_{F^*}(f^*) \\ &= \mu^* \{z \in Z^* | z \cdot (u \circ f^*) \geq z \cdot (u \circ g^*) \ \ \forall g^* \in F^*\} \\ &= \mu^* \{z \in Z^* | U_z(f) \geq U_z(g) \ \ \forall g \in F\}\end{aligned}$$

Finally, we show that U_z is continuous. Let $f_k \to f$, and first, suppose $f^- < f^+$. Let $f_\delta \in \left[\underline{f}, \overline{f}\right]$ be such that $u(f_\delta) = u(f^-) + \delta$ for $\delta > 0$ such that $\rho(f, f_\delta) \in (0,1)$. Suppose f_k is tied with f_δ for all $k > \bar{k}$ for some $\bar{k} \in \mathbb{N}$. If we let $f_\varepsilon \in \left[\underline{f}, \overline{f}\right]$ be such that $u(f_\varepsilon) = u(f^-) + \varepsilon$ for some $\varepsilon > \delta$ such that $\rho(f, f_\varepsilon) \in (0,1)$, then by continuity,

$$0 = \rho(f_k, f_\varepsilon) = \rho(f, f_\varepsilon)$$

a contradiction. Thus, by continuity, $\rho(f_k, f_\delta) \to \rho(f, f_\delta)$. Since this is true for all $\delta > 0$, we have $(f_k)^- \to f^-$. By symmetric argument, $(f_k)^+ \to f^+$ so $f_k^* \to f^*$ and $U_z(f_k) \to U_z(f)$. Now, if $f^- = f^+ = f^*$, then by similar argument, $\rho(f_k, f_\delta) \to 1$ and $\rho(f_k, f_\varepsilon) \to 0$ for $\{f_\delta, f_\varepsilon\} \subset [\underline{f}, \overline{f}]$ such that $u(f_\delta) = u(f^*) + \delta$ and $u(f_\varepsilon) = u(f^*) - \varepsilon$ for some $\delta > 0$ and $\varepsilon > 0$. Thus, we again have $f_k^* \to f^*$ so $U_z(f_k) \to U_z(f)$. Thus, U_z is continuous. □

Lemma (2B.8)**.** *For all $z \in Z^*$,*

(1) $U_z(fbg) = bU_z(f) + (1-b)U_z(g)$ *for $b \in (0,1)$ and constant g*

(2) $U_z(f) > U_z(g) > U_z(h)$ *implies there are $\{a,b\} \subset (0,1)$ such that* $U_z(fah) > U_z(g) > U_z(fbh)$

(3) $U_z(f) > U_z(g)$ *for some f and g*

Proof. Let $z \in Z^*$. Note that (2) follows directly from the continuity of U_z (Lemma 2B.7) and (3) follows immediately from the fact that $U_z(\overline{f}) = u(\overline{f}) > u(\underline{f}) = U_z(\underline{f})$. We now show (1). Let $\{f^*, h^*, g^*\} \subset H^*$ be tied with fah_0, hah_0 and gah_0 respectively where $h = fbg$ for $b \in (0,1)$ and constant g. First, suppose $g \in [\underline{f}, \overline{f}]$ so by C-linearity, $h^* = f^*bg^*$ and

$$U_z(h) = \frac{z \cdot (u \circ h^*) - (1-a)u(h_0)}{a} = bU_z(f) + (1-b)U_z(g)$$

Now, for any constant g, let $\hat{g} \in [\underline{f}, \overline{f}]$ be such that $u \circ g = u \circ \hat{g}$ so $u \circ (fbg) = u \circ (fb\hat{g})$. By S-monotonicity, g is tied with $\hat{g}$ and fbg is tied with $fb\hat{g}$ so

$$\begin{aligned} U_z(fbg) &= U_z(fb\hat{g}) = bU_z(f) + (1-b)U_z(\hat{g}) \\ &= bU_z(f) + (1-b)U_z(g) \end{aligned}$$

□

Lemma (2B.9)**.** *For every $\{f,g\} \subset H$, there exists $Z_{fg} \subset Z^*$ such that $\mu^*(Z_{fg}) = 1$ and for all $z \in Z_{fg}$,*

(1) $U_z(fbg) \geq \min\{U_z(f), U_z(g)\}$ *for all $b \in (0,1)$*

(2) $U_z(f_s) \geq U_z(g_s)$ *for all $s \in S$ implies* $U_z(f) \geq U_z(g)$

Proof. Let $\{f, g\} \subset H$ and $b \in (0,1)$. First, suppose f and g are not tied and let

$$Z^1_{fg}(b) := \{z \in Z^* | \, U_z(fbg) \geq \min\{U_z(f), U_z(g)\}\}$$

By convexity,

$$\begin{aligned} 0 &= \rho(f, fbg) - \rho(f, (fbg) \cup g) \\ &= \mu^*\{z \in Z^* | \, U_z(f) \geq U_z(fbg)\} - \mu^*\{z \in Z^* | \, U_z(f) \geq \max(U_z(fbg), U_z(g))\} \\ &= \mu^*\{z \in Z^* | \, U_z(g) > U_z(f) \geq U_z(fbg)\} \end{aligned}$$

and by symmetric argument, $\mu^*\{z \in Z^* | \, U_z(f) > U_z(g) \geq U_z(fbg)\} = 0$. Since f and g are not tied,

$$\mu^*\{z \in Z^* | \, U_z(f) = U_z(g)\} = 0$$

so $\mu^*\left(Z^1_{fg}(b)\right) = 1$.

Now, suppose f and g are tied. Let $f_k \in H$ be such that $f_k \to f$ and

$$u \circ f = u \circ f_k + \varepsilon_k \mathbf{1}$$

for $\varepsilon_k > 0$. Now, by C-linearity from Lemma 2B.8, $U_z(f) > U_z(f_k)$ for all $z \in Z^*$ so $U_z(g) > U_z(f_k)$ for all $z \in Z^*$ as f and g are tied. Thus, g and f_k are not tied, so by the above, $\mu^*\left(Z^1_{f_k g}(b)\right) = 1$ for all k. Now define

$$Z^1_{fg}(b) := \bigcap_k Z^1_{f_k g}(b)$$

so $\mu^*\left(Z^1_{fg}(b)\right) = 1$. For $z \in Z^1_{fg}(b)$, $U_z(f_k bg) \geq \min\{U_z(f_k), U_z(g)\}$ for all k so by the continuity of U_z, $U_z(fbg) \geq \min\{U_z(f), U_z(g)\}$.

Thus, for each $b \in (0,1)$, we can find a $Z^1_{fg}(b) \subset Z^*$ such that $\mu^*\left(Z^1_{fg}(b)\right) = 1$ and $U_z(fbg) \geq \min\{U_z(f), U_z(g)\}$ for all $z \in Z^1_{fg}(b)$. Now, let $\mathbb{Q}_{(0,1)}$ be all the rationals on $(0,1)$ and

$$Z^1_{fg} := \bigcap_{b \in \mathbb{Q}_{(0,1)}} Z^1_{fg}(b)$$

Note that $\mu^*\left(Z^1_{fg}\right) = 1$ and the continuity of U_z ensures that (1) is satisfied for all $z \in Z^1_{fg}$.

Finally, note that $u(h_s) = U_z(h_s)$ for all $h \in H$ and $z \in Z^*$. Define

$$Z^2_{fg} := \{z \in Z^* | U_z(f) \geq U_z(g)\}$$

if $u(f_s) \geq u(g_s)$ for all $s \in S$ and $Z^2_{fg} := Z^*$ otherwise. Let

$$Z_{fg} := Z^1_{fg} \cap Z^2_{fg}$$

First, suppose $u(f_s) \geq u(g_s)$ for all $s \in S$ so

$$1 = \mu^* \{z \in Z^* | U_z(f_s) \geq U_z(g_s)\} = \rho(f_s, g_s)$$

for all $s \in S$. By S-monotonicity, $1 = \rho(f, g) = \mu^*\left(Z^1_{fg}\right)$. Thus, $\mu^*(Z_{fg}) = 1$ and both (1) and (2) are satisfied for all $z \in Z_{fg}$. Now, if there is some $s \in S$ such that $u(f_s) < u(g_s)$, then $Z_{fg} = Z^1_{fg}$ so again $\mu^*(Z_{fg}) = 1$ and both (1) and (2) are satisfied for all $z \in Z_{fg}$. Note that in the case of the latter, (1) is trivially satisfied. ☐

For $K \in \mathcal{K}$, define

$$U_K(f) := \min_{q \in K} q \cdot (u \circ f)$$

Lemma (2B.10)**.** *There is a $K : Z \to \mathcal{K}$ for some $Z \subset Z^*$ such that $\mu^*(Z) = 1$ and*

$$\rho_F(f) = \mu^* \{z \in Z | U_{K_z}(f) \geq U_{K_z}(g) \ \ \forall g \in F\}$$

Moreover, f and g are tied iff $U_{K_z}(f) = U_{K_z}(g)$ for all $z \in Z$.

Proof. Let $H_q \subset H$ be a countable dense subset of H. For example, H_q could be the set of all acts with rational lotteries. Let $Z_{fg} \subset Z^*$ be defined as in Lemma 2B.9 so $\mu^*(Z_{fg}) = 1$. Let

$$Z_0 := \bigcap_{(f,g) \in H_q \times H_q} Z_{fg}$$

so $\mu^*(Z_0) = 1$. Let $z \in Z_0$ and $\{f, g\} \subset H$. Since H_q is dense in H, we can find $f_k \to f$ and $g_k \to g$ for $(f_k, g_k) \in H_q \times H_q$ for all k. Since $z \in Z_0$, both conditions of Lemma 2B.9

are satisfied for all f_k and g_k. The continuity of U_z then ensures that both conditions are again satisfied for f and g. Thus, all the conditions of Lemmas 2B.8 and 2B.9 are satisfied for all $z \in Z_0$. Note that they correspond exactly to the six axioms of maxmin expected utility, so by Theorem 1 of Gilboa and Schmeidler [40], for each $z \in Z_0$, there is a $K_z \in \mathcal{K}$ and non-constant affine $v_z : \Delta X \to \mathbb{R}$ such that

$$U_z(f) = \phi_z\left(\min_{q \in K_z} q \cdot (v_z \circ f)\right)$$

for some increasing monotone transformation $\phi_z : \mathbb{R} \to \mathbb{R}$. By Lemma 2B.7 as $\mu^*(Z_0) = 1$,

$$\begin{aligned}
\rho_F(f) &= \mu^*\left\{z \in Z^* \middle| U_z(f) \geq U_z(g) \ \ \forall g \in F\right\} \\
&= \mu^*\left\{z \in Z_0 \middle| \min_{q \in K_z} q \cdot (v_z \circ f) \geq \min_{q \in K_z} q \cdot (v_z \circ g) \ \ \forall g \in F\right\}
\end{aligned}$$

Now, without loss of generality, assume $v_z\left(\underline{f}\right) = 0$ and $v_z\left(\overline{f}\right) = 1$ for all $z \in Z_0$. For every $f \in H_c$, we can find $a_f \in [0,1]$ such that f and $\overline{f} a_f \underline{f}$ are tied. Since $\min_{q \in K_z} q \cdot (v_z \circ f) = v_z(f)$, we have

$$1 = \mu^*\left\{z \in Z_0 \middle| v_z(f) = v_z\left(\overline{f} a_f \underline{f}\right) = a_f\right\}$$

Let $Z_f := \left\{z \in Z_0 \middle| v_z(f) = a_f\right\}$ so $\mu^*(Z_f) = 1$ for all $f \in H_c$. Define

$$Z_1 := Z_0 \cap \bigcap_{f \in H_c \cap H_q} Z_f$$

so $\mu^*(Z_1) = 1$. Moreover, by the continuity of v_z, $v_z(f) = a_f$ for all $z \in Z_1$ and $f \in H_c$. If we fix a $z_1 \in Z_1$ and let $v := v_{z_1}$, then

$$v_z(f) = a_f = v(f)$$

for all $z \in Z_1$. Thus, by Lemma 2B.2, $v = \alpha u + \beta$ for some $\alpha > 0$. Hence,

$$\rho_F(f) = \mu^*\left\{z \in Z_1 \middle| U_{K_z}(f) \geq U_{K_z}(g) \ \ \forall g \in F\right\}$$

with $\mu^*(Z_1) = 1$.

Finally, for $\{f,g\} \subset H$, define

$$Z_{fg} := \{ z \in Z_1 | \, U_{K_z}(f) = U_{K_z}(g) \}$$

Let H^f and H^g be the corresponding H^* for $h^* = f$ and $h^* = g$ respectively and let $(f^*, g^*) \in H^f \times H^g$ be such that f^* and g^* are tied. Define

$$Z := \bigcap_{(f,g) \in H_q \times H_q} Z_{f^* g^*}$$

Since f^* and g^* are tied iff $\mu^*(Z_{f^*g^*}) = 1$, we have $\mu^*(Z) = 1$. Thus, from above,

$$\rho_F(f) = \mu^* \{ z \in Z | \, U_{K_z}(f) \geq U_{K_z}(g) \ \ \forall g \in F \}$$

Note that $U_{K_z}(f) = U_{K_z}(g)$ for all $z \in Z$ implies f and g are tied. For the converse, suppose f and g are tied and let $z \in Z$. Since we can always find $\hat{f} \in H_q \cap H^f$ and $\hat{g} \in H_q \cap H^g$, let $\hat{f}^* \in H^f$ and $\hat{g}^* \in H^g$ be tied. As $z \in Z$, $U_{K_z}\left(\hat{f}^*\right) = U_{K_z}(\hat{g}^*)$. Thus, by C-linearity, $U_{K_z}(f) = U_{K_z}(g)$. □

Theorem (2B.11)**.** *If ρ satisfies Axioms 2.1-2.7, then it is has a random ambiguity representation.*

Proof. First, note that if H^* is degenerate, then $Z \subset Z^*$ is a singleton and the result follows trivially. Thus, assume H^* is not degenerate. Fix $h^* \in H^*$ and let $\left(\underline{h}, \overline{h}\right) := \left((h^*)^-, (h^*)^+\right)$ so $u(\underline{h}) < u\left(\overline{h}\right)$. Note that

$$\begin{aligned} 1 &= \rho(h^*, \underline{h}) = \mu^* \{ z \in Z | \, U_{K_z}(h^*) \geq u(\underline{h}) \} \\ &= \rho\left(\overline{h}, h^*\right) = \mu^* \left\{ z \in Z | \, u\left(\overline{h}\right) \geq U_{K_z}(h^*) \right\} \end{aligned}$$

Thus, if we let

$$\hat{Z} := \left\{ z \in Z | \, u(\underline{h}) \leq U_{K_z}(h^*) \leq u\left(\overline{h}\right) \right\}$$

then $\mu^*\left(\hat{Z}\right) = 1$.

Now, for each $z \in \hat{Z}$, let $\lambda : \hat{Z} \to [0,1]$ be such that

$$U_{K_z}(h^*) = \lambda_z u(\overline{h}) + (1-\lambda_z) u(\underline{h})$$

Consider $\{x, y\} \subset \hat{Z}$ such that $\lambda_x \leq \lambda_y$ so $U_{K_x}(h^*) \leq U_{K_y}(h^*)$. For $f \in H$, let $h^* a h_0$ be tied with fbh for $\{a, b\} \subset [0,1]$ and $h \in [\underline{f}, \overline{f}]$. By Lemma 2B.10 and C-linearity,

$$\begin{aligned} aU_{K_z}(h^*) + (1-a) u(h_0) &= U_{K_z}(h^* a h_0) \\ &= U_{K_z}(fbh) = bU_{K_z}(f) + (1-b) u(h) \end{aligned}$$

for all $z \in \hat{Z}$. Since $U_{K_x}(h^*) \leq U_{K_y}(h^*)$, we have $U_{K_x}(f) \leq U_{K_y}(f)$ for all $z \in \hat{Z}$ and $f \in H$. Thus, $\lambda_x \leq \lambda_y$ implies $K_y \subset K_x$ for all $\{x, y\} \subset \hat{Z}$.

Hence $K_y \subset K_x$ or $K_x \subset K_y$ for all $\{x, y\} \subset \hat{Z}$. We can now let $\{\underline{K}, \overline{K}\} \subset \mathcal{K}$ be such that $U_{\underline{K}}(f) = u(f^+)$ and $U_{\overline{K}}(f) = u(f^-)$ for all $f \in H$. Let $t : \hat{Z} \to [0,1]$ be such that

$$U_{K_z}(h^*) = t_z U_{\overline{K}}(h^*) + (1-t_z) U_{\underline{K}}(h^*) = U_{t_z \overline{K} + (1-t_z)\underline{K}}(h^*)$$

For $f \in H$, as before, let $h := h^* a h_0$ be tied with $g := fbh$ for $\{a, b\} \subset [0,1]$ and $h \in [\underline{f}, \overline{f}]$. Since h and g are tied, by Lemma 2B.6, $(h^-, h^+) = (g^-, g^+)$ so $U_{\underline{K}}(h) = U_{\underline{K}}(g)$ and $U_{\overline{K}}(h) = U_{\overline{K}}(g)$. Now, by C-linearity,

$$\begin{aligned} U_{K_z}(h) = U_{K_z}(h^* a h_0) &= aU_{K_z}(h^*) + (1-a) u(h_0) \\ &= a\left(t_z U_{\overline{K}}(h^*) + (1-t_z) U_{\underline{K}}(h^*)\right) + (1-a) u(h_0) \\ &= t_z U_{\overline{K}}(h) + (1-t_z) U_{\underline{K}}(h) \end{aligned}$$

Thus, by Lemma 2B.10,

$$U_{K_z}(g) = U_{K_z}(h) = t_z U_{\overline{K}}(g) + (1-t_z) U_{\underline{K}}(g)$$

By C-linearity again, $U_{K_z}(f) = t_z U_{\overline{K}}(f) + (1-t_z) U_{\underline{K}}(f) = U_{K_{t_z}}(f)$ where $K_{t_z} := t_z \overline{K} +$

$(1-t_z)\underline{K}$. If we let $\mu := \mu^* \circ (K_{t_z})^{-1}$ be the distribution on $\mathcal{K}$ induced by K_{t_z}, then

$$\begin{aligned}\rho_F(f) &= \mu^* \left\{ z \in \hat{Z} \middle| U_{K_z}(f) \geq U_{K_z}(g) \ \ \forall g \in F \right\} \\ &= \mu \left\{ K_t \in \mathcal{K}_T \middle| U_{K_t}(f) \geq U_{K_t}(g) \ \ \forall g \in F \right\}\end{aligned}$$

□

Theorem (2B.12)**.** *If ρ has a random ambiguity representation, then it satisfies Axioms 2.1-2.7.*

Proof. Let ρ be represented by (μ, u). We prove that it satisfies Axioms 2.1-2.7 in order.

(1) Note that for constant h and $a \in (0,1)$, $U_{K_t}(fah) = aU_{K_t}(f) + (1-a)U_{K_t}(h)$ so C-linearity follows immediately from the representation.

(2) Let $F_k \to F$ for $\{F_k, F\} \subset \mathcal{D}_0$ for all k. First, consider $\{f, g\} \subset F_k$ such that $f \neq g$ and suppose $U_{K_t}(f) = U_{K_t}(g)$ μ-a.s. so f and g are tied. As ρ is monotone (see (6) below), Lemma 2A.2 implies $g \in f_{F_k}$ contradicting the fact that $F_k \in \mathcal{K}_0$. Hence, as μ is regular, $U_{K_t}(f) = U_{K_t}(f)$ with μ-measure zero and the same holds for $\{f, g\} \subset F$. Now, for $G \in \mathcal{D}$, let

$$\mathcal{K}_G := \bigcup_{\{f,g\} \subset G, \ f \neq g} \left\{ K_t \in \mathcal{K}_T \middle| U_{K_t}(f) = U_{K_t}(f) \right\}$$

and let

$$\hat{\mathcal{K}} := \mathcal{K}_F \cup \bigcup_k \mathcal{K}_{F_k}$$

Thus, $\mu\left(\hat{\mathcal{K}}\right) = 0$ so $\mu\left(\bar{\mathcal{K}}\right) = 1$ for $\bar{\mathcal{K}} := \mathcal{K}_T \backslash \hat{\mathcal{K}}$. Let $\bar{\mu}(A) = \mu(A)$ for $A \in \mathcal{B}(\mathcal{K}_T) \cap \bar{\mathcal{K}}$. Thus, $\bar{\mu}$ is the restriction of μ to $\bar{\mathcal{K}}$ (see Exercise I.3.11 of Çinlar [18]).

Now, for each F_k, let $\xi_k : \bar{\mathcal{K}} \to H$ be such that

$$\xi_k(K_t) := \arg\max_{f \in F_k} U_{K_t}(f)$$

and define ξ similarly for F. Note that both ξ_k and ξ are well-defined. For any

$B \in \mathcal{B}(H)$,

$$\begin{aligned}
\xi_k^{-1}(B) &= \left\{ K_t \in \bar{\mathcal{K}} \middle| \, \xi_k(K_t) \in B \cap F_k \right\} \\
&= \bigcup_{f \in B \cap F_k} \{ K_t \in \mathcal{K}_T | \, U_{K_t}(f) > U_{K_t}(g) \;\; \forall g \in F_k \} \cap \bar{\mathcal{K}} \in \mathcal{B}(\mathcal{K}_T) \cap \bar{\mathcal{K}}
\end{aligned}$$

Hence, ξ_k and ξ are random variables with respect to $\mathcal{B}(\mathcal{K}_T) \cap \bar{\mathcal{K}}$. Moreover,

$$\begin{aligned}
\bar{\mu} \circ \xi_k^{-1}(B) &= \sum_{f \in B \cap F_k} \bar{\mu} \left\{ K_t \in \bar{\mathcal{K}} \middle| \, U_{K_t}(f) > U_{K_t}(g) \;\; \forall g \in F_k \right\} \\
&= \sum_{f \in B \cap F_k} \mu \{ K_t \in \mathcal{K}_T | \, U_{K_t}(f) \geq U_{K_t}(g) \;\; \forall g \in F_k \} \\
&= \rho_{F_k}(B \cap F_k) = \rho_{F_k}(B)
\end{aligned}$$

so ρ_{F_k} and ρ_F are the distributions of ξ_k and ξ respectively. Finally, let $F_k \to F$ and fix $K_t \in \bar{\mathcal{K}}$. Let $f := \xi(K_t)$ so $U_{K_t}(f) > U_{K_t}(g)$ for all $g \in F$. Since U_{K_t} is continuous, there is some $l \in \mathbb{N}$ such that $U_{K_t}(f_k) > U_{K_t}(g_k)$ for all $k > l$. Thus, $\xi_k(K_t) = f_k \to f = \xi(K_t)$ so ξ_k converges to ξ $\bar{\mu}$-a.s.. Since almost sure convergence implies convergence in distribution (see Exercise III.5.29 of Çinlar [18]), $\rho_{F_k} \to \rho_F$ and continuity is satisfied.

(3) Suppose $\rho_{F_s}(f_s) = 1$ for all $s \in S$. Thus, $u(f_s) \geq u(g_s)$ for all $g \in F$ and $s \in S$ which implies

$$U_{K_t}(f) = \min_{q \in K_t} q \cdot (u \circ f) \geq \min_{q \in K_t} q \cdot (u \circ g) = U_{K_t}(g)$$

for all $g \in F$. Hence, $\rho_F(f) = 1$.

(4) Let $\{f, fah\} \subset F$ for $a \in (0,1)$. Now,

$$\begin{aligned}
\rho_F(f) - \rho_{F \cup h}(f) =& \mu \{ K_t \in \mathcal{K}_T | \, U_{K_t}(f) \geq U_{K_t}(g) \;\; \forall g \in F \} \\
& - \mu \{ K_t \in \mathcal{K}_T | \, U_{K_t}(f) \geq U_{K_t}(g) \;\; \forall g \in F \cup h \} \\
=& \mu \{ K_t \in \mathcal{K}_T | \, U_{K_t}(h) > U_{K_t}(f) \geq U_{K_t}(g) \;\; \forall g \in F \}
\end{aligned}$$

Since $fah \in F$ and U_{K_t} is concave, $U_{K_t}(f) \geq U_{K_t}(fah)$ implies

$$U_{K_t}(f) \geq U_{K_t}(fah) \geq aU_{K_t}(f) + (1-a)\,U_{K_t}(h)$$

so $U_{K_t}(f) \geq U_{K_t}(h)$ μ-a.s.. Thus, $\rho_F(f) = \rho_{F\cup h}(f)$.

(5) Let $\{f, g\} \subset H_c$ be such that $u(f) > u(g)$ as u is non-constant. Thus

$$\rho(g, f) = \mu\{K_t \in \mathcal{K}_T |\, U_{K_t}(g) \geq U_{K_t}(f)\} = 0$$

(6) Monotonicity follows immediately from the random utility representation.

(7) Note that

$$U_{K_t}(f) = tU_{\overline{K}}(f) + (1-t)\,U_{\underline{K}}(f) = U_{\underline{K}}(f) + t\,(U_{\overline{K}}(f) - U_{\underline{K}}(f))$$

Thus, if we let $a_f := U_{\underline{K}}(f)$, $b_f := U_{\overline{K}}(f) - U_{\underline{K}}(f)$ and $\nu := \mu \circ K_t$ be the measure on T induced by K_t, then

$$\begin{aligned}\rho_F(f) &= \mu\{K_t \in \mathcal{K}_T |\, U_{K_t}(f) \geq U_{K_t}(g)\ \ \forall g \in F\} \\ &= \nu\{t \in T|\, a_f + b_f t \geq a_g + b_g t\}\end{aligned}$$

Note that since $\underline{K} \subset \overline{K}$, $b_f \leq 0$. Moreover, if f is constant, then $b_f = 0$. Now, let h and h' be constant and suppose $\rho(f, h) > \rho(g, h) = 0$ and $1 = \rho(f, h') > \rho(g, h')$. Thus, $a_h \geq a_g + b_g t$ and $a_f + b_f t \geq a_{h'}$ ν-a.s. while $a_f + b_f t \geq a_h$ and $a_{h'} \geq a_g + b_g t$ with strictly positive ν-measure. Hence, $a_h > a_{h'}$ and since $b_f \leq 0$ and $b_g \leq 0$, we have $a_f + b_f t \geq a_g + b_g t$ ν-a.s.. Thus, $\rho(f, F) \geq \rho(g, F)$ for all $F \in \mathcal{D}$.

□

Theorem (2B.13)**.** *Let ρ and τ be represented by (μ, u) and (ν, v) respectively. Then the following are equivalent*

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

(2) $\rho = \tau$

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

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

We now show that (3) implies (1). By Theorem 2B.12, ρ and τ both satisfy Axioms 2.1-2.7. Thus, by Lemma 2B.2, $u = \alpha v + \beta$ for $\alpha > 0$. Without loss of generality, we can assume $1 = u(\overline{f}) = v(\overline{f})$ and $0 = u(\underline{f}) = v(\underline{f})$ so $u = v$. Let H_ρ^* be defined for ρ. Note that if H_ρ^* is degenerate, then ρ is deterministic and $\mu(K_1) = 1$ for some $K_1 \in \mathcal{K}$. By (3), τ is also deterministic so $\nu(K_2) = 1$ for some $K_2 \in \mathcal{K}$. By the uniqueness properties of maxmin expected utility (Theorem 1 of Gilboa and Schmeidler [40]), $K_1 = K_2$ so $\mu = \nu$.

Now, assume H_ρ^* is not degenerate. Let $h^* \in H_\rho^*$ so by (3), $H^* = H_\rho^* = H_\tau^{h^*}$ is also not degenerate. Thus, let μ^* and ν^* be the measures induced on Z^* by μ and ν respectively. Note that we can assume $Z^* \subset \Delta S$ and for any $F \in \mathcal{D}^*$,

$$\rho_F(f) = \mu^* \{ z \in Z^* | \, z \cdot (u \circ f) \geq z \cdot (u \circ g) \;\; \forall g \in F \}$$
$$\tau_F(f) = \nu^* \{ z \in Z^* | \, z \cdot (u \circ f) \geq z \cdot (u \circ g) \;\; \forall g \in F \}$$

By Ionescu-Tulcea's extension (Theorem IV.4.7 of Çinlar [18]), we can create a probability space on Ω with two independent random variables $X : \Omega \to Z^*$ and $Y : \Omega \to Z^*$ such that they have distributions μ^* and ν^* respectively. For $f \in H$, let $\psi_f : Z^* \to \mathbb{R}$ be such that $\psi_r(z) := z \cdot (u \circ f)$. Let $\mu_f^* = \mu^* \circ \psi_f^{-1}$ and $\nu_f^* = \nu^* \circ \psi_f^{-1}$. Note that for $a \in [0,1]$, by (3),

$$\begin{aligned} \mu_f^*(-\infty, a) &= \mu^* \circ \psi_f^{-1}(-\infty, a) = \mu^* \left\{ z \in Z^* | \, z \cdot (u \circ f) \leq u\left(\overline{f} a \underline{f}\right) \right\} \\ &= \rho\left(f, \overline{f} a \underline{f}\right) = \tau\left(f, \overline{f} a \underline{f}\right) = \nu_f^*(-\infty, a) \end{aligned}$$

so $\mu_f^* = \nu_f^*$ for all $f \in H$. Now, for all $f \in H$,

$$\begin{aligned}\mathbb{E}\left[e^{-(u\circ f)\cdot X}\right] &= \int_{Z^*} e^{-(u\circ f)\cdot z}\mu^*(dz) = \int_{[0,1]} \mu_f^*(da)\, e^{-a} \\ &= \int_{[0,1]} e^{-a}\nu_f^*(da) = \int_{Z^*} e^{-(u\circ f)\cdot z}\nu^*(dz) = \mathbb{E}\left[e^{-(u\circ f)\cdot Y}\right]\end{aligned}$$

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 $w = u \circ f \in [0,1]^S$. As Laplace transforms completely characterize distributions (see Exercise II.2.36 of Çinlar [18]), X and Y have the same distribution, so $\mu^* = \nu^*$. Thus,

$$\mu = \mu^* \circ (K_{t_z})^{-1} = \nu^* \circ (K_{t_z})^{-1} = \nu$$

so (1) is true. Thus, (1), (2) and (3) are all equivalent. □

Proposition (2B.14)**.** *Let ρ and τ be represented by (μ, u) and (ν, v) respectively. Then ρ is more uncertainty averse than τ iff $\nu \gg \mu$ and $u = \alpha + \beta v$ for $\beta > 0$.*

Proof. We first show necessity. Assume $u = v$ without loss of generality and $\nu \gg \mu$. If $f \notin F$, then $\rho_F(f) = \tau_F(f) = 0$ trivially so assume $f \in F$ for some $f \in H_c$. Now

$$\rho_F(f) = \mu\left\{K_t \in \mathcal{K}_T \middle| u(f) \geq U_{K_t}(g)\right\} = 1 - \mu\left\{K \in \mathcal{K} \middle| K \subset K_{\bar{t}}\right\}$$

for some $\bar{t} \in [0,1]$. Since $\mu \gg \nu$, $\nu\{K \subset K_{\bar{t}}\} \geq \mu\{K \subset K_{\bar{t}}\}$ so

$$\begin{aligned}\tau_F(f) &= 1 - \nu\{K \subset K_{\bar{t}}\} \\ &\leq 1 - \mu\{K \subset K_{\bar{t}}\} = \rho_F(f)\end{aligned}$$

so ρ is more uncertainty averse than τ.

Now assume ρ is more uncertainty averse than τ. Suppose $u \neq \alpha + \beta v$ for $\beta > 0$ so we can find $\{f, g\} \subset H_c$ such that $\tau(f,g) = \tau(g,f) = 1$ but $0 = \rho(f,g) < \rho(g,f) = 1$. Thus $\rho(f,g) \geq \tau(f,g)$ contradicting the fact that ρ is more uncertainty averse than τ. Hence, we can assume $u = v$ without loss of generality. Set $L \in \mathcal{K}$ and let

$$\bar{t} := \sup\left\{t \in T \middle| K_t \subset L\right\}$$

Note that

$$\{K_t \in \mathcal{K}_T |\, K_t \subset K_{\bar{t}}\} = \{K_t \in \mathcal{K}_T |\, t \leq \bar{t}\} \subset \{K_t \in \mathcal{K}_T |\, K_t \subset L\}$$

If $K_t \in \{K_t \in \mathcal{K}_T |\, K_t \subset L\}$, then $t \leq \bar{t}$ so $K_t \in \{K_t \in \mathcal{K}_T |\, K_t \subset K_{\bar{t}}\}$. Thus,

$$\{K \subset L\} = \{K_t \in \mathcal{K}_T |\, K_t \subset K_{\bar{t}}\}$$

Now, for $K \in \mathcal{K}$, let

$$S(K, v) := K \cap \left\{ q \in \Delta S |\, \min_{p \in K} p \cdot v = q \cdot v \right\}$$

denote the support set of K at $v \in [0,1]^S$. Let

$$S(K) := \bigcup_{v \in [0,1]^S} S(K, v)$$

be the union of all the support sets of K. Fix $h_0 := \underline{f} \frac{1}{2} \overline{f}$ and for each $q \in S(K_{\bar{t}})$, let $f_q \in H$ be such that

$$U_{K_{\bar{t}}}(f_q) = q \cdot (u \circ f_q) = u(h_0)$$

Let $F := \{f_q \in H |\, q \in S(K_{\bar{t}})\}$ and let $F_0 \subset F$ be a countable dense subset of F. Now, we can find $F_k \subset F$ and $F_k \in \mathcal{D}$ such that $F_k \subset F_{k+1}$ and $F_k \nearrow F_0$. Thus,

$$\begin{aligned} \rho(h_0, F_k) &= \mu\{K_t \in \mathcal{K}_T |\, U_{K_{\bar{t}}}(f) = u(h_0) \geq U_{K_t}(f) \;\; \forall f \in F_k\} \\ &\geq \tau(h_0, F_k) = \nu\{K_t \in \mathcal{K}_T |\, U_{K_{\bar{t}}}(f) = u(h_0) \geq U_{K_t}(f) \;\; \forall f \in F_k\} \end{aligned}$$

For $G \subset H$, let

$$\mathcal{K}_G := \{K_t \in \mathcal{K}_T |\, U_{K_{\bar{t}}}(f) < U_{K_t}(f) \;\; \forall f \in G\}$$

so $\mu(\mathcal{K}_{F_k}) \leq \nu(\mathcal{K}_{F_k})$ for all k. Since $G \subset G'$ implies $\mathcal{K}_{G'} \subset \mathcal{K}_G$, $F_k \nearrow F_0$ implies $\mathcal{K}_{F_k} \searrow \mathcal{K}_{F_0}$ so

$$\mu(\mathcal{K}_{F_0}) = \lim_k \mu(\mathcal{K}_{F_k}) \leq \lim_k \nu(\mathcal{K}_{F_k}) = \nu(\mathcal{K}_{F_0})$$

Since F_0 is dense in F and U_{K_t} is continuous, we have $\mu\left(\mathcal{K}_F\right) \leq \nu\left(\mathcal{K}_F\right)$. Hence, from above,

$$\begin{aligned}\mu\left\{K \subset L\right\} &= \mu\left\{K_t \in \mathcal{K}_T \middle| K_t \subset K_{\bar{t}}\right\} \\ &\leq \nu\left\{K_t \in \mathcal{K}_T \middle| K_t \subset K_{\bar{t}}\right\} = \nu\left\{K \subset L\right\}\end{aligned}$$

so $\nu \gg \mu$. □

3 Belief Persistence and the Disposition Effect

3.1 Introduction

One of the most robust findings in behavioral finance is the disposition effect. This refers to the tendency of individual investors to oversell stocks that have gone up in value (i.e. price) and to undersell stocks that gone down. Although suboptimal, this behavior has been well-documented in many situations and in various markets around the world.[46] Moreover, similar observations have also been recorded in other contexts, such as the housing market or in the exercise of executive stock options.[47]

In this chapter, we consider a belief-based explanation for the disposition effect. In particular, we study a model of heterogeneous beliefs where beliefs differ only along a single dimension measuring *persistence*. For example, some agents may believe that earnings information from this quarter will be highly correlated with earnings information from the next quarter (high persistence). On the other hand, others may believe that the earnings information from both quarters are completely uncorrelated (no persistence). In equilibrium, agents who believe in the least persistence decrease their holdings of the stock when prices rise and increase their holdings when prices fall; in essence, they exhibit the disposition effect. On the other hand, those who believe in the most persistence exhibit the opposite behavior; they employ a trading strategy based on stock price momentum and exhibit a form of the house-money effect.[48]

To be concrete, suppose that institutional investors pay close attention to certain financial indicators after each earnings release which they believe to be highly correlated with future earnings information. Individual investors on the other hand, believe these indicators to be noisy and ignore them when placing their trades. In equilibrium, our model predicts that

[46] Studies by Odean [66] in the U.S., Grinblatt and Keloharju [45] in Finland and Feng and Seasholes [32] in China confirm that the disposition effect is a global phenomenon that transcends both institutional and cultural divisions.

[47] See Genesove and Mayer [38] and Heath, Huddart and Lang [49] respectively.

[48] The house-money effect refers to the tendency of gamblers to increase their bets after a gain and to decrease their bets after a loss (see Thaler and Johnson [79]).

individual investors will exhibit the disposition effect while institutional investors will engage in momentum trading.[49]

Much of the existing research has focused on providing a preference-based explanation to rationalize the disposition effect.[50] Our results show that by introducing belief heterogeneity in a model with equilibrium trading, we can obtain behavior that exhibits the disposition effect while still retaining the standard assumption of expected utility preferences. While our approach does not insist that a belief-based equilibrium model is the *only* explanation for the disposition effect, it does suggest that in many cases, a careful study of the disposition effect should also take into account investor beliefs and how they interact with equilibrium forces. We leave the more practical exercises of testing different explanations of the disposition effect as avenues for future research.

Formally, we consider a simple two-period model where agents form beliefs about the state space S in each period. For example, S could represent all the financial information available after each earnings release. We assume that agents share the same correct prior about S, but have different beliefs about the transition probabilities between the two time periods. We model these beliefs as Markov kernels on S. For example, some agents may believe that state realizations in the two periods are highly correlated (i.e. high persistence) while others may believe that there is little correlation (i.e. low persistence). We say that the distribution of beliefs in a population has a *persistence representation* iff each agent's belief corresponds to a Markov kernel

$$K = \lambda \overline{K} + (1 - \lambda) \underline{K}$$

where $\underline{K}$ is the constant kernel, $\overline{K}$ is some other fixed kernel and $\lambda \in [0, 1]$. Thus, all agents in the population can be parametrized by some $\lambda \in [0, 1]$ measuring each agent's belief of state persistence. For example, an agent with $\lambda = 0$ believes in zero persistence. In the case

[49] This is somewhat consistent with the empirical evidence suggesting that the disposition effect is much more pronounced for investors who are financially less sophisticated (see Dhar and Zhu [22] for example).

[50] See Barberis and Xiong [5], Strahilevitz, Odean and Barber [77] and Barberis and Xiong [6] for explanations based on prospect theory, emotional regret and realization utility respectively.

where S is finite, K is simply the λ-mixture of the two matrices $\overline{K}$ and $\underline{K}$. Note that this is not a model of asymmetric information; after the first period, all agents observe the realized state $s \in S$ but form their own individual beliefs about what that means for second period state realizations.

We consider an equilibrium where agents trade claims to two securities: a risky "stock" that yields a payoff contingent on the realization of the state $s \in S$ and a risk-free "bond" that yields the same payoff regardless of which state is realized. To simplify the exposition, we assume that all agents have the same endowment in each state and that they share the same CARA utility index. An equilibrium is *ordered* iff all agents can be ranked according to the degree of disposition effect they exhibit. Our first main result shows that any equilibrium under a persistence representation is ordered. In this case, those who believe in the least persistence (i.e. smallest λ) exhibit the disposition effect while those who believe in the most persistence (i.e. largest λ) exhibit the opposite behavior (i.e. momentum trading or a form of the house-money effect). Thus, in our model, the disposition effect emerges as equilibrium behavior induced by belief heterogeneity.

We then study a special case where beliefs are Gaussian. There are two distinct groups in the population: the first group believes that states are independent and identically distributed (i.i.d.) while the second believes that states are correlated. This CARA-Gaussian setup is easily tractable and allows for simple expressions for both prices and trading strategies. In equilibrium, a simple inequality characterizes when stock prices will increase (or decrease) and when agents in the first group (i.e. those who believe in no persistence) will decrease (or increase) their stock holdings. We then proceed to analyze some of the comparative statics. We show that increasing the proportion of agents in the second group (i.e. those who believe in persistence) will inflate prices in "good" states (i.e. states where prices rise) and deflate prices in "bad" states (i.e. states where prices fall). Thus, introducing belief heterogeneity in our model does not uniformly increase prices and may result in greater dispersion (i.e. higher volatility) of stock prices.

Lastly, we consider the uniqueness properties of the persistence representation. In general, beliefs in a population are not uniquely specified given equilibrium prices and strategies. However, by varying the payoffs of the stock and observing the resulting equilibrium behavior, we can completely identify the distribution of beliefs in a population.[51] We show that without loss of generality, beliefs in a population has a persistence representation if and only if any equilibrium is ordered by the same disposition ranking over agents. In other words, for any other representation, there exists some stock where the disposition ranking is reversed in some states. One agent exhibits a greater disposition effect than another agent under one equilibrium but not in another. Thus, the persistence representation is the only distribution of beliefs that permits a consistent disposition ranking of agents in all equilibria. This is a complete characterization of the persistence representation. It also allows us to equate observable equilibrium prices and trading strategies with the unobservable beliefs of agents in a population.

This chapter is related to a long literature on the disposition effect. Shefrin and Statman [76] first used the term the "disposition effect" to describe this behavior. Odean [66] provided the first comprehensive study of the disposition effect and ruled out various explanations including portfolio re-balancing, trading costs and tax considerations. Other studies include Grinblatt and Keloharju [45] in Finland and Feng and Seasholes [32] in China. Weber and Camerer [81] demonstrated that the effect is robust even in experimental studies where subjects behave in a manner inconsistent with Bayesian updating. Barberis and Xiong [5] illustrated that under certain parametric assumptions, the prospect theory of Kahneman and Tversky [51] could explain the disposition effect although under other assumptions the theory predicts the opposite effect. Other preference-based explanations include emotional regret by Strahilevitz, Odean and Barber [77] and realization utility by Barberis and Xiong [6]. In contrast, our belief-based model can be viewed as the "dual" approach to addressing the disposition effect, analogous to how the dual theory of Yaari [82] addresses violations of

[51] This approach is similar in spirit to that of Savage [73] where one varies the state-contingent payoffs (i.e. "acts") to uniquely identify beliefs under individual decision-making.

expected utility under individual decision-making.

This chapter also fits in the large literature on heterogeneous beliefs. Harrison and Kreps [48] introduce belief heterogeneity to obtain speculative pricing. Morris [62] relates belief heterogeneity with short-term IPO overpricing while Scheinkman and Xiong [74] study heterogeneous beliefs with Gaussian learning. Eyster and Piccione [30] consider a model where agents have incomplete knowledge but are all convinced about their own "theories" about how states transition. In all these models, risk-neutrality and the absence of short-selling of the risky security imply that belief heterogeneity generates inflated prices. In contrast, our model allows for both risk-aversion and short-selling. As a result, the implications on prices are more subtle as demonstrated in our Gaussian special case.

3.2 The Persistence Representation

Let S be a Polish space and let Π be the set of all probability measures on $(S, \mathcal{F})$. Consider a simple two-period model $S \times S$. We assume that all agents share the same prior belief $p \in \Pi$ on S at time 0. However, after the realization of some $s \in S$ at time 1, agents update their beliefs differently and may have various posterior beliefs about the realization of $s \in S$ at time 2. For example, S could represent earnings information about a company where agents differ in their beliefs about how correlated earnings are over time.

Formally, we model these conditional beliefs as Markov kernels $K : S \times \mathcal{F} \to [0,1]$ that have p as the invariant measure.[52]

Definition. The Markov kernel $K : S \times \mathcal{F} \to [0,1]$ is p-invariant iff K_s is absolutely continuous with respect to p and for all $A \in \mathcal{F}$

$$p(A) = \int_S p(ds)\, K_s(A)$$

[52] Given $p \in \Pi$ and measurable $f : S \to \mathbb{R}$, we use the notation $\int_S p(ds)\, f(s)$ to denote its integral if it exists.

If K is p-invariant, then each K_s has a density κ_s with respect to p. In what follows, almost surely (a.s.) always mean almost surely with respect to the measure p.[53] We say that a p-invariant Markov kernel exhibits *persistence* iff the conditional density of the same state occurring is greater than unity.

Definition. K exhibits *persistence* iff $\kappa_s(s) \geq 1$ a.s.

For example, if S is finite, then persistence implies that $K_s\{s\} \geq p\{s\}$ for all $s \in S$. In other words, after observing the realization of $s \in S$ at time 1, all agents increase their beliefs about $s \in S$ occurring again at time 2.

Let $\mathcal{K}$ denote the set of all p-invariant kernels that exhibit persistence. Since we are interested in beliefs that differ only in the degree of persistence they exhibit, we only consider kernels in $\mathcal{K}$. One extreme example is the constant kernel $\underline{K} \in \mathcal{K}$ such that $\underline{K}_s = p$ for all $s \in S$. Thus, $\underline{K}$ corresponds to believing that realizations of $s \in S$ in both time periods are completely independent and identically distributed (i.i.d.). This is an example of zero persistence.

Consider a population with heterogeneous beliefs about persistence. Formally, we model this as a probability measure μ on $\mathcal{K}$ with finite support. Each $K \in \mathcal{K}$ such that $\mu\{K\} > 0$ represents the belief of an agent (or group of agents with the same belief) in the population. A *persistence representation* is a linear one-dimensional parametrization of the degree of persistence in the population.

Definition. μ has a *persistence representation* iff there is some $\overline{K} \in \mathcal{K}$ such that for all $\mu\{K\} > 0$, there is some $\lambda \in [0,1]$ where a.s.

$$K = \lambda \overline{K} + (1-\lambda)\underline{K}$$

Under a persistence representation, each agent's belief $K \in \mathcal{K}$ is characterized by a parameter $\lambda \in [0,1]$ which specifies the level of persistence of S over time. For example, if $\lambda =$

[53] This is without loss of generality since we only consider measures that are absolutely continuous with respect to p.

0, then $K = \underline{K}$ is the constant kernel and there is no persistence. Note that μ is completely characterized by a scalar distribution on $[0,1]$ representing persistence. Hence, beliefs in the population are heterogeneous only along this dimension measuring the persistence level of realizations of S. Note that the model is completely agnostic as to what beliefs should be, that is, the true distribution of states are irrelevant. Moreover, there is no asymmetric information. For example, an agent with belief $\underline{K}$ also observes realizations of S in time 1 but believes they are irrelevant and chooses to ignore them.

We end this section with two examples of persistence representations.

Example 3.1. Let $S = \{s_1, s_2, s_3\}$, $p = \left[\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right]$ and $\overline{K}$ be the identity matrix. Note that $p\overline{K} = p$ so $\overline{K}$ is p-invariant. Also, for all $s \in S$,

$$\frac{\overline{K}_s\{s\}}{p\{s\}} = 3 \geq 1$$

so $\overline{K} \in \mathcal{K}$. In fact, $\overline{K}$ is the identity kernel representing full persistence. Let $K := \frac{1}{2}\underline{K} + \frac{1}{2}\overline{K}$ and

$$\mu := \frac{1}{3}\delta_{\underline{K}} + \frac{1}{3}\delta_K + \frac{1}{3}\delta_{\overline{K}}$$

Thus, μ is a persistence representation with equal masses on three beliefs of increasing persistence: $\underline{K}$, K and $\overline{K}$.

Example 3.2 (Gaussian Case)**.** Let $S = \mathbb{R}$ and suppose that an agent believes that the joint distribution on $S \times S$ is Gaussian (or normal). Let p be a Gaussian distribution with mean m and variance σ^2. If we let τ be the correlation coefficient, then for $s \in S$, K_s is Gaussian with mean $m(1-\tau) + \tau s$ and variance $(1-\tau^2)\sigma^2$. Thus,

$$p(ds')\,\kappa_s(s') = K_s(ds') = ds'\frac{1}{\sigma\sqrt{2\pi(1-\tau^2)}}e^{-\frac{(s'-m(1-\tau)-\tau s)^2}{2(1-\tau^2)\sigma^2}}$$

Note that

$$p(ds') = ds'\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(s'-m)^2}{2\sigma^2}}$$

Hence, for $s = s'$, we have

$$\kappa_s(s) = \frac{1}{\sqrt{1-\tau^2}} e^{\left(\frac{s-m}{\sigma}\right)^2 \frac{\tau}{1+\tau}} \geq 1$$

iff $\tau \geq 0$. Thus, $K \in \mathcal{K}$. Let $\overline{K} = K$ and $\mu := \frac{1}{2}\delta_{\underline{K}} + \frac{1}{2}\delta_{\bar{K}}$. This represents a population where half of the agents believe that realizations of S over time are completely uncorrelated while the other half of agents believe that realizations of S over time are correlated with correlation $\tau \geq 0$.

3.3 Trading Equilibrium

We now consider the trading equilibrium for a population with beliefs distributed according to μ. Let $I \subset \mathbb{N}$ be finite and for each $i \in I$, let $K^i \in \mathcal{K}$ be such that $\mu\{K^i\} > 0$. Thus, each $i \in I$ represents an agent (or group of agents) with belief $K^i \in \mathcal{K}$ about the realization of S in both time periods.

Let Z_0 denote the set of all bounded and measurable $z : S \to \mathbb{R}_+$. We interpret each $z \in Z_0$ as a security that gives payoff $z(s)$ if $s \in S$ is realized. For example, $1 \in Z_0$ and we call this risk-free security that pays one unit in every $s \in S$ a "bond". Let Z denote the set of all securities that have non-zero variance.[54] We call any $z \in Z$ a "stock" and note that $1 \notin Z$.

Fix some stock $z \in Z$. Consider a trading equilibrium where at time 0, agents trade claims to both the stock and bond. We let $s_0 \notin S$ represent the initial state at time 0 and $S_0 := \{s_0\} \cup S$. The price density be given by a measurable function $\psi : S_0 \to \mathbb{R}^2$. For $s \in S_0$, we interpret $\psi_0(s)$ and $\psi_1(s)$ as the price densities for the bond and stock respectively. Note that $\psi(s_0)$ is the price vector for both the stock and the bond at time 0.

Let Θ denote the set of all measurable functions $\theta : S_0 \to \mathbb{R}^2$. Each $\theta \in \Theta$ represents a

[54] That is, the variance of z with respect to the measure p. Note that this is equivalent to requiring that z is not constant a.s..

trading strategy. For $s \in S_0$, we interpret $\theta_0(s)$ and $\theta_1(s)$ as the holdings for the bond and stock respectively. As before, $\theta(s_0)$ represents the initial holdings for both the stock and the bond at time 0.

Given a price ψ, the pricing functional $\Psi : \Theta \to \mathbb{R}$ is given by

$$\Psi(\theta) := \psi(0) \cdot \theta(0) + \int_S p(ds)\, \psi(s) \cdot \theta(s)$$

for all $\theta \in \Theta$. Hence, $\Psi(\theta)$ specifies the total price for executing the trading strategy $\theta \in \Theta$. We assume that agents have constant endowment 1 in all states. Thus, the budget set given price ψ is

$$B(\psi) := \{\theta \in \Theta \mid \Psi(\theta) \leq \Psi(1)\}$$

Let u be a CARA utility index with constant risk aversion $\rho > 0$ and let $\delta \in (0,1)$ denote the discount rate.[55] We let $\mathbf{z} := (1, z)$ denote the asset vector. The utility of an agent with belief $K \in \mathcal{K}$ is given by the function $U_K : \Theta \to \mathbb{R}$ where

$$U_K(\theta) := \int_S p(ds)\, u(\theta(s_0) \cdot \mathbf{z}(s)) + \delta \int_S p(ds) \int_S K_s(ds')\, u(\theta(s) \cdot \mathbf{z}(s'))$$

Thus, an agent with belief $K \in \mathcal{K}$ and chooses trading strategy $\theta \in \Theta$ obtains utility $U_K(\theta)$. We say $\theta \in B(\psi)$ is *optimal for* $K \in \mathcal{K}$ iff $U_K(\theta) \geq U_K(\theta')$ for all $\theta' \in B(\psi)$. For $i \in I$, we let $\theta^i \in B(\psi)$ be optimal for $K^i \in \mathcal{K}$.

An *allocation* is a probability ν on Θ. We let (ψ, ν) denote the price and allocation pair. We now define an equilibrium given a stock z and a distribution of beliefs μ as follows.

Definition. (ψ, ν) is an *equilibrium* for (z, μ) iff $\nu\{\theta^i\} = \mu\{K^i\}$ for all $i \in I$ and a.s.

$$\sum_{i \in I} \nu\left\{\theta^i\right\} \theta^i(s) = \mathbf{1}$$

Thus, in an equilibrium, all agents elect their optimal strategies and all claim markets clear. Note that since μ has finite support, ν must also have finite support on Θ.

Given a price ψ, we can consider the normalized price density $\bar{\psi} : S_0 \to \mathbb{R}$ such that for

[55] The CARA utility index is given by $u(x) = -e^{-\rho x}$ for some $\rho > 0$.

all $s \in S_0$

$$\bar{\psi}(s) := \frac{\psi_1(s)}{\psi_0(s)}$$

The normalized price density gives the relative price of the stock to the bond in every realization of $s \in S$.[56] We interpret this as the interest-adjusted price of the stock.

We now address the behavioral characteristics of an equilibrium. Let $\succeq$ be a complete binary relation on I.

Definition. An equilibrium (ψ, ν) is *ordered by* $\succeq$ iff a.s. the following are equivalent:

(1) $i \succeq j$

(2) $\bar{\psi}(s_0) \geq \bar{\psi}(s)$ iff $\theta_1^i(s) \geq \theta_1^j(s)$

We say an equilibrium is *ordered* iff it is ordered by some $\succeq$. To illustrate this definition, consider two agents i and j such that $i \succeq j$. Now, in all states where the normalized price decreases (increases), agent i has greater (less) stock holdings than agent j. In other words, agent i exhibits a *greater disposition effect* than agent j. If an equilibrium is ordered, then this ranking on I is complete. In other words, all agents can be ranked according to the degree of disposition effect that they exhibit. Note that this is a behavioral characterization of the equilibrium that is completely observable.

Theorem 3.1 below asserts that every μ with a persistence representation has an ordered equilibrium.

Theorem 3.1. *If μ has a persistence representation, then any equilibrium of (z, μ) is ordered by some $\succeq$. Moreover, $\lambda^i \leq \lambda^j$ iff $i \succeq j$ for all $\{i, j\} \subset I$.*

Proof. See Appendix. □

In an equilibrium where μ has a persistence representation, agents who exhibit the greatest disposition effect are exactly those who have beliefs that exhibit the least persistence. This is because at time 0, all agents share the same prior on S so they all hold the same

[56] Lemma 3A.1 in the Appendix ensures that we can always define $\bar{\psi}$ without loss of generality.

amount of the stock. In states where the price of the stock increases (decreases), the agent with the least persistent belief holds the most (least) amount of stock in equilibrium. Market clearing ensures that this agent exhibits the disposition effect.

On the other hand, agents with beliefs that have the greatest persistence exhibit the opposite of the disposition effect. They increase stock holdings when prices rise and decrease holdings when prices fall. Thus, they trade based on stock price momentum. Note that this is exactly the house-money effect if we consider the limit case where prices are constant.[57] Agents who believe in any state persistence increase holdings of the stock after "good" realizations of $s \in S$ and decrease their holdings after "bad" realizations.

3.4 Special Case: Gaussian Beliefs

In this section, we consider a special case where beliefs are Gaussian. Let $S = \mathbb{R}$ and assume that p is Gaussian with mean m and variance $\sigma^2 > 0$. Recall from Example 3.2 that if we let $\tau \in (-1, 1)$ be the correlation coefficient between the two periods, then for every $s \in S$, K_s is Gaussian with mean $m(1-\tau) + \tau s$ and variance $(1-\tau^2)\sigma^2$.

Definition. $K \in \mathcal{K}$ is *Gaussian* iff there is some $\tau \geq 0$ such that K_s is a Gaussian distribution with mean $m(1-\tau) + \tau s$ and variance $(1-\tau^2)\sigma^2$.

Note that the constant kernel $\underline{K} \in \mathcal{K}$ is Gaussian with $\tau = 0$. Suppose that the measure μ only puts strictly positive mass on the constant kernel $\underline{K}$ and some other Gaussian kernel $K \in \mathcal{K}$. We call such a μ *simple Gaussian*.

Definition. μ is *simple Gaussian* iff there is some Gaussian $K \in \mathcal{K}$ and $\alpha \in [0, 1]$ such that

$$\mu = (1-\alpha)\,\delta_{\underline{K}} + \alpha\delta_K$$

Thus, a population with a simple Gaussian μ consists of agents (α proportion) who believe in that S is correlated with coefficient τ while the rest ($1-\alpha$ proportion) believe that

[57] For example, by assuming that $\mu\{\underline{K}\} \to 1$.

there is no correlation. Note that μ is trivially a persistence representation. Moreover, since mixtures of Gaussian distributions are in general not Gaussian, any μ that has a persistence representation and only puts weight on Gaussian beliefs must be simple Gaussian.

Now, suppose the stock is given by $z(s) = s$. In other words, payoffs are increasing in $s \in S$. The Proposition below characterizes the equilibrium prices and strategies for the stock.

Proposition 3.1. *Let μ be simple Gaussian and θ be the equilibrium strategy for $K \in \mathcal{K}$. Then for all $s \in S$,*

$$\bar{\psi}(s) = \frac{\alpha\tau(s-(1-\tau)m) + (1-\tau^2)(m-\rho\sigma^2)}{1-(1-\alpha)\tau^2}$$
$$\theta_1(s) = \frac{\rho\sigma^2 + (s-m)(1-\alpha)\tau}{\rho(1-(1-\alpha)\tau^2)\sigma^2}$$

Proof. See Appendix. □

This immediately implies the following corollary below.

Corollary 3.1. *Let μ be simple Gaussian and θ be the equilibrium strategy for $K \in \mathcal{K}$. Then the following are equivalent for all $s \in S$.*

(1) $s \geq m - \tau\rho\sigma^2$

(2) $\bar{\psi}(s) \geq \bar{\psi}(s_0)$

(3) $\theta_1(s) \leq \theta_1(s_0)$

Proof. See Appendix. □

Corollary 3.1 provides a very precise illustration of Theorem 3.1. A realization $s \in S$ at time 1 is considered "good" for the stock iff $s \geq m - \tau\rho\sigma^2$ and the price density increases. In this case, agents who believe that is no persistence ($\tau = 0$) decrease their holdings of the stock. Vice-versa, a realization $s \in S$ at time 1 is considered "bad" for the stock iff $s \leq m - \tau\rho\sigma^2$, the price density decreases and the agents who believe in no persistence increase their holdings. This is exactly the disposition effect in this special case with Gaussian beliefs.

Corollary 3.2 below summarizes some comparative statistics of pricing in this model.

Corollary 3.2. *If μ is simple Gaussian, then*

(1) $\bar{\psi}$ *is increasing in* α *iff* $s \geq m - \tau\rho\sigma^2$

(2) $\bar{\psi}$ *is decreasing in* ρ

Proof. Follows directly from Proposition 3.1. □

Thus, as more agents believe in persistence, the stock prices increase under "good" states and decrease under "bad" states. In other words, increasing the relative mass of agents in the population who believe in persistence results in prices that are more dispersed. This is in contrast to many other models of heterogeneous beliefs where introducing belief heterogeneity uniformly increases prices. In our model, risk aversion and the absence of any short-sale constraints result in a more subtle interaction between belief heterogeneity and prices. Note that on the other hand, increasing risk aversion uniformly lowers stock prices.

3.5 Characterization and Uniqueness

In this section, we consider the uniqueness properties of persistence representations. First, note that we can also view each kernel $K \in \mathcal{K}$ as an operator $K : \Pi \to \Pi$ where for any $q \in \Pi$, $K(q) \in \Pi$ is the measure that satisfies

$$(K(q))(A) = \int_S q(ds)\, K_s(A)$$

for all $A \in \mathcal{F}$. Since K is p-invariant, p is a fixed point of this operator. We say K is *generic* iff the operator is injective. If S is finite, then this is equivalent to requiring that all K_s are linearly independent. Thus, in this case, an agent with a generic kernel possesses conditional beliefs with enough persistence that they span the entire probability simplex. For example, the matrix corresponding to the kernel $\overline{K}$ in Example 3.1 is invertible so $\overline{K}$ is generic. We say μ is *generic* iff it puts strictly positive mass on some generic kernel.

Definition. μ is *generic* iff $\mu\{K\} > 0$ for some generic $K \in \mathcal{K}$.

If μ is generic, then there is at least some non-trivial proportion of agents in the population we exhibit enough variation and persistence in beliefs. The following is an example of a non-generic μ.

Example 3.3. Let $S = \{s_1, s_2, s_3\}$, $p = \left[\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right]$ and

$$K := \begin{bmatrix} \frac{2}{3} & \frac{1}{6} & \frac{1}{6} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ 0 & \frac{1}{2} & \frac{1}{2} \end{bmatrix}$$

Note that $pK = p$ so K is p-invariant. Also, for all $s \in S$,

$$\frac{K_s\{s\}}{p\{s\}} \in \left\{2,\ 1,\ \frac{3}{2}\right\}$$

so K is persistent and $K \in \mathcal{K}$. However, note that K is not an invertible matrix so it is not generic. This is because an agent with belief K does not update her beliefs if s_2 occurs. If we let

$$\mu := \frac{1}{2}\delta_{\underline{K}} + \frac{1}{2}\delta_K$$

then μ is not generic.

Fix some $z \in Z$ and consider the equilibrium (ψ, ν) for some (z, μ). Clearly, there are cases where we can find some other μ' such that (ψ, ν) is also the equilibrium for (z, μ'). In other words, μ is not unique given its equilibrium. However, suppose we were able to vary the security z and observe the equilibria for (z, μ) for each $z \in Z$. We say all the equilibria of μ are ordered by some $\succeq$ iff $\succeq$ ranks all agents by the degree of disposition effect they exhibit for all $z \in Z$.

Definition. The equilibria of μ are *ordered by* $\succeq$ iff a.s. for all $z \in Z$, the following are equivalent:

(1) $i \succeq j$

(2) $\bar{\psi}(0) \geq \bar{\psi}(s)$ iff $\theta_1^i(s) \geq \theta_1^j(s)$

As before, we say that equilibria of μ are *ordered* iff it is ordered by some $\succeq$. Theorem 3.2 below asserts that persistence representations are the only representations that ensure that all equilibria of μ are ordered by the same $\succeq$.

Theorem 3.2. *The equilibria of a generic μ are ordered iff μ has a persistence representation.*

Proof. See Appendix. □

Thus, if μ is generic, then observing a consistent disposition ranking $\succeq$ for all equilibria completely characterizes persistence representations. For any other representation of μ, we can find some $z \in Z$ such that the disposition ranking is violated. This allows us to completely identify the unobservable μ with observable characteristics of equilibrium prices and strategies. By varying the stock payoffs $z \in Z$, we can uniquely pin down the distribution of beliefs in the population.

3.6 Summary

We introduce a model where beliefs are heterogeneous along a single dimension measuring persistence. In equilibrium, agents can all be ordered by the degree of disposition effect they exhibit. In particular, those who hold the most persistent beliefs exhibit the disposition effect while those who hold the least persistent beliefs engage in momentum trading (a form of the house-money effect). Although we only consider a simple two-period trading model, our results could be generalized to a dynamic infinite-period setup with Markov trading strategies.

Appendix 3A

In this appendix, we prove the main results for the model.

Lemma (3A.1)**.** *If* (ψ, ν) *is an equilibrium for* (z, μ)*, then* $\psi_0(s_0) > 0$ *and* $\psi_0(s) > 0$ a.s.

Proof. Consider $K \in \mathcal{K}$ such that $\mu\{K\} > 0$ and let $\theta \in \Theta$ be optimal for K. First, suppose $\psi_0(s_0) \leq 0$ and consider $\hat{\theta} \in \Theta$ such that $\hat{\theta}(s) = \theta(s)$ for all $s \in S$ and $\hat{\theta}(s_0) = (\theta_0(s_0) + \varepsilon,\ \theta_1(s_0))$ for some $\varepsilon > 0$. Now,

$$\Psi\left(\hat{\theta}\right) = \psi_0(s_0)\,\varepsilon + \Psi(\theta) \leq \Psi(1)$$

so $\hat{\theta} \in B(\psi)$. Since u is CARA,

$$\begin{aligned} U_K\left(\hat{\theta}\right) - U_K(\theta) &= \int_S p(ds)\left[u(\theta(s_0)\cdot\mathbf{z}(s) + \varepsilon) - u(\theta(s_0)\cdot\mathbf{z}(s))\right] \\ &= \left(e^{-\rho\varepsilon} - 1\right)\int_S p(ds)\,u(\theta(s_0)\cdot\mathbf{z}(s)) \end{aligned}$$

Since z is bounded, $U_K\left(\hat{\theta}\right) > U_K(\theta)$ contradicting the fact that θ is optimal. Hence, $\psi_0(s_0) > 0$.

Define

$$E := \{s \in S \mid \psi_0(s) \leq 0\}$$

and suppose $p(E) > 0$. Let $\hat{\theta} \in B(\psi)$ be such that $\hat{\theta}(s) = \theta(s)$ if $s \in \{s_0\} \cup (S \backslash E)$ and $\hat{\theta}(s) = (\theta_0(s) + \varepsilon,\ \theta_1(s))$ if $s \in E$ for some $\varepsilon > 0$. Now,

$$\Psi\left(\hat{\theta}\right) = \Psi(\theta) + \varepsilon\int_E p(ds)\,\psi_0(s) \leq \Psi(1)$$

so $\hat{\theta} \in B(\psi)$. Again, as u is CARA,

$$U_K\left(\hat{\theta}\right) - U_K(\theta) = \delta\left(e^{-\rho\varepsilon} - 1\right)\int_E p(ds)\int_S K_s(ds')\,u(\theta(s)\cdot\mathbf{z}(s'))$$

Let

$$\zeta(s) := \int_S K_s(ds')\,u(\theta(s)\cdot\mathbf{z}(s'))$$

and note that since z is bounded, $\zeta(s) < 0$ for all $s \in S$. For $\eta > 0$, let

$$E_\eta := \{s \in E \mid \zeta(s) < -\eta\}$$

Now, for all $\eta > 0$,

$$\int_E p(ds)\,\zeta(s) \leq -p(E_\eta)\,\eta$$

Suppose $p(E_\eta) = 0$ for all $\eta > 0$. As $\eta \to 0$, $E_\eta \nearrow E$ so $p(E_\eta) \to p(E) > 0$ a contradiction. Thus, $\exists \eta > 0$ such that $p(E_\eta) > 0$ so $\int_E p(ds)\,\zeta(s) < 0$. Hence, $U_K(b) > U_K(a)$ again contradicting the optimality of θ. We thus have $\psi_0(s) > 0$ a.s.. □

Lemma 3A.1 ensures that we can define the normalized price density $\bar{\psi} : S_0 \to \mathbb{R}$ such that

$$\bar{\psi}(s) := \frac{\psi_1(s)}{\psi_0(s)}$$

For ease of notation, we let $\mathbb{E}^q$ denote the expectation operator with respect to the measure $q \in \Pi$. For $q = K_s^i$, we let $\mathbb{E}_s^i := \mathbb{E}^{K_s^i}$. Let Π_0 be the set of probability measures on S absolutely continuous with respect to p. Note that by definition, $K \in \mathcal{K}$ implies $K_s \in \Pi_0$ for all $s \in S$. Fix $z \in Z$ and let $\xi : \Pi_0 \times \mathbb{R} \to \mathbb{R}$ be such that

$$\xi(q, a) := \frac{\mathbb{E}^q[u(az)\,z]}{\mathbb{E}^q[u(az)]}$$

Note that since z is bounded and u is CARA, $\mathbb{E}^q[u(az)] < 0$ so ξ is well-defined.

Lemma (3A.2)**.** *Fix $z \in Z$.*

(1) *ξ is strictly decreasing in a*

(2) $\lim_{a\to\infty} \xi(q, a) = \sup_{s\in S} z(s)$ *and* $\lim_{a\to-\infty} \xi(q, a) = \inf_{s\in S} z(s)$

(3) *If $\xi(q, a) = \xi(r, a)$, then $\xi(\lambda q + (1-\lambda) r, a) = \xi(q, a)$ for all $\lambda \in [0, 1]$.*

Proof. Fix $z \in Z$. We prove the lemma in order.

(1) Since u is CARA,

$$\frac{\partial u(az)}{\partial a} = u'(az)\,z = -\rho u(az)\,z$$

As z is bounded, by Theorem 16.8 of Billingsley [9], ξ is differentiable in a and

$$\frac{\partial \xi}{\partial a} = -\rho \frac{\mathbb{E}^q \left[u\left(az\right) z^2\right]}{\mathbb{E}^q \left[u\left(az\right)\right]} + \rho \left(\frac{\mathbb{E}^q \left[u\left(az\right) z\right]}{\mathbb{E}^q \left[u\left(az\right)\right]}\right)^2$$

For $a \in \mathbb{R}$, define $q_a \in \Pi$ such that

$$q_a\left(A\right) := \frac{\mathbb{E}^q \left[\mathbf{1}_A u\left(az\right)\right]}{\mathbb{E}^q \left[u\left(az\right)\right]}$$

for all $A \in \mathcal{F}$. Now,

$$\frac{\partial \xi}{\partial a} = \rho \left(\mathbb{E}^{q_a}\left[z\right]\right)^2 - \rho \mathbb{E}^{q_a}\left[z^2\right] = -\rho \mathbb{E}^{q_a}\left[\left(z - \mathbb{E}^{q_a}\left[z\right]\right)^2\right] \leq 0$$

If the last inequality is an equality, then $z = \mathbb{E}^{q_a}\left[z\right]$ a constant q_a-a.s.. Note that since $q \in \Pi_0$, $p\left(A\right) = 0$ implies $q\left(A\right) = 0$ which implies $q_a\left(A\right) = 0$. Hence, $q_a \in \Pi_0$ for all $a \in \mathbb{R}$ so z is non-constant q_a-a.s.. Thus, ξ must be strictly decreasing in a.

(2) Since z is measurable, we can approximate z by a sequence of increasing simple functions. In other words, $z = \lim_n z_n$ where the z_n are increasing and

$$z_n = \sum_t c_t^n \mathbf{1}_{A_t^n}$$

for $A_t^n \in \mathcal{F}$. Now,

$$\frac{\mathbb{E}^q \left[u\left(az_n\right) z_n\right]}{\mathbb{E}^q \left[u\left(az_n\right)\right]} = \frac{\sum_t q\left(A_t^n\right) u\left(ac_t^n\right) c_t^n}{\sum_t q\left(A_t^n\right) u\left(ac_t^n\right)} = \frac{\sum_t q\left(A_t^n\right) e^{-\rho a c_t^n} c_t^n}{\sum_t q\left(A_t^n\right) e^{-\rho a c_t^n}}$$

Clearly, if $\bar{c}_t^n = \sup_{s \in S} z_n\left(s\right)$, then

$$\lim_{a \to \infty} \frac{\mathbb{E}^q \left[u\left(az_n\right) z_n\right]}{\mathbb{E}^q \left[u\left(az_n\right)\right]} = \bar{c}_t^n = \sup_{s \in S} z_n\left(s\right)$$

By dominated convergence (see Theorem I.4.16 of Çinlar [18]),

$$\xi\left(q, a\right) = \frac{\mathbb{E}^q \left[u\left(az\right) z\right]}{\mathbb{E}^q \left[u\left(az\right)\right]} = \lim_n \frac{\mathbb{E}^q \left[u\left(az_n\right) z_n\right]}{\mathbb{E}^q \left[u\left(az_n\right)\right]}$$

Hence,

$$\lim_{a\to\infty}\xi(q,a)=\lim_{n}\lim_{a\to\infty}\frac{\mathbb{E}^{q}\left[u(az_n)z_n\right]}{\mathbb{E}^{q}\left[u(az_n)\right]}=\lim_{n}\sup_{s\in S}z_n(s)$$
$$=\sup_{s\in S}\lim_{n}z_n(s)=\sup_{s\in S}z(s)$$

The case for $\lim_{\alpha\to-\infty}\xi(q,\alpha)=\inf_{s\in S}z(s)$ is symmetric.

(3) Suppose $\xi(q,a)=\xi(r,a)=\xi^{*}$ for $\{r,q\}\subset\Pi_0$ and $a\in\mathbb{R}$. Thus,

$$\mathbb{E}^{q}\left[u(az)(z-\xi^{*})\right]=\mathbb{E}^{r}\left[u(az)(z-\xi^{*})\right]=0$$

If we let $q_\lambda:=\lambda q+(1-\lambda)r$ for $\lambda\in[0,1]$, then

$$\mathbb{E}^{q_\lambda}\left[u(az)(z-\xi^{*})\right]=0$$

which implies

$$\xi^{*}=\frac{\mathbb{E}^{q_\lambda}\left[u(az)z\right]}{\mathbb{E}^{q_\lambda}\left[u(az)\right]}=\xi(q_\lambda,a)$$

□

Lemma (3A.3)**.** *Let $q^{j}=\lambda q^{i}+(1-\lambda)q^{k}$ for some $\lambda\in(0,1)$, and suppose*

$$\xi\left(q^{k},a^{k}\right)=\xi\left(q^{j},a^{j}\right)=\xi\left(q^{i},a^{i}\right)$$

Then $a^{k}=a^{i}$ implies $a^{k}=a^{j}=a^{i}$ and $a^{k}>a^{i}$ implies $a^{k}>a^{j}>a^{i}$.

Proof. Let

$$\xi^{*}:=\xi\left(q^{k},a^{k}\right)=\xi\left(q^{j},a^{j}\right)=\xi\left(q^{i},a^{i}\right)$$

First, suppose $a:=a^{i}=a^{k}$. Thus, from Lemma 3A.2, $\xi(q^{j},a)=\xi^{*}$ and $a=a^{j}$.

Now, assume $a^{k}>a^{i}$. Again, by Lemma 3A.2,

$$\xi\left(q^{k},a^{i}\right)>\xi^{*}>\xi\left(q^{i},a^{k}\right)$$

Suppose $a^j > a^k$ so $\xi\left(q^j, a^k\right) > \xi^*$. By continuity, we can find some $\gamma \in (0,1)$ such that

$$\xi^* = \xi\left(\gamma q^j + (1-\gamma) q^i, a^k\right) = \xi\left(q^k, a^k\right)$$

Now, if we let $\hat{\gamma} := \frac{\lambda}{\lambda+(1-\gamma)(1-\lambda)}$, then

$$\hat{\gamma}\left(\gamma q^j + (1-\gamma) q^i\right) + (1-\hat{\gamma}) q^k = \lambda q^i + (1-\lambda) q^k = q^j$$

Thus, by Lemma 3A.2, $\xi^* = \xi\left(q^j, a^k\right)$ which implies $a^j = a^k$ a contradiction. The case for $a^i > a^j$ is symmetric, so we have $a^k \geq a^j \geq a^i$. Lastly, suppose $a := a^k = \alpha^j$ so by Lemma 3A.2,

$$\xi^* = \xi\left(q^i, a^i\right) > \xi\left(q^i, a\right) = \frac{\mathbb{E}^{q^i}\left[u(az)z\right]}{\mathbb{E}^{q^i}\left[u(az)\right]}$$

Thus,

$$\begin{aligned}\mathbb{E}^{q^i}\left[u(az)(z-\xi^*)\right] > 0 &= \mathbb{E}^{q^k}\left[u(az)(z-\xi^*)\right] \\ &= \mathbb{E}^{q^j}\left[u(az)(z-\xi^*)\right]\end{aligned}$$

However, $q^j = \lambda q^i + (1-\lambda) q^k$ for $\lambda \in (0,1)$ yielding a contradiction. The case for $a^i = a^j$ is symmetric, so $a^k > a^j > a^i$. □

Theorem (3A.4)**.** *If μ has a persistence representation, then any equilibrium of (z,μ) is ordered by some $\succeq$. Moreover, $\lambda^i \leq \lambda^j$ iff $i \succeq j$ for all $\{i,j\} \subset I$.*

Proof. Let μ have a persistence representation and (ψ,ν) be an equilibrium for (z,μ). From the optimality conditions and the fact that u is CARA, we have for all $i \in I$,

$$\begin{aligned}\bar{\psi}(s_0) = \frac{\psi_1(s_0)}{\psi_0(s_0)} &= \frac{\mathbb{E}^p\left[u'\left(\theta^i(s_0)\cdot\mathbf{z}\right)z\right]}{\mathbb{E}^p\left[u'\left(\theta^i(s_0)\cdot\mathbf{z}\right)\right]} \\ &= \frac{\mathbb{E}^p\left[u\left(\theta_1^i(s_0)z\right)z\right]}{\mathbb{E}^p\left[u\left(\theta_1^i(s_0)z\right)\right]} = \xi\left(p, \theta_1^i(s_0)\right)\end{aligned}$$

Thus, by Lemma 3A.2, $\theta_1^i(s_0)$ is the same for all $i \in I$ so by market clearing, $\theta_1^i(s_0) = 1$ for all $i \in I$. Note that by similar reasoning, we have a.s.

$$\bar{\psi}(s) = \frac{\psi_1(s)}{\psi_0(s)} = \frac{\mathbb{E}_s^i\left[u\left(\theta_1^i(s)\, z\right) z\right]}{\mathbb{E}_s^i\left[u\left(\theta_1^i(s)\, z\right)\right]} = \xi\left(K_s^i, \theta_1^i(s)\right)$$

Define $i \succeq j$ iff $\lambda^i \leq \lambda^j$ iff $i \leq j$ for all $\{i, j\} \subset I$. We show that (z, μ) is ordered by $\succeq$. Let $\{1, n\} \subset I$ be such that $1 \leq i \leq n$ for all $i \in I$. Let

$$E := \left\{ s \in S \;\middle|\; \bar{\psi}(s_0) \geq \bar{\psi}(s) \text{ and } \theta_1^1(s) < 1 \right\}$$

and suppose $p(E) > 0$. For $s \in E$, the Lemma 3A.2 ensures that we can always find some $a_s \in \mathbb{R}$ such that $\xi(p, a_s) = \bar{\psi}(s)$. Note that we can assume $\bar{\psi}(s) = \xi\left(K_s^i, \theta_1^i(s)\right)$ for all $i \in I$ without loss of generality, so

$$\xi\left(p, \theta_1^1(s)\right) > \xi(p, 1) = \bar{\psi}(0) \geq \bar{\psi}(s) = \xi(p, a_s) = \xi\left(K_s^1, \theta_1^1(s)\right) = \xi\left(K_s^n, \theta_1^n(s)\right)$$

By Lemma 3A.2 and 3A.3, we have $a_s \geq 1 > \theta_1^1(s)$ so $a_s > \theta_1^1(s) > \theta_1^n(s)$ as $K_s^1 = \lambda^1 p + (1 - \lambda^1) K_s^n$. Since this is true for all $s \in E$, we have $1 > \theta_1^i(s)$ for all $i \in I$ on a set of strictly positive measure, contradicting market clearing. Thus, $\bar{\psi}(s_0) \geq \bar{\psi}(s)$ implies $\theta_1^1(s) \geq 1$ which implies $\theta_1^i(s) \geq \theta_1^j(s)$ for $j \geq i$ a.s.. By symmetric argument, we have $\bar{\psi}(s_0) \leq \bar{\psi}(s)$ implies $\theta_1^i(s) \leq \theta_1^j(s)$ for $j \geq i$ a.s.. Thus, $\bar{\psi}(s_0) \geq \bar{\psi}(s)$ iff $\theta_1^i(s) \geq \theta_1^j(s)$ for $j \geq i$ a.s. so (z, μ) is ordered by $\succeq$. □

Lemma (3A.5)**.** *Suppose $L_s = \lambda p + (1 - \lambda) K_s$ a.s. for $\{K, L\} \subset \mathcal{K}$ and $\lambda \in (0, 1)$. Then K is generic iff L is generic.*

Proof. First, suppose K is generic. Let $\{r, q\} \subset \Pi$ be such that

$$\int_S q(ds)\, L_s(A) = \int_S r(ds)\, L_s(A)$$

for all $A \in \mathcal{F}$. Now,

$$\int_S q(ds)\left(\lambda p(A) + (1 - \lambda) K_s(A)\right) = \int_S r(ds)\left(\lambda p(A) + (1 - \lambda) K_s(A)\right)$$
$$\int_S q(ds)\, K_s(A) = \int_S r(ds)\, K_s(A)$$

implying $q = r$ so L is generic. If L is generic, then let $\{r, q\} \subset \Pi$ be such that

$$\int_S q\,(ds)\, K_s\,(A) = \int_S r\,(ds)\, K_s\,(A)$$
$$\int_S q\,(ds)\,(\lambda p\,(A) + (1-\lambda)\, K_s\,(A)) = \int_S r\,(ds)\,(\lambda p\,(A) + (1-\lambda)\, K_s\,(A))$$

for all $A \in \mathcal{F}$. Thus, $r = q$ and K is generic. □

For $\{q, r\} \subset \mathbb{R}^d$ where $d \in \mathbb{N}$, define

$$[q, r] := \left\{ q\lambda + (1-\lambda)\, r \in \mathbb{R}^d \;\middle|\; \lambda \in [0, 1] \right\}$$

Similarly, for $\{q, r\} \subset \Pi_0$, define

$$[q, r] := \{ q\lambda + (1-\lambda)\, r \in \Pi_0 \mid \lambda \in [0, 1] \}$$

Lemma (3A.6)**.** *If μ has a persistence representation, then its equilibria are ordered.*

Proof. Suppose μ has a persistence representation, and for any $z \in Z$, define $E_z \subset S$ such that $s \in E_z$ iff $i \succeq j$ is equivalent to $\bar{\psi}\,(s_0) \geq \bar{\psi}\,(s)$ iff $\theta_1^i\,(s) \geq \theta_1^j\,(s)$. By Theorem 3A.4, $p\,(E_z) = 1$. Now, let $\bar{S} \subset S$ be the set such that $s \in \bar{S}$ iff for all $z \in Z$, $i \succeq j$ is equivalent to $\bar{\psi}\,(s_0) \geq \bar{\psi}\,(s)$ iff $\theta_1^i\,(s) \geq \theta_1^j\,(s)$. Note that

$$\bar{S} = \bigcap_{z \in Z} E_z$$

Let $Z^* \subset Z$ be some dense countable subset of Z so

$$\bigcap_{z \in Z} E_z \subset \bigcap_{z \in Z^*} E_z$$

Now, let $s \in \bigcap_{z \in Z^*} E_z$. Thus, for all $z \in Z^*$, $i \succeq j$ is equivalent to $\bar{\psi}\,(s_0) \geq \bar{\psi}\,(s)$ iff $\theta_1^i\,(s) \geq \theta_1^j\,(s)$. By the continuity of prices and holdings, we have $i \succeq j$ is equivalent to $\bar{\psi}\,(s_0) \geq \bar{\psi}\,(s)$ iff $\theta_1^i\,(s) \geq \theta_1^j\,(s)$ for all $z \in Z$. Thus, $s \in \bar{S}$ so $\bar{S}$ is measurable. Hence,

$$p\,(\bar{S}) = p\left(\bigcap_{z \in Z^*} E_z\right) = 1$$

□

Theorem (3A.7)**.** *The equilibria of a generic μ are ordered iff μ has a persistence representation.*

Proof. Note that necessity follows from Lemma 3A.6, so we prove sufficiency. Let μ be generic with equilibria ordered by $\succeq$. Without loss of generality, let $1 \succeq i \succ i+1 \succeq n$ for all $i \in I$. Define $\bar{S} \subset S$ to be the a.s. set such that $s \in \bar{S}$ iff for all $z \in Z$, $i \succeq j$ is equivalent to $\bar{\psi}(s_0) \geq \bar{\psi}(s)$ iff $\theta_1^i(s) \geq \theta_1^j(s)$.

Fix $s \in \bar{S}$ and first consider $i \succ j \succ k$ for $\{i,j,k\} \subset I$. Let $\{A_1, \dots, A_d\}$ be a measurable partition of S for some $d \in \mathbb{N}$ and $c \in \mathbb{R}_+^d$. Let $z = \sum_t \mathbf{1}_{A_t} c_t$ for $t \in \{1, \dots, d\}$. Note that $\theta_1^i(s) \geq \theta_1^j(s) \geq \theta_1^k(s)$ or $\theta_1^i(s) \leq \theta_1^j(s) \leq \theta_1^k(s)$. Assume the former without loss of generality. Let $q^i := K_s^i$ for all $i \in I$ and note that also without loss of generality

$$\bar{\psi}(s) = \xi\left(q^i, \theta_1^i(s)\right) = \xi\left(q^j, \theta_1^j(s)\right) = \xi\left(q^k, \theta_1^k(s)\right)$$

If we let $a_c := \theta_1^j(s)$ and $\xi_c := \bar{\psi}(s)$, then by the lemma above,

$$\xi\left(q^i, a_c\right) \leq \xi_c = \xi\left(q^j, a_c\right) \leq \xi\left(q^k, a_c\right)$$

Thus,

$$\mathbb{E}^{q^i}\left[u\left(a_c z\right)\left(z - \xi_c\right)\right] \geq 0 = \mathbb{E}^{q^j}\left[u\left(a_c z\right)\left(z - \xi_c\right)\right] \geq \mathbb{E}^{q^k}\left[u\left(a_c z\right)\left(z - \xi_c\right)\right]$$

If we let $v_c(t) = u(a_c c_t)(c_t - \xi_c)$ for all $t \in \{1, \dots, m\}$, then $q^i \cdot v_c \geq q^j \cdot v_c \geq q^k \cdot v_c$ where $\{q^i, v_c\} \subset \mathbb{R}^m$.

For $c = \mathbf{1}_t$,

$$q^i \cdot v_c = q^i(t)\, u(a_c)(1 - \xi_z) + \left(1 - q^i(t)\right) u(0)(-\xi_z)$$

If $q^i(t) = q^k(t)$, then $q^j(t) = q^i(t)$. Otherwise, we can find a $t' \in \{1, \dots, d\}$ such that $q^i(t) > q^k(t)$ and $q^i(t') < q^k(t')$ without loss of generality. By continuity, there is some $\hat{c} = \beta \mathbf{1}_t + (1-\beta)\mathbf{1}_{t'}$ such that $q^i \cdot v_{\hat{c}} = q^j \cdot v_{\hat{c}} = q^k \cdot v_{\hat{c}}$. Note that we can always find $m-2$ such $\hat{c}$ where $v_{\hat{c}}$ are linearly independent. Since $\sum_t q^i(t) = 1$, we must have $q^j \in \left[q^i, q^k\right]$.

Since this is true for all $\{i, j, k\} \subset I$, we must have $q^i \in [q^1, q^n]$ for all $i \in I$. If $p \notin [q^1, q^n]$, we can find some c such that $\xi_c \geq \bar{\psi}(s_0)$ and $\mathbb{E}^{q^i}[u(az)(z - \xi_c)] \geq \mathbb{E}^p[u(az)(z - \xi_c)]$ for all $i \in I$ where $\xi(p, a) = \xi_c$. Hence,

$$\xi\left(q^i, a\right) \leq \xi(p, a) = \xi\left(q^i, \theta_1^i(s)\right) = \xi_c \geq \bar{\psi}(s_0) = \xi(p, 1)$$

Thus, $\theta_1^i(s) \leq a \leq 1$ contradicting market clearing. Hence, $p \in [q^1, q^n]$.

Suppose $q^j \notin \left[q^i, q^k\right]$. By a standard Separating Hyperplane Theorem (see Theorem 5.61 of Aliprantis and Border [2]), we can find some measurable $\zeta : S \to \mathbb{R}$ such that

$$\mathbb{E}^{q^j}[\zeta] \notin \left[\mathbb{E}^{q^i}[\zeta],\ \mathbb{E}^{q^k}[\zeta]\right]$$

Since ζ is measurable, it is the limit of a sequence of increasing simple functions. Hence, $\zeta = \lim_l \zeta_l$ where

$$\zeta_l = \sum_t b_t^l \mathbf{1}_{A_t^l}$$

Now, for each ζ_l, we have

$$\mathbb{E}^{q^j}[\zeta_l] = \sum_t b_t^l q^j\left(A_t^l\right) = \lambda \sum_t b_t^l q^i\left(A_t^l\right) + (1 - \lambda) \sum_t b_t^l q^k\left(A_t^l\right)$$

so $\mathbb{E}^{q^j}[\zeta_l] \in \left[\mathbb{E}^{q^i}[\zeta_l],\ \mathbb{E}^{q^k}[\zeta_l]\right]$ for all ζ_l. By monotone convergence, we must have $\mathbb{E}^{q^j}[\zeta] \in \left[\mathbb{E}^{q^i}[\zeta],\ \mathbb{E}^{q^k}[\zeta]\right]$ a contradiction. By similar argument, $p \in [q^1, q^n]$.

Thus, we have $p \in [K_s^1, K_s^n]$ a.s.. Hence, we have a.s.

$$\begin{aligned} p(ds') &= \lambda_s K_s^1(ds') + \lambda_s K_s^n(ds') \\ &= p(ds')\,\kappa_s^1(s')\,\lambda_s + p(ds')\,\kappa_s^n(s')\,(1 - \lambda_s) \end{aligned}$$

Hence, we have $\kappa_s^1(s)\lambda_s + \kappa_s^n(s)(1 - \lambda_s) = 1$ a.s.. Since $\{K^1, K^n\} \subset \mathcal{K}$ we have $\kappa_s^1(s) \geq 1$ and $\kappa_s^n(s) \geq 1$ so $\lambda_s \in \{0, 1\}$ a.s.

Now, consider $K^i \in [p, K^n]$ so for all $A \in \mathcal{F}$,

$$\begin{aligned} p(A) &= \int_S p(ds) K_s^i(A) = \int_S p(ds) \left(\lambda_s^i p(A) + \left(1 - \lambda_s^i\right) K_s^n(A)\right) \\ &= p(A) \int_S p(ds) \lambda_s^i + p(A) - \int_S p(ds) K_s^n(A) \lambda_s^i \end{aligned}$$

Note that if $\lambda_s^i = 0$ a.s. then $K_s^i = p$ a.s.. Hence, suppose the $\lambda_s^i > 0$ on some set of strictly positive p-measure. We can now define $p^i \in \Pi$ such that

$$p^i(A) := \int_A \frac{p(ds) \lambda_s^i}{\int_S p(ds) \lambda_s^i}$$

so

$$p(A) = \int_S p^i(ds) K_s^n(A) = \int_S p(ds) K_s^n(A)$$

Since μ is generic, by Lemma 3A.5, K^n is generic. Hence, $p = p^i$ so by the Radon-Nikodym Theorem (Theorem I.5.11 of Çinlar [18]), we have

$$1 = \frac{dp^i}{dp} = \frac{\lambda_s^i}{\int_S p(ds) \lambda_s^i}$$

a.s. so $\lambda_s^i = \lambda^i$ a.s.. The case for $K^i \in [p, K^1]$ is symmetric, so we have $\lambda_s^i = \lambda^i$ for all $i \in I$. Thus, μ has a persistence representation. □

Appendix 3B

In this appendix, we prove the results for Gaussian model. Let $z \in Z$ be such that $z(s) = s$.

Lemma (3B.1)**.** *Let $q \in \Pi$ be Gaussian distributed with mean m and variance $\sigma^2 > 0$. Then*

$$\begin{aligned} \mathbb{E}^q[u(az)] &= u\left(ma - \rho\frac{\sigma^2}{2}a^2\right) \\ \mathbb{E}^q[u(az)z] &= u\left(ma - \rho\frac{\sigma^2}{2}a^2\right)\left(m - a\rho\sigma^2\right) \end{aligned}$$

Proof. A simple computation yields

$$\mathbb{E}^q\left[u\left(az\right)\right] = -e^{-\rho a\left(m - a\rho\frac{\sigma^2}{2}\right)}$$
$$\mathbb{E}^q\left[u\left(az\right)z\right] = -e^{-\rho a\left(m - a\rho\frac{\sigma^2}{2}\right)}\left(m - a\rho\sigma^2\right)$$

The result then follows from the definition of u. □

Proposition (3B.2)**.** *Let μ be simple Gaussian and θ be the equilibrium strategy for $K \in \mathcal{K}$. Then for all $s \in S$,*

$$\bar{\psi}\left(s\right) = \frac{\alpha\tau\left(s - \left(1-\tau\right)m\right) + \left(1-\tau^2\right)\left(m - \rho\sigma^2\right)}{1 - \left(1-\alpha\right)\tau^2}$$
$$\theta_1\left(s\right) = \frac{\rho\sigma^2 + \left(s - m\right)\left(1-\alpha\right)\tau}{\rho\left(1 - \left(1-\alpha\right)\tau^2\right)\sigma^2}$$

Proof. Let $K \in \mathcal{K}$ be Gaussian and θ be optimal for K. Fix $s \in S$ and let $a = \theta_1\left(s\right)$. Now, by Lemma 3B.1,

$$\bar{\psi}\left(s\right) = \xi\left(K_s, a\right) = \frac{\mathbb{E}^{K_s}\left[u\left(az\right)z\right]}{\mathbb{E}^{K_s}\left[u\left(az\right)\right]}$$
$$= m\left(1-\tau\right) + \tau s - a\rho\left(1-\tau^2\right)\sigma^2$$

Thus, we have

$$\theta_1\left(s\right) = a = \frac{m\left(1-\tau\right) + \tau s - \bar{\psi}\left(s\right)}{\rho\left(1-\tau^2\right)\sigma^2}$$

Note that if $\tau = 0$, then $\theta_1\left(s\right) = \frac{m - \bar{\psi}(s)}{\rho\sigma^2}$. Since $\theta_1\left(s_0\right) = 1$ by market clearing, we have

$$\bar{\psi}\left(s_0\right) = m - \rho\sigma^2$$

Also by market clearing,

$$1 = \alpha\theta_1^{\tau}\left(s\right) + \left(1-\alpha\right)\theta_1^0\left(s\right) = \alpha\frac{m\left(1-\tau\right) + \tau s - \bar{\psi}\left(s\right)}{\rho\left(1-\tau^2\right)\sigma^2} + \left(1-\alpha\right)\frac{m - \bar{\psi}\left(s\right)}{\rho\sigma^2}$$

Hence

$$\bar{\psi}\left(s\right) = \frac{\alpha\tau\left(s - \left(1-\tau\right)m\right) + \left(1-\tau^2\right)\left(m - \rho\sigma^2\right)}{1 - \left(1-\alpha\right)\tau^2}$$

Substituting the formula for $\bar{\psi}$ yields

$$\theta_1(s) = \frac{\rho\sigma^2 + (s-m)(1-\alpha)\tau}{\rho(1-(1-\alpha)\tau^2)\sigma^2}$$

□

Corollary (3B.3)**.** *Let μ be simple Gaussian and θ be the equilibrium strategy for $K \in \mathcal{K}$. Then the following are equivalent for all $s \in S$*

(1) $s \geq m - \tau\rho\sigma^2$

(2) $\bar{\psi}(s) \geq \bar{\psi}(s_0)$

(3) $\theta_1(s) \leq \theta_1(s_0)$

Proof. From Proposition 3B.2, we have $\bar{\psi}(s) \geq \bar{\psi}(s_0)$ iff

$$\alpha\tau(s-(1-\tau)m) + \left(1-\tau^2\right)\left(m-\rho\sigma^2\right) \geq \left(1-(1-\alpha)\tau^2\right)\left(m-\rho\sigma^2\right)$$
$$s \geq m - \tau\rho\sigma^2$$

so (1) and (2) are equivalent. Since it also follows readily that $\theta_1(s) \geq 1$ iff $s \geq m - \tau\rho\sigma^2$ we have (1), (2) and (3) are all equivalent. □

References

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

[2] C. Aliprantis and K. Border. *Infinite Dimensional Analysis*. Springer, 2006.

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

[4] S. Athey and J. Levin. The value of information in monotone decision problems. Mimeo, 2001.

[5] N. Barberis and W. Xiong. What drives the disposition effect? an analysis of a long-standing preference-based explanation. *The Journal of Finance*, 64(2):751–784, 2009.

[6] N. Barberis and W. Xiong. Realization utility. *Journal of Financial Economics*, 104:251–271, 2012.

[7] G. Becker, M. DeGroot, and J. Marschak. An experimental study of some stochastic models for wagers. *Behavioral Science*, 8:199–202, 1963.

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

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

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

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

[12] 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.

[13] S. Bose, E. Ozdenoren, and A. Pape. Optimal auctions with ambiguity. *Theoretical Economics*, 1(4):411–438, 2006.

[14] R. Caballero and A. Krishnamurthy. Collective risk management in a flight to quality episode. *The Journal of Finance*, 63:2195–2230, 2008.

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

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

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

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

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

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

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

[22] R. Dhar and N. Zhu. Up close and personal: Investor sophistication and the disposition effect. *Management Science*, 52(5):726–740, 2006.

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

[24] J. Dow and S. da Costa Werlang. Uncertainty aversion, risk aversion, and the optimal choice of portfolio. *Econometrica*, 60:197–204, 1992.

[25] D. Easley and M. O'Hara. Ambiguity and nonparticipation: The role of regulation. *The Review of Financial Studies*, 22:1817–1843, 2009.

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

[27] D. Ellsberg. Risk, ambiguity, and the savage axioms. *The Quarterly Journal of Economics*, 75(4):643–669, 1961.

[28] L. Epstein and I. Kopylov. Cold feet. *Theoretical Economics*, 2:231–259, 2007.

[29] L. Epstein and M. Schneider. Ambiguity and asset markets. *Annual Review of Financial Economics*, 2:315–346, 2010.

[30] E. Eyster and M. Piccione. An approach to asset pricing under incomplete and diverse perceptions. *Econometrica*, 81(4):1483–1506, 2013.

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

[32] L. Feng and M. Seasholes. Do investor sophistication and trading experience eliminate behavioral biases in financial markets. *Review of Finance*, 9:305–351, 2005.

[33] 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.

[34] 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.

[35] P. Fishburn. *Utility Theory for Decision Making*. John Wiley and Sons, Inc., 1970.

[36] D. Fudenberg and T. Strzalecki. Recursive stochastic choice. Mimeo, 2013.

[37] T. Gajdos, T. Hayashi, J.-M. Tallon, and J.-C. Vergnaud. Attitude toward imprecise information. *Journal of Economic Theory*, 140:27–65, 2008.

[38] D. Genesove and M. Christopher. Loss aversion and seller behavior: Evidence from the housing market. *The Quarterly Journal of Economics*, 116(4):1233–1260, 2001.

[39] P. Ghirardato and M. Marinacci. Ambiguity made precise: A comparative foundation. *Journal of Economic Theory*, 102:251–289, 2002.

[40] I. Gilboa and D. Schmeidler. Maxmin expected utility with non-unique prior. *Journal of Mathematical Economics*, 18:141–153, 1989.

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

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

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

[44] S. Grant and K. Kajii. The epsilon-gini-contamination multiple priors model admits a linear-mean-standard-deviation utility representation. *Economics Letters*, 95:39–47, 2007.

[45] M. Grinblatt and M. Keloharju. What makes investors trade? *The Journal of Finance*, 56(2):589–616, 2001.

[46] F. Gul, P. Natenzon, and W. Pesendorfer. Random choice and behavioral optimization. Mimeo, 2012.

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

[48] M. Harrison and D. Kreps. Speculative investor behavior in a stock market with heterogeneous expectations. *The Quarterly Journal of Economics*, 92(2):323–336, 1978.

[49] C. Heath, S. Huddart, and M. Lang. Psychological factors and stock option exercise. *Quarterly Journal of Economics*, 114:601–627, 1991.

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

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

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

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

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

[55] I. Kopylov. Choice deferral and ambiguity aversion. *Theoretical Economics*, 4:199–225, 2009.

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

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

[58] E. Lehmann. Comparing location experiments. *The Annals of Statistics*, 16(2):521–533, 1988.

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

[60] 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.

[61] 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.

[62] S. Morris. Speculative investor behavior and learning. *The Quarterly Journal of Economics*, 111(4):1111–1133, 1996.

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

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

[65] K. Nehring. Preference for flexibility in a savage framework. *Econometrica*, 67(1):101–119, 1999.

[66] T. Odean. Are investors reluctant to realize their losses? *The Journal of Finance*, 53(5):1775–1798, 1998.

[67] N. Persico. Information acquisition in auctions. *Econometrica*, 68(1):135–148, 2000.

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

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

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

[71] 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.

[72] W. Rudin. *Real and Complex Analysis*. McGraw-Hill, 1987.

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

[74] J. Scheinkman and W. Xiong. Overconfidence and speculative bubbles. *The Journal of Political Economy*, 111(6):1183–1219, 2003.

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

[76] H. Shefrin and M. Statman. The disposition to sell winners too early and ride losers too long: Theory and evidence. *The Journal of Finance*, 40(3):777–790, 1985.

[77] M. Strahilevitz, T. Odean, and B. Barber. Once burned, twice shy: How naive learning, counterfactuals and regret affect the repurchase of stocks previously sold. *Journal of Marketing Research*, 48:S102–S120, 2011.

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

[79] R. Thaler and E. Johnson. Gambling with the house money and trying to break even: The effects of prior outcomes on risky choice. *Management Science*, 36(6):643–660, 1990.

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

[81] M. Weber and C. Camerer. The disposition effect in securities trading: An experimental analysis. *Journal of Economic Behavior and Organization*, 33:167–184, 1998.

[82] M. Yaari. The dual theory of choice under risk. *Econometrica*, 55(1):95–115, 1987.