ECONOMETRICA

JOURNAL OF THE ECONOMETRIC SOCIETY

An International Society for the Advancement of Economic Theory in its Relation to Statistics and Mathematics

http://www.econometricsociety.org/

Econometrica, Vol. 84, No. 6 (November, 2016), 1983–2027

RANDOM CHOICE AND PRIVATE INFORMATION

JAY LU

UCLA, Los Angeles, CA 90095, U.S.A.

Econometrica, Vol. 84, No. 6 (November, 2016), 1983–2027

RANDOM CHOICE AND PRIVATE INFORMATION

BY JAY LU[1]

We consider an agent who chooses an option 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 menus from random choice and vice versa, (2) assess which agent has better information by using choice dispersion as a measure of informativeness, (3) determine if the agent's beliefs about information are dynamically consistent, and (4) test to see if these beliefs are well-calibrated or rational.

KEYWORDS: Random utility, Blackwell informativeness, informational dynamic consistency.

1. INTRODUCTION

1.1. *Overview and Motivation*

IN MANY ECONOMIC SITUATIONS, an agent's private information is not observable. If this information affects the agent's choice behavior, then from the perspective of an analyst who does not see the agent's information, these choices appear to be *random*. In this paper, we provide a theory in which the analyst can use the agent's random choice to identify her private information and also perform various exercises of inference.

We motivate our theory with two examples. In the first example, a lender (agent) is faced with a pool of loan applications and has to decide whether to approve each applicant. Approving an applicant results in a low payoff for the lender if the applicant defaults and a high payoff otherwise. There is a fixed information structure (e.g., credit check, socioeconomic data) that the lender uses to learn about the likelihood of default before making a decision. A regulator (analyst) wants to check whether the lender is following proper anti-discrimination policies. For instance, the lender could be using demographic information such as race or gender that should not be taken into account when evaluating loan applications. In many cases, this information is private to the lender and not observable to the regulator. Hence, from the regulator's perspective, the only observable data are the approval rates for the overall pool of applicants. Our theory shows how the regulator can use these approval rates

[1]I am deeply indebted to both Faruk Gul and Wolfgang Pesendorfer for their advice, guidance, and encouragement. I am very grateful to Stephen Morris for some highly valuable and insightful discussions. I would also like to thank Roland Bénabou, Simon Board, Simone Cerreia-Vioglio, Sylvain Chassang, Mark Dean, Leandro Gorno, Doron Ravid, Kota Saito, seminar audiences at Yale, UCLA, Northwestern, Kellogg MEDS, Cornell, MIT, UBC, BU, UPenn, SAET (Waseda), USC, Caltech, Duke-UNC, NYU, Wisconsin, Stanford, CETC (Western), RUD (Bocconi), a coeditor, and three anonymous referees for their helpful suggestions and feedback.

 DOI: 10.3982/ECTA12821

(random choice) to identify the lender's private information and determine if she is following relevant policies.

In the second example, a group of consumers (agents) choose whether to book a hotel online. Each consumer has some private information, such as a prior belief about how much she will be using the various amenities at the hotel (e.g., proximity to the beach). A website (analyst) wants to identify the consumers' private information in order to inform pricing strategies and advertising policies. For instance, suppose the website wants to assess the value of disclosing additional information in the form of online reviews. The optimal amount of information disclosure would then depend on the distribution of consumer beliefs. Since the consumers' information is private, the only observable data to the website are the booking frequencies of everyone in the group. Our theory shows how the website can use these booking frequencies (random choice) to identify the consumers' private information in order to implement the optimal disclosure policy.

In both examples above, the analyst is faced with the problem of identifying the agent's private information from random choice. In the first example, the random choice is the result of the lender (single agent) making choices in repetition across applicants. In the second example, the random choice is the result of many consumers (group of agents) choosing whether to book a hotel. We refer to these as the *individual* and *group interpretations* of random choice, respectively. There are many economic applications that fit in this framework. A policy-maker wants to identify the information used by a college for admission decisions. A government wants to evaluate the effects of providing tax-code information to low-income families. An insurance company wants to assess the amount of private information in a particular group of customers.[2] In all these examples, the identification of private information is an important and useful exercise for the analyst.

This paper provides a theoretical methodology for identifying private information from random choice. We also show how the analyst can perform standard exercises of inference. First, the analyst can use random choice to directly compute the agent's ex ante valuations of choice menus. In the loan example, this allows the regulator to use approval rates to determine how the lender would price a menu of loan products ex ante, that is, before slotting applicants into different loan products based on the lender's private information. Call this exercise *evaluating menus*. Second, the analyst can use choice dispersion to measure the informativeness of the agent's signal. This allows the regulator to distinguish between a discriminatory lender and one who is ignoring racial and gender information and following protocol. Call this *assessing informativeness*. Third, if the analyst observes both ex ante valuations and random choice,

[2]In health insurance, Finkelstein and McGarry (2006) and Hendren (2013) identified private information by directly eliciting it from survey data. In contrast, our approach follows the revealed preference methodology of Savage (1954) by inferring it from choice behavior.

then he can compare them to determine when the agent may be dynamically inconsistent with respect to her information. For example, the lender may anticipate to follow anti-discriminatory policies ex ante but violate them ex post. Call this *detecting biases*. Finally, if the analyst also observes the joint distribution of choices and payoff-relevant states, then he can test to see if the agent's beliefs are rational. A rational lender would have signals about applicants that are consistent with actual default rates. Call this *calibrating beliefs*. While we focus on the loan example for expository purposes, our results equally apply to the other examples (e.g., hotel bookings).

When information is observable, identification and inferences are important and well-understood exercises in information theory and information economics. Our main contribution is showing how to carry out these analyses even when information is *not* observable and can only be inferred from choice behavior. When the available choice data are relatively rich, such as with loan approvals or online hotel bookings, we show that identification and inferences can be performed just as effectively as in the case when information is observable. In the loan example, this means that the regulator can carry out all the analysis without observing any socioeconomic or demographic data on applicants.

When the available choice data are less rich, the application of our results is less immediate. Nevertheless, our theory provides a useful benchmark by highlighting the key conceptual issues that are relevant for identifying information and performing inference. It also provides a unifying methodology to help inform and guide what choice data should be collected in the lab or field when trying to identify private information. Moreover, in many settings, the nature of the environment will allow the analyst to make additional assumptions in order to reduce the dimensionality of the problem. With additional structure, our results imply sharper predictions that can be tested with less data. Combined with the fact that our main identification result only requires binary choices, this greatly reduces the burden on data availability. We leave the practical questions of performing these exercises when choice data are only partially available as interesting avenues for future research.

Our model follows the standard setup of Anscombe and Aumann (1963). There is an objective payoff-relevant state space and each choice option corresponds to a state-contingent payoff called an *act*. In the loan example, the state space is whether an applicant defaults or not. Approving a loan corresponds to choosing the act that gives the lender a low payoff if the applicant defaults and a high payoff otherwise. A *menu* is a finite set of acts. Since the agent's private information is not observed by the analyst, the only observable data (i.e., the primitive of the model) consist of a *random choice rule (RCR)* that specifies a choice distribution over acts for each menu. We consider an RCR that can be rationalized by a random utility maximization (RUM) model where the random utilities are subjective expected utilities that depend on the realization of the agent's beliefs.[3] The probability that an act is chosen is exactly the probability that it attains the highest subjective expected utility in the menu. Call this an *information representation* of the RCR. Theorem 1, our main identification result, shows that the analyst can completely identify the agent's private information from random choice over binary menus.

We introduce *test functions* as a technical innovation for using random choice to perform identification and inference. Given any menu, consider adding a *test act* that yields the same payoff in all states. As the value of the test act changes, the probability that acts in the original menu will be chosen over the test act will also vary. Call this graph the *test function* for that menu. In the loan example, variation in the test act could correspond to historical fluctuations in the value of the lender's outside option. Test functions are cumulative distribution functions that characterize the utility distributions of menus and serve as sufficient statistics for identifying information. This identification strategy assumes that the information structure does not vary with the choice menu, a natural premise for loan approvals or college admissions where a fixed signal structure is used by the agent for all decisions.[4] In the group interpretation, this is the standard assumption that beliefs are menu-independent as in Savage (1954) and Anscombe and Aumann (1963).

We now turn to the four exercises of inference. First, we show how the analyst can *evaluate menus*. In the individual interpretation, the valuation of a menu is the agent's ex ante utility of the menu before receiving any information. For example, this reflects how the lender would price a menu of loan products prior to receiving applicant credit reports. In the group interpretation, the valuation of a menu is its total utility for all agents in the group.[5] Menu valuations correspond to the subjective learning representation of Dillenberger, Lleras, Sadowski, and Takeoka (2014) (henceforth DLST). We provide tools for the analyst to directly compute valuations from random choice and vice versa. Theorem 2 shows that integrating test functions recovers valuations, while Theorem 3 shows that differentiating valuations with respect to test acts recovers random choice. These operations are mathematical inverses and provide a precise connection between menu choice and random choice. They offer a methodology for elicitation and identification that are in the same spirit as classical results from consumer and producer theory (for instance, Theorem 3 is the random choice analog of Hotelling's lemma).

[3]For more about RUM, see Block and Marschak (1960), Falmagne (1978), McFadden and Richter (1990), Gul and Pesendorfer (2006), and Gul, Natenzon, and Pesendorfer (2014). RUM is also used extensively in discrete choice estimation with specific parameterizations such as the logit, the probit, the nested logit, etc. (see Train (2009)).

[4]In loan approvals, the signal structure may be a protocol for obtaining credit scores or socioeconomic information that is set by senior management. Even if the lender is discriminatory, our results hold as long as the lender is uniformly discriminatory across all decisions. In college admissions, notions of fairness may prevent the college from changing the signal structure when the choice menu changes.

[5]McFadden (1981) called this the "social surplus."

Second, we show how to *assess informativeness*. In the classical approach of Blackwell (1951, 1953), better information is characterized by higher ex ante valuations. Theorem 4 shows that under random choice, better information is characterized by second-order stochastic dominance of test functions. One agent is better informed than another if and only if test functions under the latter second-order stochastic dominate those of the former. This allows the analyst to use a stochastic dominance relation over observable choice data to characterize a multidimensional ordering over unobservable private information. Intuitively, a more informative signal structure (or more private information in a group of agents) is characterized by greater dispersion or randomness in choice behavior.[6] In the loan example, the regulator can use approval rates to distinguish between a discriminatory lender who uses racial or gender information and a lender who follows protocol.[7]

Third, we apply these results to *detect biases*. Consider an analyst who observes both random choice and ex ante valuations. He can then detect when the agent may be dynamically inconsistent with respect to her information. In other words, the agent may have ex ante preferences (reflecting valuations) that suggest a more (or less) informative signal than that implied by random choice. In *prospective overconfidence*, she initially prefers large menus in anticipation of an informative signal but subsequently exhibits more deterministic choice indicating a less informative signal. In *prospective underconfidence*, the ordering is reversed. For example, a lender initially anticipates to ignore race and gender information but then ends up using it for loan approvals. In either case, we call this dynamic inconsistency *subjective misconfidence*. Our results show how the regulator can detect these biases even when the lender's signals are not directly observable.

Finally, we *calibrate beliefs*. We show that given joint data on choices *and* actual state realizations, the analyst can test whether the agent has well-calibrated (i.e., consistent) beliefs. In the individual interpretation, this implies that the agent has rational expectations with respect to her signals. For example, the lender's beliefs about her signals are consistent with actual default rates. In the group interpretation, this implies that agents' beliefs are predictive of actual state realizations so there is genuine private information in the group.[8] Using *conditional test acts* with payoffs that vary only in a given state,

[6]In the group interpretation, Theorem 4 is the revealed preference analog of Hendren (2013) who used elicited beliefs from survey data to test whether there is more private information in one group of agents (insurance rejectees) than another (non-rejectees).

[7]Our theory cannot distinguish between a non-discriminatory lender and one who gathers discriminatory information but chooses not to use it for loan approvals. Since we are ultimately interested in actions, the distinction between using and not using discriminatory information is the relevant comparison.

[8]See Chiappori and Salanié (2000) and Finkelstein and McGarry (2006) for empirical tests of the presence of private information in the context of health insurance.

we construct *conditional test functions*. Theorem 5 shows that beliefs are well-calibrated if and only if conditional and unconditional test functions share the same mean. Combined with previous results on bias detection, this allows the analyst to determine when the agent is objectively misconfident with respect to her information.

RUM models in general have difficulty dealing with indifferences in the random utility. Indifferences are useful in that they allow the analyst to leave select data unmodeled for purposes of tractability and flexibility. Classic deterministic choice permits this flexibility but only admits static choice. Traditional random choice models allow for more richness but impose too much rigidity by insisting all choice probabilities be fully specified. In our model, we relax this latter restriction. We thus bridge the gap by retaining both the flexibility of deterministic choice and the richness of random choice. This broadens the set of environments in which choice models can be applied. In particular, it allows us to include the subjective expected utility model of Anscombe and Aumann (1963) as a special case of ours.

Section 2 introduces the model and the main identification result. Test functions are introduced in Section 3 and the four inference exercises are addressed in Sections 4 to 7. All proofs are in the Appendix unless otherwise stated. In the Supplemental Material to this paper (Lu (2016)), we provide an axiomatic characterization of our model. It consists of four axioms (*monotonicity*, *linearity*, *extremeness*, and *continuity*) that are direct analogs of the random expected utility axioms from Gul and Pesendorfer (2006) and three new ones: *non-degeneracy* which ensures no universal indifference, *S-monotonicity* which ensures that acts that dominate in every state are chosen for sure, and *C-determinism* which ensures deterministic choice over constant acts.

1.2. *Related Literature*

This paper is related to a long literature on stochastic choice. Recent papers that have specifically studied the relationship between stochastic choice and information include Natenzon (2013), Caplin and Dean (2015), Matějka and McKay (2015), Ellis (2012), and Fudenberg, Strack, and Strzalecki (2015). In these models, the information structure varies with the menu so the resulting RCR may not necessarily have a RUM representation. In contrast, the information structure in our model is fixed, which conforms to the benchmark model of information processing and choice. Caplin and Martin (2015) also studied a RUM model with a fixed information structure. If we recast their model in our Anscombe–Aumann setup, our use of test functions to calibrate beliefs coincides with checking their NIAS inequalities.[9] While their model

[9]See Section S.3 in the Supplemental Material. While NIAS takes utilities as given, we take the random choice rule as the primitive.

requires less data, the richer setup in our model allows information to be uniquely identified from the RCR.

This paper is also related to the literature on menu choice which includes Kreps (1979), Dekel, Lipman, and Rustichini (2001) (henceforth DLR), and DLST. Our main contribution to this literature is showing that there is an intimate connection between ex ante choice *over* menus and ex post random choice *from* menus. Ahn and Sarver (2013) also studied this relationship where choice options are lotteries. Their work connecting DLR with Gul and Pesendorfer (2006) random expected utility is analogous to our results connecting DLST with our model. As our choice options reside in the richer Anscombe–Aumann space, we are able to achieve a tighter connection between menu and random choice (see Section S.2 in the Supplemental Material). Fudenberg and Strzalecki (2015) also analyzed the relationship between preference for flexibility and random choice but in a dynamic setting with a generalized logit model. Saito (2015) established a connection between greater preference for flexibility and more randomness in a model where the agent deliberately randomizes due to ambiguity aversion. Working with very different primitives, Karni and Safra (2016) also studied menu choice and implied random choice.

Grant, Kajii, and Polak (1998, 2000) studied decision-theoretic models involving information. They considered generalizations of the Kreps and Porteus (1978) model where the agent has an intrinsic preference for information even when she is unable or unwilling to act on that information. In contrast, the agent in our model prefers information only due to its instrumental value as in the classical sense of Blackwell.

In the strategic setting, Bergemann and Morris (2016) studied information structures in Bayes correlated equilibria. In the special case where there is a single bidder, our results translate directly to their setup for a single-person game. Kamenica and Gentzkow (2011) and Rayo and Segal (2010) characterized optimal information structures where senders commit to an information disclosure policy. In these models, the sender's ex ante utility is a function of the receiver's random choice rule, so our results relating random choice with valuations provide a technique for expressing the sender's utility in terms of the receiver's utility and vice versa. Chambers and Lambert (2014) also studied the elicitation of an expert's private information where they used a properly incentivized protocol for identification.

2. AN INFORMATIONAL MODEL OF RANDOM CHOICE

We now describe the basic setup of the model. Let S be a finite state space and ΔS be the set of beliefs about S. In the loan example, the state space is whether the applicant defaults or not. In the hotel example, the state space is whether the consumer makes use of the hotel amenities (e.g., visits the beach). Let X be a finite prize space. Payoffs are in the form of lotteries over prizes which we denote by ΔX. Each choice option of the agent corresponds to a state-contingent payoff $f : S \to \Delta X$ called an Anscombe–Aumann (1963) *act*. For example, approving a loan results in a low payoff for the lender if the applicant defaults and a high payoff otherwise. Booking an expensive hotel close to the beach results in a high payoff for the consumer if she visits the beach and a low payoff otherwise.

Let H be the set of all acts. A *menu* is a finite set of acts, and let $\mathcal{K}$ denote the set of all menus.[10] In the loan example, a menu could be a set of loan products that also includes the option to reject the applicant. In the hotel example, a menu could be a set of different hotel packages that also includes the option to not book the hotel. An act f is *constant* if it has the same payoff in all states. For notational convenience, we also let f denote the singleton menu $\{f\}$ whenever there is no risk of confusion.

The analyst observes a *random choice rule (RCR)* that specifies choice probabilities for acts in every menu. In the *individual interpretation* of random choice, the RCR specifies the choice frequency by the agent if she chooses from a given menu repeatedly. If the menu consists of whether or not to approve an applicant, then this corresponds to the lender's approval rate across applicants. In the *group interpretation* of random choice, the RCR specifies the frequency distribution of choices in the group if every agent in the group chooses from the given menu. If the menu consists of whether or not to book a hotel, then this corresponds to the booking frequency across all consumers in the group.

Under classic deterministic choice, if two acts are indifferent (i.e., they have the same utility), then the model is silent about which act the agent will choose. This allows the analyst to be agnostic about data that are orthogonal to the model at hand. Traditional random choice models lack this flexibility by insisting that all choice probabilities be fully specified. We introduce an innovation to model indifferences that retains this flexibility for random choice. If two acts are indifferent (i.e., they have the same random utility), then we allow the RCR to be silent about each act's individual choice probability. This provides the analyst with additional freedom to interpret data and allows for just enough flexibility so that we can include the deterministic Anscombe–Aumann (1963) model as a special case.

Formally, we model indifferences as non-measurable sets with respect to the RCR. Let $\mathcal{H}$ be a collection of measurable sets on H. For example, if $\mathcal{H}$ is the Borel σ-algebra,[11] then this is the benchmark case where every act is measurable and there are no indifferences. Indifferences occur when $\mathcal{H}$ is coarser than the Borel σ-algebra. Since the agent will choose some act from the menu, the

[10] We endow $\mathcal{K}$ with the Hausdorff metric. For two sets F and G, the Hausdorff metric is given by

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

[11] That is, the Borel σ-algebra generated by the Euclidean metric.

menu itself must be measurable. Hence, given any menu F, the corresponding random choice must be a probability on the σ-algebra generated by $\mathcal{H} \cup \{F\}$ which we denote by $\mathcal{H}_F$.[12] Let ΔH be the set of all probability measures on H. We formally define an RCR as follows.

DEFINITION: A *random choice rule (RCR)* is a $(\rho, \mathcal{H})$ where $\rho : \mathcal{K} \to \Delta H$ and ρ_F is a measure on $(H, \mathcal{H}_F)$ such that $\rho_F(F) = 1$.

For each menu F, the RCR ρ assigns a probability measure on $(H, \mathcal{H}_F)$ with support in F. We interpret $\rho_F(G)$ as the probability that some act in G will be chosen given the menu F. For ease of exposition, we denote RCRs by ρ with the implicit understanding of its association with some $\mathcal{H}$. To address the fact that G may not be $\mathcal{H}_F$-measurable, define the outer measure

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

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

An RCR is *deterministic* if all choice probabilities are either 0 or 1. The following example highlights the use of non-measurability to model indifferences and how classic deterministic choice is a special case of random choice.

EXAMPLE 1: Suppose there are two states $S = \{s_1, s_2\}$ and two prizes $X = \{x, y\}$. Without loss of generality, we can associate each act $f \in H$ with the point $f \in [0, 1] \times [0, 1]$ where $f_i = f(s_i)(x)$ for $i \in \{1, 2\}$. Let $\mathcal{H}$ be the σ-algebra generated by sets of the form $B \times [0, 1]$ where B is a Borel set on $[0, 1]$. Consider the RCR $(\rho, \mathcal{H})$ where $\rho_F(f) = 1$ if $f_1 \geq g_1$ for all $g \in F$. Hence, acts are ranked based on how likely they will yield prize x if state s_1 occurs. This could describe an agent who prefers x to y and believes that s_1 will realize for sure. If we consider a menu F that consists of two acts f and g where $f_1 = g_1$, then neither f nor g is $\mathcal{H}_F$-measurable; the RCR is unable to specify choice probabilities for f or g and they are indifferent. Observe that ρ corresponds exactly to classic deterministic choice where f is preferred to g if and only if $f_1 \geq g_1$.

We now describe an *information representation* of an RCR. Recall the timing of our model. First, the agent receives some private information about the underlying state. She then chooses the best act in the menu given her updated belief. Since her private information is unobservable to the analyst, choice is probabilistic and can be modeled as an RCR.

[12]This definition imposes a common measurability across all menus which can be relaxed if some of our axioms are strengthened (see Section S.1 in the Supplemental Material).

Since each signal realization corresponds to a posterior belief $q \in \Delta S$, we model private information as a *signal distribution* μ over the canonical signal space ΔS. This approach allows us to work directly with posterior beliefs and be agnostic about updating rules. In the loan example, μ is the distribution of the lender's beliefs about the likelihood of default. In the hotel example, μ is the distribution of consumer beliefs that they will use hotel amenities (e.g., visit the beach). If $\mu = \delta_q$ for some $q \in \Delta S$, then the signal distribution is degenerate. In the first example, this corresponds to a completely uninformative signal; in the second example, this corresponds to all consumers sharing the same belief. Note that even though μ could be the result of an information acquisition problem by the lender, as long as it is independent of the menu, the regulator can assume μ is exogenous without loss of generality.

Let $u : \Delta X \to \mathbb{R}$ be an affine utility function. An agent's subjective expected utility of an act f given her belief q is $q \cdot (u \circ f)$.[13] Given a utility function, a signal distribution is *regular* if 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 where utilities are never equal and allows us to handle indifferences.

DEFINITION: μ is *regular* if $q \cdot (u \circ f) = q \cdot (u \circ g)$ with μ-measure 0 or 1 for any f, g.

Going forward, we let (μ, u) denote a regular μ and a non-constant u. We can now define an *information representation* as follows.

DEFINITION—Information Representation: ρ is *represented by* (μ, u) if for any $f \in F \in \mathcal{K}$,

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

This is a RUM model where the random utilities are subjective expected utilities that depend on the agent's private information. The probability of choosing an act f is exactly the measure of beliefs that rank f higher than every other act in the menu. In the loan example, the approval rate is exactly the probability that the lender receives a good signal about an applicant. In the hotel example, the booking frequency is exactly the proportion of consumers who believe they will be using hotel amenities (e.g., visiting the beach). Note that from the analyst's perspective, both μ and u are unobserved and the only observable data is the RCR. When μ is degenerate, this reduces to the standard subjective expected utility model of Anscombe and Aumann (1963).

[13]We let $u \circ f \in \mathbb{R}^S$ denote the utility vector of an act $f \in H$ where $(u \circ f)(s) = u(f(s))$ for all $s \in S$.

As we are interested in studying the role of information in random choice, the taste utility u is fixed in an information representation. In the individual interpretation, this means that signals only affect beliefs but not tastes. In the group interpretation, this means that from the analyst's perspective, there is unobservable heterogeneity in beliefs but observable tastes. For instance, the RCR could be conditional on all the taste-relevant socioeconomic and demographic data on the group of agents. In the Supplemental Material to this paper, we study a more general model that allows for unobserved heterogeneity in tastes as well.[14] We also assume the standard assumption of state-independent (taste) utilities. This can be addressed in our model by considering random choice generalizations of classic solutions to state-dependent utility as in Karni, Schmeidler, and Vind (1983) and Karni (2007).[15]

Theorem 1 below states that by studying binary choices, the analyst can completely identify private information.

THEOREM 1: *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* $\alpha > 0$.

In other words, given two agents (or two groups of agents), comparing binary choices is sufficient to completely differentiate between the two information structures. In the loan example, the regulator can completely identify the lender's signal structure by only looking at approval rates for different loan products. In the hotel example, the website can completely identify the distribution of consumer beliefs by only looking at booking frequencies for different hotel offerings.

We end this section with a technical remark about regularity. As mentioned above, indifferences in traditional RUM must occur with probability zero. Since all choice probabilities are specified, these models run into difficulty when dealing with indifferences in the random utility. Our definition of regularity circumvents this issue by allowing for enough flexibility so that we can model indifferences using non-measurability. Note that our definition still imposes certain restrictions on μ. For instance, multiple mass points are not allowed if μ is regular.[16]

[14]See Section S.1 of the Supplemental Material.

[15]In practice, however, most of the applied literature has largely assumed state independence due to lack of better empirical evidence (see Finkelstein, Luttmer, and Notowidigdo (2009), for example).

[16]More precisely, our definition of regularity permits strictly positive measures on sets in ΔS that have less than full dimension. This relaxes the definition of regularity from Gul and Pesendorfer (2006) which requires μ to be full-dimensional (see their Lemma 2). See Block and Marschak (1960) for the case of finite alternatives.

3. TEST FUNCTIONS

We now introduce a key technical tool that will play an important role in our subsequent analysis. We show how the analyst can use the agent's RCR to construct *test functions* that will serve as powerful tools for performing identification and inference.

First, we show how these test functions can be constructed from choice data. Given any menu, consider the enlarged menu that contains an additional *test act* which yields a constant payoff in all states. In the loan example, a test act could correspond to the lender's outside option from rejecting an applicant. In the hotel example, it could correspond to the consumer's outside option from not booking a hotel. The analyst can then record the probability that the agent will choose some act in the original menu *over* the test act. For example, if the test act is very valuable (i.e., it yields a high payoff), then this probability will presumably be low. On the other hand, if the value of the test act is low, then this probability will be high. In both loan approvals and hotel bookings, variation in the test act could correspond to historical fluctuations in the value of the agent's outside option.

As the value of the test act changes, the probability that the agent will choose something in the menu over the test act will also change. We call this graph the *test function* for the original menu. Given any menu, the analyst can always record these choice probabilities in order to construct a test function for that menu. Test functions are derived completely from the observable choice data (i.e., the RCR). While they require the choice data to contain enough variation in test acts, once obtained, test functions allow the analyst to perform all the identification and inference exercises.

To gain some intuition for why test functions are useful constructs, note that each value of the test function corresponds to the probability that the utility of the menu is above some cutoff value. Hence, a menu's test function essentially maps out its utility distribution. Since choice is random due to information, this utility distribution is induced by the distribution of beliefs. Test functions thus provide a direct way of measuring the agent's information using the observable RCR.

We now formally define test functions. Call an act the *best* (*worst*) act under ρ if in any binary choice comparison, the act (other act) is chosen with certainty. In other words, letting $\overline{f}$ and $\underline{f}$ denote the best and worst acts, respectively, we have $\rho_{f\cup\overline{f}}(\overline{f}) = \rho_{f\cup\underline{f}}(f) = 1$ for any act f. If ρ is represented by (μ, u), then there exist constant best and worst acts.[17] *Test acts* are mixtures between the best and worst acts.

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

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

Note that test acts are also constant acts. Define *test functions* as follows.

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

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

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

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

Under deterministic choice, test functions reduce to mixtures between the best and worst acts that yield indifference. What follows is an example of a test function.

EXAMPLE 2: Consider the loan example with $S = \{s_1, s_2\}$ where s_1 corresponds to the applicant defaulting and s_2 otherwise. Consider a risk-neutral lender with normalized utility u such that the payoff of the best act is 1 and that of the worst act is 0. Suppose the lender's signal structure results in a uniform signal distribution μ. The lender's RCR ρ is represented by (μ, u). Suppose approving an applicant corresponds to choosing the act f that yields a low payoff of $\frac{1}{4}$ if the applicant defaults and a high payoff of $\frac{3}{4}$ otherwise. Rejecting an applicant corresponds to choosing the test act f^a with payoff $1 - a$ for $a \in [0,1]$. Hence, one can interpret a as a parameter reflecting variation in the lender's payoff for rejecting an applicant. The test function for f is given by

$$\begin{aligned} f_\rho(a) = \rho_{f \cup f^a}(f) &= \mu\left\{ q \in [0,1] \,\middle|\, q\frac{1}{4} + (1-q)\frac{3}{4} \geq 1 - a \right\} \\ &= \begin{cases} 0 & \text{if } a < \frac{1}{4}, \\ 2a - \frac{1}{2} & \text{if } \frac{1}{4} \leq a < \frac{3}{4}, \\ 1 & \text{if } \frac{3}{4} \leq a. \end{cases} \end{aligned}$$

Note that this is a cumulative distribution function and the approval rate increases with a.

A corollary of Theorem 1 is that knowing test functions is enough for the analyst to completely identify the agent's information structure.

COROLLARY 1: *Let ρ and τ have information representations. Then $\rho = \tau$ if and only if $f_\rho = f_\tau$ for all $f \in H$.*

Thus, test functions are "sufficient statistics" for identifying information. In other words, an analyst who knows all test functions can do just as good a job of identifying the signal distribution μ as an analyst who knows the entire sample of all random choices ρ.

4. EVALUATING MENUS

We now address our first exercise of inference and show that there is an intimate relationship between random choice and ex ante valuations of menus. Consider a *valuation preference relation* $\succeq$ over menus.

DEFINITION—Subjective Learning: $\succeq$ is *represented by* (μ, u) if it is represented by

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

In the individual interpretation, V gives the agent's ex ante valuation of menus prior to receiving her signal. This is the subjective learning representation axiomatized by DLST. In the loan example, this reflects how the lender would price a menu of loan products prior to receiving the credit reports on applicants. If she expects the reports to be very informative, then she will prefer larger menus ex ante, that is, she exhibits a *preference for flexibility*. In the group interpretation, V gives the total utility or "social surplus" for all agents in the group (see McFadden (1981)). In the hotel example, a consumer advocate who internalizes the utilities of all the consumers would be interested in calculating V. An advocate who thinks that consumer beliefs are more dispersed would prefer more flexible (i.e., larger) menus.

In this section, assume that ρ has a best and worst act and F_ρ is a well-defined cumulative distribution function for all $F \in \mathcal{K}$. Let $\mathcal{K}_0$ denote the subset of menus where every act in the menu is measurable with respect to the RCR.[18] Call an RCR *standard* if it satisfies the following properties.

DEFINITION: ρ is *standard* if it is
(1) *monotone*: $G \subset F$ implies $\rho_G(f) \geq \rho_F(f)$,
(2) *linear*: $\rho_F(f) = \rho_{aF+(1-a)g}(af + (1-a)g)$ for $a \in (0, 1)$,
(3) *continuous*: ρ is continuous on $\mathcal{K}_0$.[19]

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

[19]In other words, $\rho : \mathcal{K}_0 \to \Delta_0 H$ is continuous where $\Delta_0 H$ is the set of all Borel measures on H endowed with the topology of weak convergence. Note that $\rho_F \in \Delta_0 H$ for all $F \in \mathcal{K}_0$ without loss of generality.

Monotonicity says that the probability of choosing an act must not increase as more acts are added to the menu. It is necessary for any RUM model. Linearity is the random choice analog of the standard independence axiom. Continuity is the usual continuity axiom adjusted for indifferences. Any RCR that has an information representation is standard, while being standard is a relatively weak restriction.[20]

Given an RCR, consider ranking menus as follows.

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

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

A menu that is more valuable ex ante will have acts that are chosen more frequently over a test act. This yields a test function that takes on high values. Theorem 2 shows that $\succeq_\rho$ is exactly the valuation preference relation corresponding to the RCR ρ.

THEOREM 2: *The following are equivalent*:
(1) *ρ is represented by* (μ, u),
(2) *ρ is standard and $\succeq_\rho$ is represented by* (μ, u).

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

Theorem 2 also demonstrates that if a standard RCR induces a preference relation that has a subjective learning representation, then that RCR must have an information representation. In fact, both the RCR and the preference relation are represented by the same (μ, u). This serves as an alternate characterization of information representations using properties of its induced preference relation.

The discussion above suggests a converse: given valuations, can the analyst directly compute random choice? First, given any act f and state s, let $f(s)$ also denote the constant act that yields $f(s)$ in every state.

DEFINITION: $\succeq$ is *dominant* if $f(s) \succeq g(s)$ for all $s \in S$ implies $F \sim F \cup g$ for $f \in F$.

[20]See the axiomatic treatment in Section S.1 of the Supplemental Material. In fact, standardness may not be enough to ensure that a random utility representation even exists. In particular, an additional restriction called extremeness is needed.

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

DEFINITION: Given $\succeq$, let $\rho_{\succeq}$ denote any standard ρ such that almost everywhere (a.e.)

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

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

Given any preference relation $\succeq$, the RCR $\rho_{\succeq}$ may not even exist. On the other hand, there could be a multiplicity of RCRs that satisfy this definition. Theorem 3 shows that if $\succeq$ has a subjective learning representation, then these issues are irrelevant: $\rho_{\succeq}$ exists and is the unique RCR corresponding to $\succeq$.

THEOREM 3: *The following are equivalent*:
(1) $\succeq$ *is represented by* (μ, u),
(2) $\succeq$ *is dominant and* $\rho_{\succeq}$ *is represented by* (μ, u).

The probability that an act f_a is chosen is exactly its marginal contribution to the valuation of the menu. The more often the act is chosen, the greater its effect on the menu's overall valuation.[21] For instance, if the act is never chosen, then it will have no effect on valuations. One could interpret this as a cardinal version of Axiom 1 in Ahn and Sarver (2013) (see Section S.2 in the Supplemental Material). Any violation of this would indicate some form of inconsistency, which we explore in Section 6.

The analyst can now use $\rho_{\succeq}$ to directly compute random choice from valuations. To see how, first define ρ so that it coincides with $\succeq$ over all constant acts. Then use the definition of $\rho_{\succeq}$ to specify $\rho_{F\cup f_a}(f_a)$ for all $a\in[0,1]$ and $F\in\mathcal{K}$.[22] Linearity then extends ρ to all menus. By Theorem 3, the ρ so constructed is represented by (μ, u).

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

[21]In the econometrics literature, this is related to the Williams–Daly–Zachary theorem (see McFadden (1981)). The use of constant acts in the Anscombe–Aumann setup, however, means that Theorem 2 has no counterpart.

[22]Technically, $V(F\cup f_a)$ may be undifferentiable, but this occurs at most once so $\rho_{\succeq}$ is well-defined.

induced RCR. Moreover, combined with Theorem 1, this means that binary menus are sufficient for identification in the subjective learning representation.

Theorem 3 is the random choice version of Hotelling's lemma from classical producer theory. The analogy follows if we interpret choice probabilities as "outputs," conditional utilities as "prices," and valuations as "profits."[23] Similarly to how Hotelling's lemma is used to compute firm outputs from the profit function, Theorem 3 can be used to compute random choice from valuations. Combining these results, we obtain the following.

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

Similarly to how classical results from consumer and producer theory (e.g., Hotelling's lemma) provide a methodology for relating data, Corollary 2 allows an analyst to relate valuations with random choice and vice versa. Integrating test functions yields valuations, while differentiating valuations yields random choice. In the loan example, this means the regulator can use approval rates to calculate how the lender would price a set of loan products ex ante and vice versa. In the hotel example, this means a consumer advocate can use booking frequencies to calculate consumer welfare and vice versa. These operations allow the analyst to compute one type of data from another without the need to identify signal distributions or utilities explicitly.

5. ASSESSING INFORMATIVENESS

We now show how the analyst can infer which agent uses a more informative signal even when information is not directly observable. First, consider the classic methodology when information *is* observable. A transition kernel on the set of beliefs ΔS is *mean-preserving* if it preserves average beliefs.[24]

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

$$\int_{\Delta S} pK(q, dp) = q.$$

[23]Formally, interpret each act $f \in F$ as a "good," $y_f = \mu(Q_f)$ as the "output" given a partition $\mathcal{Q}_y = \{Q_f\}_{f\in F}$, and $p_f = \int_{Q_f} \frac{q\cdot(u\circ f)}{\mu(Q_f)}\mu(dq)$ as the "price." Hence, $V(F) = \sup_{y,\mathcal{Q}_y} p \cdot y$ is the maximizing "profit." Note that $a = p_{f_a}$ is exactly the price of f_a. The caveat is that in Hotelling's lemma, prices are fixed, while in our case, they depend on the partition $\mathcal{Q}_y$.

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

Let μ and ν be two signal distributions. We say μ is *more informative* than ν if 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 if there is a mean-preserving transition kernel K such that, for all $Q \in \mathcal{B}(\Delta S)$,

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

If μ is more informative than ν, then the information structure of ν can be generated by adding noise or "garbling" μ. This is Blackwell's (1951, 1953) ordering of informativeness based on signal sufficiency. For example, if K is the identity kernel, then no information is lost and $\nu = \mu$. In the loan example, a discriminatory lender who uses additional race and gender information would have a signal distribution that is more informative than that of a lender who follows protocol. In the hotel example, a group of consumers with more heterogeneous beliefs would have a signal distribution that is more informative than that of a group with less belief heterogeneity.

In the classical approach, Blackwell (1951, 1953) showed that better information is characterized by higher ex ante valuations. We now show how to characterize better information using random choice. First, consider a degenerate signal distribution corresponding to an uninformative signal (or a group of agents all with the same belief). Choice is deterministic in this case, so the test function of an act corresponds to a single mass point. Another agent (or group of agents) with more information will have test functions that have a more dispersed distribution. This is captured exactly by *second-order stochastic dominance*, that is, $F \geq_{\text{SOSD}} G$ if $\int_{\mathbb{R}} \phi \, dF \geq \int_{\mathbb{R}} \phi \, dG$ for all increasing concave $\phi : \mathbb{R} \to \mathbb{R}$.

THEOREM 4: *Let ρ and τ be represented by (μ, u) and (ν, u), respectively. Then μ is more informative than ν if and only if $F_\tau \geq_{\text{SOSD}} F_\rho$ for all $F \in \mathcal{K}$.*

Theorem 4 equates an unobservable multidimensional information ordering with an observable single-dimensional stochastic dominance relation. An analyst can assess informativeness simply by comparing test functions via second-order stochastic dominance. It is the random choice characterization of better information. The intuition is that better information corresponds to more dispersed (i.e., random) choice, while worse information corresponds to more concentrated (i.e., deterministic) choice. We illustrate this with an example.

EXAMPLE 3: Recall Example 2 from above and consider a second lender with a signal distribution ν that has density $6q(1-q)$. Recall that the first lender's RCR ρ is represented by (μ, u) and let the second lender's RCR τ be represented by (ν, u). As before, let f be the act corresponding to approving an applicant. Now, the test function of f under the second lender's RCR τ is

$$f_\tau(a) = \begin{cases} 0 & \text{if } a < \dfrac{1}{4}, \\ (1-a)(1-4a)^2 & \text{if } \dfrac{1}{4} \leq a < \dfrac{3}{4}, \\ 1 & \text{if } \dfrac{3}{4} \leq a. \end{cases}$$

Since $f_\tau \geq_{\text{SOSD}} f_\rho$, the first lender's signal distribution μ is more informative than that of the second lender ν. In fact, f_ρ is a mean-preserving spread of f_τ in this example. Thus, by looking at the approval rates of the two lenders and comparing test functions, the regulator can identify the lender who is using additional discriminatory information.

In DLST, better information is characterized by a greater preference for flexibility in the valuation preference relation. This is the choice-theoretic version of Blackwell's (1951, 1953) result. A preference relation exhibits *more preference for flexibility* than another if whenever the other prefers a menu to a singleton, the first must do so as well. Corollary 3 relates our random choice characterization of better information with more preference for flexibility.

DEFINITION: $\succeq_1$ has *more preference for flexibility* than $\succeq_2$ if $F \succeq_2 f$ implies $F \succeq_1 f$.

COROLLARY 3: *Let* ρ *and* τ *be represented by* (μ, u) *and* (ν, u), *respectively. Then the following are equivalent*:
(1) $F_\tau \geq_{\text{SOSD}} F_\rho$ *for all* $F \in \mathcal{K}$,
(2) $\succeq_\rho$ *has more preference for flexibility than* $\succeq_\tau$,
(3) μ *is more informative than* ν.

PROOF: By Theorem 4, (1) and (3) are equivalent. By Corollary 2, $\succeq_\rho$ and $\succeq_\tau$ are represented by (μ, u) and (ν, u), respectively. Hence, by Theorem 2 of DLST, (2) is equivalent to (3). *Q.E.D.*

Greater preference for flexibility and greater choice dispersion are the behavioral manifestations of better information. In the individual interpretation, a more informative signal corresponds to greater preference for flexibility (ex ante) and more randomness in choice (ex post). In the group interpretation, more private information corresponds to a greater social preference for flexibility and more heterogeneity in choice. Note the prominent role of test functions: computing their integrals evaluates menus, while comparing them via second-order stochastic dominance assesses informativeness.

If μ is more informative than ν, then it follows from the Blackwell ordering that the two distributions must have the same average belief. Combined with Theorem 4, this implies that test functions of singleton menus under the more informative signal is a mean-preserving spread of those under the less informative signal as in Example 3. This condition, however, is insufficient for assessing informativeness; it corresponds to a stochastic order over distributions that is strictly weaker than the Blackwell order.[25]

6. DETECTING BIASES

In this section, we study situations when an agent's private information inferred from valuations is inconsistent with that inferred from her random choice. In the individual interpretation, this inconsistency describes an agent whose *prospective* ex ante (pre-signal) beliefs about her signals are misaligned with her *retrospective* ex post (post-signal) beliefs. This is an informational version of the dynamic inconsistency as seen in Strotz (1955). In the group interpretation, this describes the situation when social preferences indicate a more (or less) dispersed distribution of beliefs in a group than that implied by random choice.

Consider the loan example where a lender expects to use credit scores that are not very informative. As a result, ex ante, the lender does not consider menus containing a variety of loan products to be particularly valuable. Ex post, however, she decides to use very detailed reports on applicants. As a result, the lender updates her beliefs about applicants by more than what she anticipated initially and exhibits greater randomness or dispersion in choice. This is exactly the case if the lender expects to follow protocol initially but then decides to use discriminatory information when approving loans. Call this *prospective underconfidence*, where confidence here refers to signal precision.

On the flip side, there may be situations where ex post choice reflects less confidence than that implied by ex ante preferences. For example, suppose the lender initially expects to use very detailed credit reports. Hence, she ex ante prefers large menus and may be willing to pay a cost in order to postpone her decision and "keep her options open." Ex post, however, she decides to ignore the reports all together and approve all applicants in the pool. The analyst can deduce from the lender's behavior that she incorrectly anticipated a much more informative signal. Call this *prospective overconfidence.*

Since beliefs in our model are completely subjective, we are silent as to which period's behavior is more "correct." Both prospective overconfidence and underconfidence are relative comparisons involving *subjective misconfidence*. This

[25]Formally, test functions of singleton menus under ρ is a mean-preserving spread of those under τ if and only if the signal distribution of τ stochastically dominates that of ρ under the linear concave order. Although the Blackwell order implies the linear concave order, the converse is not true (see Section 3.5 of Muller and Stoyan (2002)).

is a form of belief misalignment that is independent of the true information structure and is in a sense more fundamental as it indicates a violation of intrapersonal consistency.[26]

Let $(\succeq, \rho)$ denote the valuation preference relation $\succeq$ and the RCR ρ. Motivated by Theorem 4, define subjective misconfidence as follows.[27]

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

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

Corollary 4 follows immediately from Corollary 2 and Theorem 4. It provides a choice-theoretic foundation for subjective misconfidence. Returning to the loan example, Corollary 4 allows a regulator to use the lender's ex ante valuations and ex post approval rates to detect when the lender may have initially planned to follow protocol but ultimately decided to use discriminatory information (or vice versa).

We can also apply Corollary 4 to study other behavioral biases. In the diversification bias (see Read and Loewenstein (1995)), an agent ex ante prefers large menus with a wide variety of food options, but ex post always chooses her favorite food. Hence, the agent initially overestimates the degree to which she will be taking advantage of the wide variety of options. If her choices are driven by informational reasons, then this corresponds exactly to prospective overconfidence.[28] Other biases involving information processing include the hot-hand fallacy and the gambler's fallacy (see Gilovich, Vallone, and Tversky (1985) and Rabin (2002), respectively). If we assume that the agent is unaffected by these biases ex ante but becomes afflicted ex post, then the hot-hand and gambler's fallacies correspond to prospective underconfidence and overconfidence, respectively. Corollary 4 also allows us to rank the severity of these biases via the Blackwell ordering of information structures and provides a unifying methodology to study a wide variety of informational biases.

7. CALIBRATING BELIEFS

Following Savage (1954) and Anscombe and Aumann (1963), we have adopted a purely subjective treatment of beliefs. Our theory identifies when

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

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

[28]Although the diversification bias was originally in reference to uncertainty over tastes, there are many informational reasons why one would prefer one food over another (news of a candy recall, for example).

observed choice behavior is consistent with some distribution of beliefs, but is unable to recognize when these beliefs may be incorrect.[29] For example, our notions of misconfidence in the previous section are descriptions of *subjective* belief misalignment and not measures of *objective* misconfidence.

In this section, we incorporate additional data to achieve this distinction. By studying the joint distribution over choices and state realizations, the analyst can test whether agents' beliefs are objectively well-calibrated. In the individual interpretation, this implies that the agent has rational expectations about her signals. In the loan example, this means that the lender's beliefs about her signals are consistent with actual default rates. 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 the hotel example, this means that consumers' beliefs correctly predict their use of hotel amenities (e.g., visiting the beach). If information is observable, then calibrating beliefs is a well-understood statistical exercise.[30] We show how the analyst can use test functions to calibrate beliefs even when information is not observable.

Let $r \in \Delta S$ be some observed distribution over states. Assume that r has full support without loss of generality. For example, in the loan example, r is the unconditional default rate of applicants.[31] In this section, the primitive consists of r and a *state-dependent random choice rule (sRCR)* $\rho := (\rho_s)_{s\in S}$ that specifies a RCR for each state.[32] Let $\rho_{s,F}(f)$ denote the probability of choosing $f \in F$ given state $s \in S$. The unconditional RCR is

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

Note that r in conjunction with the sRCR ρ completely specify the joint distribution over choices and state realizations.[33]

Information now corresponds to a joint distribution over beliefs and state realizations. In this section, let $\mu := (\mu_s)_{s\in S}$ be a *state-dependent signal distribution* where μ_s is the signal distribution conditional on $s \in S$. Let (μ, u) denote a state-dependent signal distribution μ and a non-constant u. We now define a *state-dependent information representation* and say ρ is represented by

[29]In Section S.1 of the Supplemental Material, we also provide an axiomatic treatment that incorporates the observed distribution of states as part of the primitive.

[30]For example, see Dawid (1982).

[31]This is the default rate unconditional on the decision of the lender. For instance, if not defaulting corresponds to the applicant's project for the loan succeeding, then this is the unconditional rate of project success.

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

[33]State-dependent stochastic choice was studied by Caplin and Martin (2015) and Caplin and Dean (2015) who first demonstrated the feasibility of collecting such data for individuals. In the group interpretation, these data are also readily available (see Chiappori and Salanié (2000)).

(μ, u) if ρ_s is represented by (μ_s, u) for all $s \in S$. Note that this does not imply beliefs are well-calibrated since μ may not be consistent with the observed frequency r. Define the unconditional signal distribution as

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

DEFINITION: μ is *well-calibrated* if, for all $s \in S$ and $Q \in \mathcal{B}(\Delta S)$,

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

Well-calibration implies that μ satisfies Bayes's rule. For each $s \in S$, μ_s is exactly the conditional signal distribution as implied by μ. In other words, choice behavior implies beliefs that agree with the observed joint data on choices and state realizations.

We now show how the analyst can test for well-calibrated beliefs using test functions. Let ρ be represented by (μ, u). Since u is fixed, both the best and worst acts are well-defined for ρ. Given a state $s \in S$, define a *conditional worst act* $\underline{f}^s$ as the act that coincides with the worst act if s occurs and with the best act otherwise.[34] Let $f_s^a := a\underline{f}^s + (1-a)\overline{f}$ be a *conditional test act* and define a *conditional test function* as follows.

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

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

A conditional test function specifies the conditional choice probability as we vary the conditional test act from the best to the conditional worst act. As with unconditional test functions, conditional test functions are increasing functions that are cumulative distribution functions if $F_\rho^s(r_s) = 1$. Let $\mathcal{K}_s$ denote all menus with conditional test functions that are cumulatives.

THEOREM 5: *Let ρ be represented by (μ, u). Then μ is well-calibrated if and only if F_ρ^s and $F_{\overline{\rho}}$ share the same mean for all $F \in \mathcal{K}_s$ and $s \in S$.*

Theorem 5 equates well-calibrated beliefs with the requirement that both conditional and unconditional test functions have the same mean. If beliefs are well-calibrated, then the ex ante value of a menu calculated from the state-dependent RCR should be the same once adjusted for the prior in that

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

state. This is a random choice characterization of rational beliefs using test functions.[35] In the loan example, this means that the regulator can use state-dependent approval rates to determine if the lender has unbiased beliefs about her signals. In the hotel example, this means that the website can use state-dependent booking frequencies to determine if consumers have genuine private information.

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

APPENDIX A: MODEL AND IDENTIFICATION

A.1. *Preliminaries*

In this section, we introduce some preliminary notation and results regarding indifferences. Recall that $\mathcal{H}$ is a σ-algebra on H corresponding to indifferences, and for any menu F, $\mathcal{H}_F$ is the σ-algebra generated by $\mathcal{H} \cup \{F\}$. For $G \subset F \in \mathcal{K}$, let G_F denote the smallest $\mathcal{H}_F$-measurable set containing G, that is,

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

Since menus are finite, if $G \subset F$, then G_F must also be $\mathcal{H}_F$-measurable. We first show the following useful property.

LEMMA A.1: *If $G \subset F \subset J \in \mathcal{K}$, then $G_F = G_J \cap F$.*

PROOF: Let $G \subset F \subset J \in \mathcal{K}$. First, note that since $G \subset F \in \mathcal{H}_F$,

$$G_F = \bigcap_{G \subset E \in \mathcal{H}_F} (E \cap F) = \left(\bigcap_{G \subset E \subset F; E \in \mathcal{H}} E \right) \cap F.$$

By the same argument, we have

$$G_J = \left(\bigcap_{G \subset E \subset J; E \in \mathcal{H}} E \right) \cap J.$$

[35]This approach is also related to the NIAS restrictions of Caplin and Martin (2015) (see Section S.3 in the Supplemental Material).

Since $F \subset J$, this implies that

$$G_J \cap F = \left(\bigcap_{G \subset E \subset J; E \in \mathcal{H}} E \right) \cap F = G_F,$$

as desired. *Q.E.D.*

Since G_F is well-defined, we can define the outer measure

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

and let ρ denote ρ^* without loss of generality. We will also use the condensed notation $\rho(F, G) := \rho_{F \cup G}(F)$.

Two acts are *tied* if $\rho(f, g) = \rho(g, f) = 1$. Note that if ρ is deterministic, then this is exactly the traditional notion of indifference. The following are three equivalent characterizations of indifference that will be useful in our subsequent analysis.

LEMMA A.2: *Let* $\{f, g\} \subset F \in \mathcal{K}$. *The following are equivalent*:
(1) $\rho(f, g) = \rho(g, f) = 1$,
(2) $g \in f_F$,
(3) $f_F = g_F$.

PROOF: We will prove that (1) implies (2) implies (3) implies (1). First, suppose (1) is true so $\rho(f, g) = \rho(g, f) = 1$. We will show that $f_{f \cup g} = f \cup g$. Suppose otherwise, so $f_{f \cup g}$ is just the singleton f. Now, since $f \cup g \in \mathcal{H}_{f \cup g}$, $g \in \mathcal{H}_{f \cup g}$ so $g_{f \cup g}$ is just the singleton g. This implies that $\rho(f, g) + \rho(g, f) = 2 > 1$, yielding a contradiction. Thus, $f_{f \cup g} = f \cup g$. Now, since $f \cup g \subset F$, applying Lemma A.1 yields

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

Thus, $g \in f_F$ so (1) implies (2).

Now, suppose (2) is true, so $g \in f_F$ and thus $g \in g_F \cap f_F$. Since $g_F \cap f_F \in \mathcal{H}_F$, the definition of g_F means that $g_F \subset g_F \cap f_F$, which implies that $g_F \subset f_F$. We will now show that $f \in g_F$. Suppose otherwise, so $f \in f_F \setminus g_F \in \mathcal{H}_F$. As a result, $f_F \subset f_F \setminus g_F$, which implies that g_F must be empty, yielding a contradiction. Thus, $f \in g_F$ so $f \in g_F \cap f_F$. By the definition of f_F, this implies that $f_F \subset g_F \cap f_F$ so $f_F \subset g_F$. Hence, $f_F = g_F$ so (2) implies (3).

Finally, suppose (3) is true so $f_F = g_F$. This implies that $f \cup g \subset f_F$. Applying Lemma A.1 yields

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

Hence, $\rho(f,g) = \rho_{f\cup g}(f \cup g) = 1$. By symmetric reasoning, $\rho(g,f) = 1$ so (3) implies (1). This shows that (1), (2), and (3) are all equivalent. *Q.E.D.*

We now show that replacing acts with ties does not change choice probabilities; in other words, RCRs are insensitive to ties.

LEMMA A.3: *Let ρ be monotonic. Suppose $F = \bigcup_i f_i$, $G = \bigcup_i g_i$ where 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 first show that if g is tied with some act $h \in F$, then $\rho_F(f) = \rho_{F\cup g}(f)$ for any $f \in F$. By Lemma A.2, we can find $h^j \in F$ such that $\{h_F^1,\dots,h_F^m\}$ forms a unique partition of F. Without loss of generality, assume g is tied with h^1. By Lemma A.2 again, $h_{F\cup g}^1 = h_F^1 \cup g$ and $h_{F\cup g}^j = h_F^j$ for all $j > 1$. By monotonicity,

$$\rho_F(h_F^j) = \rho_F(h^j) \ge \rho_{F\cup g}(h^j) = \rho_{F\cup g}(h_{F\cup g}^j).$$

Now, for any $f \in h_F^k$, we know that $f \in h_{F\cup g}^k$ and

$$\rho_F(f) = 1 - \sum_{j\neq k} \rho_F(h_F^j) \le 1 - \sum_{j\neq k} \rho_{F\cup g}(h_{F\cup g}^j) = \rho_{F\cup g}(f).$$

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

We now prove the main result. 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 above and the fact that f_i and g_i are tied, 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. *Q.E.D.*

Finally, we show that ties are insensitive to (Minkowski) mixing. For conciseness, let $GaF := aG + (1-a)F$ denote the a-mixture of menu G with menu F.

LEMMA A.4: *Let ρ be monotonic and linear. Then $(fah)_{Fah} = (f_F)ah$ for $f \in F \in \mathcal{K}$.*

PROOF: Suppose $gah \in (f_F)ah$ for some $g \in f_F$. By Lemma A.2, g is tied with f. Hence, by linearity,

$$\rho(fah, gah) = \rho(f,g) = 1,$$
$$\rho(gah, fah) = \rho(g,f) = 1,$$

so gah is tied with fah. By Lemma A.2, $gah \in (fah)_{Fah}$ so $(f_F)ah \subset (fah)_{Fah}$. Now, suppose $gah \in (fah)_{Fah}$ so gah is tied with fah. By Lemma A.2 and linearity again, this implies that $\rho(f,g) = \rho(g,f) = 1$ so f and g are tied. Thus, $g \in f_F$ and $gah \in (f_F)ah$. This shows that $(f_F)ah = (fah)_{Fah}$. *Q.E.D.*

A.2. *Proof of Lemma 1*

To simplify notation, denote test acts by $f^a := \underline{f}a\overline{f}$ for $a \in [0, 1]$. First, we show that test functions are well-defined, that is, they are insensitive to the choice of any particular best or worst acts.

LEMMA A.5: *Suppose $\overline{f}$ and $\overline{g}$ are best acts and $\underline{f}$ and $\underline{g}$ are worst acts. Then $\rho(F, \underline{f}a\overline{f}) = \rho(F, \underline{g}a\overline{g})$ for all $F \in \mathcal{K}$.*

PROOF: Since $\overline{f}$ and $\overline{g}$ are best acts, $\rho(\overline{f}, \overline{g}) = \rho(\overline{g}, \overline{f}) = 1$. Since $\underline{f}$ and $\underline{g}$ are worst acts, $\rho(\underline{f}, \underline{g}) = \rho(\underline{g}, \underline{f}) = 1$. Thus, $\overline{f}$ is tied with $\overline{g}$ and $\underline{f}$ is tied with $\underline{g}$. By Lemma A.4, f^a is tied with $g^a = \underline{g}a\overline{g}$, so by Lemma A.3, $\rho(F, f^a) = \rho(F, g^a)$ as desired. *Q.E.D.*

We will now prove the following result characterizing expectations with respect to test functions. This will automatically imply Lemma 1.

LEMMA A.6: *Let ρ be represented by (μ, u). Then for any measurable $\phi : \mathbb{R} \to \mathbb{R}$,*

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

PROOF: Fix some menu F and define the function $\psi : \Delta S \to [0, 1]$ such that

$$\psi(q) := \frac{q \cdot (u \circ \overline{f}) - \sup_{f \in F} q \cdot (u \circ f)}{q \cdot (u \circ \overline{f}) - q \cdot (u \circ \underline{f})}.$$

Note that ψ is well-defined as u is non-constant. Let $\lambda := \mu \circ \psi^{-1}$ be the image measure of ψ on $[0, 1]$. By a standard change of variables, for any measurable $\phi : \mathbb{R} \to \mathbb{R}$,

$$\int_{[0,1]} \phi(x)\lambda(dx) = \int_{\Delta S} \phi\big(\psi(q)\big)\mu(dq).$$

We now show that F_ρ is exactly the cumulative of λ. Now,

$$\begin{aligned}\lambda[0, a] &= \mu \circ \psi^{-1}[0, a] = \mu\big\{q \in \Delta S | a \geq \psi(q) \geq 0\big\} \\ &= \mu\Big\{q \in \Delta S | \sup_{f \in F} q \cdot (u \circ f) \geq q \cdot \big(u \circ f^a\big)\Big\}.\end{aligned}$$

First, assume f^a is tied with nothing in F. Since μ is regular, for all $f \in F$, $q \cdot (u \circ f^a) = q \cdot (u \circ f)$ with μ-measure 0. Thus,

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

Now, assume f^a is tied with some $g \in F$ so $q \cdot (u \circ f^a) = q \cdot (u \circ g)$ μ-a.s. By Lemma A.2, $f^a \in g_{F \cup f^a}$ so

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

Hence, $\lambda[0, a] = F_\rho(a)$ for all $a \in [0, 1]$. Note that $\lambda[0, 1] = 1 = F_\rho(1)$ so F_ρ is the cumulative of λ as desired. *Q.E.D.*

A.3. *Proofs of Theorem 1 and Corollary 1*

Let ρ and τ be represented by (μ, u) and (ν, v), respectively. Let $\overline{f}$, $\underline{f}$ and $\overline{g}$, $\underline{g}$ be the best and worst acts of ρ and τ, respectively. We will show that the following are all equivalent:

(1) $\rho(f, g) = \tau(f, g)$ for all f and g,
(2) $\rho = \tau$,
(3) $(\mu, u) = (\nu, \alpha v + \beta)$ for $\alpha > 0$,
(4) $f_\rho = f_\tau$ for all f.

Note that the equivalence of (1) and (3) proves Theorem 1 and the equivalence of (4) proves Corollary 1. We will show that (3) implies (2) implies (1) implies (4) implies (3).

First, suppose (3) is true so $(\mu, u) = (\nu, \alpha v + \beta)$ for $\alpha > 0$. Thus, $\rho_F(f) = \tau_F(f)$ 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 (1) trivially and implies (4) by Lemma A.5.

Hence, all that remains is to prove that (4) implies (3). Suppose (4) is true so $f_\rho = f_\tau$ for all $f \in H$. First, we will show that $u = \alpha v + \beta$ for $\alpha > 0$. By Lemma A.5, we can assume that both sets of best and worst acts are constant acts. Note that for any two constant acts $u(f) \geq u(g)$ iff $\rho(f, g) = 1$ and $v(f) \geq v(g)$ iff $\tau(f, g) = 1$. Now, for any constant f, suppose $u(f) = u(f^a)$ for some $a \in [0, 1]$. Hence,

$$1 = f_\rho(a) = f_\tau(a) = \tau(f, g^a),$$

so $v(f) \geq v(g^a)$. If $a = 0$, then clearly $v(f) = v(g^0) = v(\overline{g})$. If $a > 0$, then for any $c < a$, $u(f) < u(f^c)$ so

$$0 = f_\rho(c) = f_\tau(c) = \tau(f, g^c),$$

so $v(f) < v(g^a)$ for all $c < a$. By continuity, $v(f) = v(g^a)$. Now, suppose $u(f) \geq u(g)$ and let $u(f) = u(f^a)$ and $u(g) = u(f^b)$ so $a \leq b$. Since $v(f) = v(g^a)$ and $v(g) = v(g^b)$, $v(f) \geq v(g)$. By symmetric argument, we have $u(f) \geq u(g)$ iff $v(f) \geq v(g)$. Since both u and v are linear, $u = \alpha v + \beta$ for $\alpha > 0$.

Thus, without loss of generality, we can assume $1 = u(\overline{f}) = v(\overline{g})$ and $0 = u(\underline{f}) = v(\underline{g})$ and $u = v$. Now,

$$\begin{aligned}
&\mu\{q \in \Delta S | q \cdot (u \circ f) \geq 1 - a\} \\
&\quad = \rho(f, f^a) = f_\rho(a) = f_\tau(a) \\
&\quad = \tau(f, g^a) = \nu\{q \in \Delta S | q \cdot (v \circ f) \geq 1 - a\}.
\end{aligned}$$

Since this is true for any $f \in H$ and $a \in [0, 1]$, the distributions of $q \cdot w$ under μ and ν are the same for any $w \in [0, 1]^S$. Hence, by the Cramér–Wold theorem, $\mu = \nu$ so (3) is true, as desired.

APPENDIX B: INFERENCES

This section includes proofs for the inference results. For any two cumulatives F and G on $[0, 1]$, let $F \geq_m G$ denote that F has a higher mean than G, that is, $\int_{[0,1]} x\, dF(x) \geq \int_{[0,1]} x\, dG(x)$. The following properties follow from standard results (see Billingsley (1986)).

LEMMA B.1: *Let F and G be cumulatives on* $[0, 1]$.
(1) $\int_{[0,1]} F(a)\, da = 1 - \int_{[0,1]} a\, dF(a)$.
(2) $F = G$ if and only if $F = G$ *a.e.*

B.1. *Proof of Theorem 2*

Before proving Theorem 2, we first show a few useful lemmas. The first provides a convenient expression for test functions of menus containing test acts. Recall the notation $f^a := \underline{f} a \overline{f}$ and $f_a := \overline{f} a \underline{f}$ for $a \in [0, 1]$.

LEMMA B.2: *If ρ is standard, then* $(F \cup f^b)_\rho = \max\{F_\rho, f^b_\rho\}$ *for all* $b \in [0, 1]$.

PROOF: Let ρ be standard and $\overline{f}$ and $\underline{f}$ be the best and worst acts of ρ. Note that if $\rho(\underline{f}, \overline{f}) > 0$, then $\underline{f}$ and $\overline{f}$ are tied, so by Lemma A.3, $\rho(\underline{f}, f) = \rho(f, \underline{f}) = 1$ for all $f \in H$. Thus, all acts are tied, so $(F \cup f^b)_\rho = 1 = \max\{F_\rho, f^b_\rho\}$ trivially.

Now, suppose $\rho(\underline{f}, \overline{f}) = 0$. Note that linearity implies $\rho(f^b, f^a) = 1$ if $a \geq b$ and $\rho(f^b, f^a) = 0$ otherwise. Thus

$$f^b_\rho(a) = \begin{cases} 1 & \text{if } a \geq b, \\ 0 & \text{otherwise,} \end{cases}$$

so for any $F \in \mathcal{K}$,

$$\max\{F_\rho(a), f_\rho^b(a)\} = \begin{cases} 1 & \text{if } a \geq b, \\ F_\rho(a) & \text{otherwise.} \end{cases}$$

Consider the menu $G := F \cup \{f^b, f^a\}$ so

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

First, suppose $a \geq b$. If $a > b$, then $\rho(f^a, f^b) = 0$ so $\rho_G(f^a) = 0$ by monotonicity. Hence, $\rho_G(F \cup f^b) = 1$. If $a = b$, then $\rho_G(F \cup f^b) = 1$ trivially. Thus, $(F \cup f^b)_\rho(a) = 1$ for all $a \geq b$. Now consider $a < b$ so $\rho(f^b, f^a) = 0$, which implies $\rho_G(f^b) = 0$ by monotonicity. First, suppose f^a is tied with nothing in F. Thus, by Lemma A.2, $f_G^a = f_{F \cup f^a}^a = f^a$ so

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

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

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

Finally, suppose f^a is tied with some $h \in F$. Thus, by Lemma A.3,

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

This implies that

$$(F \cup f^b)_\rho(a) = \begin{cases} 1 & \text{if } a \geq b, \\ F_\rho(a) & \text{otherwise,} \end{cases}$$

so $(F \cup f^b)_\rho = \max\{F_\rho, f_\rho^b\}$ as desired. *Q.E.D.*

Next, we show that if two standard RCRs share the same test functions, then they are the same.

LEMMA B.3: *Let ρ and τ be standard. If $F_\rho = F_\tau$ for all $F \in \mathcal{K}$, then $\rho = \tau$.*

PROOF: Let $\overline{f}$, $\underline{f}$ and $\overline{g}$, $\underline{g}$ be the best and worst acts of ρ and τ, respectively, so for any act f,

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

Suppose $F_\rho = F_\tau$ for all $F \in \mathcal{K}$. Note that

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

so $\overline{f}$ and $\overline{g}$ are τ-tied. Also, note that for any $a < 1$,

$$\tau(\underline{f}, g^a) = \underline{f}_\tau(a) = \underline{f}_\rho(a) = \rho(\underline{f}, f^a) = 0.$$

This implies that $\tau(g^a, \underline{f}) = 1$ for all $a < 1$. Since τ is continuous, this means that $\tau(\underline{g}, \underline{f}) = 1$ so $\underline{f}$ and $\underline{g}$ are also τ-tied. Note that this implies f^a and g^a are τ-tied for all $a \in [0, 1]$.

We now show that ties under ρ and τ agree, that is, $\rho(f, g) = 1$ iff $\tau(f, g) = 1$. Note that for any f and g, by linearity and Lemma A.4, we can assume that $g = f^a$ for some $a \in [0, 1]$ without loss of generality. Hence

$$\rho(f, f^a) = f_\rho(a) = f_\tau(a) = \tau(f, g^a) = \tau(f, f^a),$$

where the last equality follows from Lemma A.3. Thus, $\rho(f, f^a) = 1$ iff $\tau(f, f^a) = 1$. This establishes that f and g are ρ-tied iff they are τ-tied, so $\mathcal{H}_F^\rho = \mathcal{H}_F^\tau$ for all $F \in \mathcal{K}$.

Finally, we show that the two RCRs are the same. Consider $f \in G$. Again by linearity and Lemma A.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 $F := G \setminus f^a$. 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 $F' := F \setminus f_G^a$. By Lemma A.3, $\rho_G(f) = \rho_{F' \cup f}(f)$ and $\tau_G(f) = \tau_{F' \cup f}(f)$, 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, $\rho = \tau$ as desired. *Q.E.D.*

Finally, the following useful result shows that the integral of a menu's test function is exactly its ex ante utility.

LEMMA B.4: *If ρ is represented by (μ, u), then for all $F \in \mathcal{K}$,*

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

PROOF: Let ρ be represented by (μ, u) and assume without loss of generality that u is normalized. By Lemma A.5, without loss of generality, we can let $\overline{f}$ and $\underline{f}$ be constant acts that are the best and worst acts of ρ. Thus, $u(\overline{f}) = 1$

and $u(\underline{f}) = 0$. Now, from part (1) of Lemma B.1 and Lemma A.6,

$$\begin{aligned} V_\rho(F) &= \int_{[0,1]} F_\rho(a)\, da = 1 - \int_{[0,1]} a\, dF_\rho(a) \\ &= 1 - \int_{\Delta S} \Big(1 - \sup_{f \in F} q \cdot (u \circ f)\Big) \mu(dq) \\ &= \int_{\Delta S} \sup_{f \in F} q \cdot (u \circ f) \mu(dq), \end{aligned}$$

as desired. *Q.E.D.*

We are now ready to prove Theorem 2. We wish to show the following are equivalent:

(1) ρ is represented by (μ, u).

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

First suppose (1) is true. Lemma B.4 immediately implies that $\succeq_\rho$ is represented by (μ, u), and ρ is standard follows from the information representation (see Theorem S.2 in the Supplemental Material). This proves (2).

Now, suppose (2) is true so ρ is standard and $\succeq_\rho$ is represented by (μ, u). Define V such that

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

with u normalized so V represents $\succeq_\rho$. Let $\overline{f}$ and $\underline{f}$ be the best and worst acts of ρ. We first show that $\rho(\underline{f}, \overline{f}) = 0$. Suppose otherwise, so $\underline{f}$ and $\overline{f}$ must be tied. By Lemma A.3,

$$\rho(f, \overline{f}) = \rho(f, \underline{f}) = 1,$$

so every f is tied with $\overline{f}$. This implies that $f_\rho(a) = 1$ for all $a \in [0,1]$ so $V_\rho(f) = V_\rho(g)$ for all acts f and g. Since $\succeq_\rho$ is represented by (μ, u), this contradicts the fact that u is non-constant. This proves that $\rho(\underline{f}, \overline{f}) = 0$.

We now show that $V(\overline{f}) = 1$ and $V(\underline{f}) = 0$. Note that

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

which implies $\underline{f} \preceq_\rho f \preceq_\rho \overline{f}$. Thus, $V(\underline{f}) \leq V(f) \leq V(\overline{f})$ for all $f \in H$ so $V(\overline{f}) = 1$ and $V(\underline{f}) = 0$.

Define τ as the RCR represented by (μ, u), so by Lemma B.4, $V_\tau = V$. We show that $F_\rho = F_\tau$ for all $F \in \mathcal{K}$. Since $\succeq_\rho$ is represented by V, $V_\rho(F) =$

$\phi(V(F))$ for some monotonic transformation $\phi:\mathbb{R}\to\mathbb{R}$. Since $\rho(\underline{f},\overline{f})=0$, for any $b\in[0,1]$,

$$\begin{aligned}1-b&=\int_{[0,1]} f^b_\rho(a)\,da=V_\rho\big(f^b\big)=\phi\big(V\big(f^b\big)\big)\\&=\phi\big((1-b)V(\overline{f})+bV(\underline{f})\big)=\phi(1-b).\end{aligned}$$

Thus, $V_\rho=V=V_\tau$ so for all $F\in\mathcal{K}$,

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

By Lemma B.2, for all $b\in[0,1]$,

$$\begin{aligned}\int_{[0,1]}\big(F\cup f^b\big)_\rho(a)\,da&=\int_{[0,1]}\max\big\{F_\rho(a),f^b_\rho(a)\big\}\,da\\&=\int_{[0,1]}\max\big\{F_\rho(a),1_{[b,1]}(a)\big\}\,da\\&=\int_{[0,b]}F_\rho(a)\,da+1-b.\end{aligned}$$

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

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

By the Radon–Nikodym theorem, $F_\rho(a)=F_\tau(a)$ a.e. so by Part (2) of Lemma B.1, $F_\rho=F_\tau$ for all $F\in\mathcal{K}$. By Lemma B.3, $\rho=\tau$, as desired.

B.2. *Proof of Theorem 3*

Before proving Theorem 3, we first show two useful lemmas. The first shows that if some RCR is induced by a preference relation, then the best and worst acts of the RCR are also the best and worst acts of the preference relation.

LEMMA B.5: *Suppose $\overline{f}$ and $\underline{f}$ are the best and worst acts of $\rho_\succeq$ for some $\succeq$. Then $\overline{f}\succeq F\succeq\underline{f}$ for all $F\in\mathcal{K}$ and $\rho(\underline{f},\overline{f})=0$.*

PROOF: Let $\rho=\rho_\succeq$ so $\rho(f_a,F)=\frac{dW(F\cup f_a)}{da}$ for some $W:\mathcal{K}\to[0,1]$ that represent $\succeq$. Let $\overline{f}$ and $\underline{f}$ be the best and worst acts of ρ. Note that

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

If we let F be the singleton $\underline{f}$, then

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

where the last equality follows from the definition of $\underline{f}$. Since the range of W is [0, 1], this implies that $W(\underline{f} \cup \overline{f}) = 1$ and $W(\underline{f}) = 0$. Now, if we let F be the singleton $\overline{f}$, then

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

$$W(\overline{f}) = 1 + \int_{[0,1]} \rho(f_a, \overline{f})\, da.$$

Hence, $W(\overline{f}) = 1$ and $\int_{[0,1]} \rho(f_a, \overline{f})\, da = 0$. Thus, $\overline{f} \succeq F \succeq \underline{f}$ for all $F \in \mathcal{K}$. Note that this also proves that $\rho(\underline{f}, \overline{f}) = 0$ since otherwise, $\underline{f}$ and $\overline{f}$ would be tied, which would contradict the above equality. *Q.E.D.*

The next lemma shows that if a RCR ρ is standard, then it is induced by V_ρ.

LEMMA B.6: *If ρ is standard and $\rho(\underline{f}, \overline{f}) = 0$, then a.e.*,

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

PROOF: Let ρ be standard and $\rho(\underline{f}, \overline{f}) = 0$. We first show that a.e.,

$$1 = \rho(f^b, F) + F_\rho(b) = \rho(f^b, F) + \rho(F, f^b).$$

By Lemma A.2, this is violated iff $\rho(f^b, F) > 0$ 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 $\rho(\underline{f}, \overline{f}) = 0$. Thus, $\rho(f^b, F) + F_\rho(b) \neq 1$ at most a finite number of points as F is finite. The result follows.

Now, by Lemma B.2,

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

$$= \int_{[0,1-b]} F_\rho(a)\, da + b = \int_{[b,1]} F_\rho(1-a)\, da + b.$$

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

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

Thus, we have a.e.

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

as desired. *Q.E.D.*

We are now ready to prove Theorem 3. We wish to show the following are equivalent:

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

(2) $\succeq$ is dominant and $\rho_\succeq$ is represented by (μ, u).

First, suppose (1) is true so let u be normalized and

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

represent $\succeq$. Let $\rho = \rho_\succeq$ and $\overline{f}$ and $\underline{f}$ are the best and worst acts of ρ. Thus,

$$\rho(f_a, F) = \frac{dW(F \cup f_a)}{da},$$

where $W : \mathcal{K} \to [0,1]$ represents $\succeq$. Let τ be represented by (μ, u). Note that $\succeq$ is dominant follows from Theorem 1 of DLST. In order to prove (2), we will show that $\rho = \tau$.

First, we show that $q \cdot (u \circ \overline{f}) = 1$ and $q \cdot (u \circ \underline{f}) = 0$ μ-a.s. By Lemma B.5, we know that $\rho(\underline{f}, \overline{f}) = 0$ and $\overline{f} \succeq F \succeq \underline{f}$ for all $F \in \mathcal{K}$. Since V also represents $\succeq$ and u is normalized, this means that

$$0 = V(\underline{f}) = \int_{\Delta S} q \cdot (u \circ \underline{f}) \mu(dq),$$
$$1 = V(\overline{f}) = \int_{\Delta S} q \cdot (u \circ \overline{f}) \mu(dq),$$

which proves that $q \cdot (u \circ \overline{f}) = 1$ and $q \cdot (u \circ \underline{f}) = 0$ μ-a.s. Note that by Lemma A.5, we can let $\overline{f}$ and $\underline{f}$ be the best and worst acts of τ as well without loss of generality.

Next, we show that $W = V$. Since both W and V represent $\succeq$, $W = \phi \circ V$ for some monotonic $\phi : \mathbb{R} \to \mathbb{R}$. Hence

$$W(f_a) = \phi\big(V(f_a)\big) = \phi(a).$$

Since $\rho = \rho_{\succeq}$,

$$W(f_0 \cup f_a) - W(f_0) = \int_{[0,a]} \rho(f_b, f_0)\, db = a.$$

Since $q \cdot (u \circ f_0) = 0$ μ-a.s., $V(f_0 \cup f_a) = V(f_a)$ so

$$\phi(a) = W(f_a) = W(f_0 \cup f_a) = a.$$

Thus, $W = V$.

Note that by Lemma B.4, $V_\tau = V$. Thus, by Lemma B.6,

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

This implies that $F_\tau = F_\rho$ for all $F \in \mathcal{K}$ so by Lemma B.3, $\rho = \tau$. This proves (2).

Now, suppose (2) is true so $\succeq$ is dominant and $\rho = \rho_{\succeq}$ is represented by (μ, u). Let $\overline{f}$ and $\underline{f}$ be best and worst acts of ρ so

$$\rho(f_a, F) = \frac{dW(F \cup f_a)}{da},$$

where $W : \mathcal{K} \to [0,1]$ represents $\succeq$. Assume u is normalized so we have $q \cdot (u \circ \overline{f}) = 1$ and $q \cdot (u \circ \underline{f}) = 0$ μ-a.s. Note that this also implies that for all $s \in S$, $q_s u(\overline{f}(s)) = 1$ and $q_s u(\underline{f}(s)) = 0$ μ-a.s.

Define V as above, so by Lemma B.4, $V = V_\rho$. Since ρ is standard, by Lemma B.6,

$$\frac{dV(F \cup f_a)}{da} = \rho(f_a, F) = \frac{dW(F \cup f_a)}{da},$$

which implies that

$$W(F \cup \overline{f}) - W(F \cup \underline{f}) = V(F \cup \overline{f}) - V(F \cup \underline{f}).$$

By Lemma B.5, $\overline{f} \succeq F \succeq \underline{f}$ for all $F \in \mathcal{K}$. Let $s^* \in S$ be such that $\underline{f}(s^*) \succeq \underline{f}(s)$ for all $s \in S$. Consider the constant menu

$$F := \bigcup_{s \in S} \underline{f}(s).$$

Since $q_s u(\underline{f}(s)) = 0$ μ-a.s. for all $s \in S$,

$$W(F \cup \overline{f}) - W(F \cup \underline{f}) = V(F \cup \overline{f}) - V(F \cup \underline{f}) = 1.$$

This implies that $W(F \cup \underline{f}) = 0 = W(\underline{f})$ so $F \cup \underline{f} \sim \underline{f}$. Since $\underline{f}(s^*) \succeq h(s)$ for all $h \in F \cup \underline{f}$, dominance implies that

$$\underline{f}\big(s^*\big) \sim F \cup \underline{f} \sim \underline{f}.$$

Thus, $\underline{f}(s) \sim \underline{f}$ for all $s \in S$. By a symmetric argument, $\overline{f}(s) \sim \overline{f}$ for all $s \in S$. By dominance again, for any $F \in \mathcal{K}$, $F \cup \underline{f} \sim F$ and $F \cup \overline{f} \sim \overline{f}$. Thus,

$$\begin{aligned}1 - W(F) &= W(F \cup \overline{f}) - W(F \cup \underline{f}) \\ &= V(F \cup \overline{f}) - V(F \cup \underline{f}) = 1 - V(F),\end{aligned}$$

so $W = V$ and (1) is true.

B.3. *Proof of Theorem 4*

Let ρ and τ be represented by (μ, u) and (ν, u), respectively, and assume u is normalized. We will prove that the following are equivalent:

(1) μ is more informative than ν,
(2) $F_\tau \geq_{\text{SOSD}} F_\rho$ for all $F \in \mathcal{K}$,
(3) $F_\tau \geq_m F_\rho$ for all $F \in \mathcal{K}$.

We will show that (1) implies (2) implies (3) implies (1). First, suppose (1) is true so μ is more informative than ν. Fix $F \in \mathcal{K}$ and let $U := u \circ F$. Define

$$\psi_F(q) := 1 - \sup_{v \in U} q \cdot v.$$

Since support functions are convex, ψ_F is concave in $q \in \Delta S$.[36] Let $\phi : \mathbb{R} \to \mathbb{R}$ be increasing concave. Since u is normalized, by Lemma A.6,

$$\begin{aligned}\int_{[0,1]} \phi(a)\, dF_\rho(a) &= \int_{\Delta S} \phi\Big(1 - \sup_{v \in U} q \cdot v\Big)\mu(dq) \\ &= \int_{\Delta S} \phi \circ \psi_F(q)\mu(dq).\end{aligned}$$

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

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

[36]See Theorem 1.7.5 of Schneider (1993) for elementary properties of support functions.

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

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

so $\int_{[0,1]} \phi\, dF_\rho \leq \int_{[0,1]} \phi\, dF_\tau$ and $F_\tau \geq_{\text{SOSD}} F_\rho$ for all $F \in \mathcal{K}$. This proves (2).

Since $\geq_{\text{SOSD}}$ implies $\geq_m$, (2) implies (3) 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,

$$\int_{\Delta S}\left(\sup_{v\in U} q\cdot v\right)\mu(dq) \geq \int_{\Delta S}\left(\sup_{v\in U} q\cdot v\right)\nu(dq)$$

for all $U = u \circ F \subset [0,1]^S$. By Blackwell (1951, 1953), μ is more informative than ν, proving (1). Theorem 4 now follows immediately from the equivalence of (1) and (2).

B.4. *Proof of Theorem 5*

Before proving Theorem 5, we first show two useful lemmas. Recall that $\rho = (\rho_s)_{s\in S}$ and $\mu = (\mu_s)_{s\in S}$ are the state-dependent RCR and state-dependent signal distribution, respectively, and $\overline{\rho} = \sum_s r_s\rho_s$ and $\overline{\mu} = \sum_s r_s\mu_s$ are the corresponding unconditional RCR and unconditional signal distribution, respectively. We first show that if a state-dependent RCR has a state-dependent information representation, then its unconditional RCR has an unconditional information representation. Recall that $\underline{f}^s$ is the conditional worst act that coincides with the worst act if $s \in S$ occurs and the best act otherwise. Throughout this section, we will assume without loss of generality that u is normalized and that the best and worst acts are constant (see Lemma A.5).

LEMMA B.7: *If ρ is represented by (μ, u), then $\overline{\rho}$ is represented by $(\overline{\mu}, u)$. Moreover, for any $s \in S$, $q_s > 0$ $\overline{\mu}$-a.s. if and only if $q_s > 0$ $\mu_{s'}$-a.s. for all $s' \in S$.*

PROOF: Let ρ be represented by (μ, u). We first show that $\overline{\rho}$ is represented by $(\overline{\mu}, u)$. Note that by definition, ρ_s is represented by (μ_s, u_s) for all $s \in S$ and the measurable sets of $\rho_{s,F}$ and $\overline{\rho}_F$ coincide for each $F \in \mathcal{K}$. Since 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

$$\overline{\rho}_F(f) = \overline{\rho}_F(f_F) = \sum_s r_s \rho_{s,F}(f_F) = \sum_s r_s \mu_s(Q_{f,F}) = \overline{\mu}(Q_{f,F}),$$

so $\overline{\rho}$ is represented by $(\overline{\mu}, u)$.

We now prove the last part of the lemma. Let $s \in S$ and define

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

Since $\overline{f}$ is the best act, for any $s' \in S$, we have $\rho_{s'}(\overline{f}, \underline{f}^s) = 1 = \overline{\rho}(\overline{f}, \underline{f}^s)$ where the second equality follows from the fact that $\overline{\rho}$ is represented by $(\overline{\mu}, u)$. Note that if $\underline{f}^s$ is tied with $\overline{f}$, then $\mu_{s'}(Q) = \mu(Q) = 1$. On the other hand, if $\underline{f}^s$ is not tied with $\overline{f}$, then $\mu_{s'}(Q) = \mu(Q) = 0$. Hence, $q_s > 0$ $\overline{\mu}$-a.s. iff $q_s > 0$ $\mu_{s'}$-a.s. for all $s' \in S$, as desired. *Q.E.D.*

The next lemma provides a useful expression for the means of conditional test functions.

LEMMA B.8: *Let ρ_s be represented by (μ_s, u) and $\rho_s(\underline{f}^s, \overline{f}) = 0$. Then $q_s > 0$ μ_s-a.s. and for all $F \in \mathcal{K}_s$,*

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

PROOF: Note that

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

so $q_s > 0$ μ_s-a.s. Define

$$\psi_F^s(q) := \frac{r_s}{q_s}\Big(1 - \sup_{f \in F} q \cdot (u \circ f)\Big),$$

and let $\lambda_s^F := \mu_s \circ (\psi_F^s)^{-1}$ 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 since $q_s > 0$ μ_s-a.s., 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,

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

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

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

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 of λ_s^F. *Q.E.D.*

We are now ready to prove Theorem 5. Let ρ be represented by (μ, u). We wish to show the following are equivalent:

(1) $F_\rho^s =_m F_{\overline{\rho}}$ for all $F \in \mathcal{K}$,

(2) μ is well-calibrated.

First, suppose (1) is true so $F_\rho^s =_m F_{\overline{\rho}}$ for all $F \in \mathcal{K}$. Let S_+ be the set of states $s \in S$ such that $\rho_s(\underline{f}^s, \overline{f}) = 0$. Let $s \in S_+$, so by Lemma B.8, $q_s > 0$ μ_s-a.s. Define the measure ν_s on ΔS such that, for any measurable set $Q \subset \Delta S$,

$$\nu_s(Q) := \int_Q \frac{r_s}{q_s}\mu_s(dq).$$

We will show that $\overline{\mu} = \nu_s$ for all $s \in S_+$.

Since $F_\rho^s =_m F_{\overline{\rho}}$, by Lemmas A.6, B.7, and B.8, for all $F \in \mathcal{K}_s$

$$\int_{[0,1]} a\, dF_{\overline{\rho}}(a) = \int_{[0,r_s]} a\, dF_\rho^s(a),$$

$$\begin{aligned}\int_{\Delta S} \Big(1 - \sup_{f\in F} q\cdot(u\circ f)\Big)\overline{\mu}(dq) &= \int_{\Delta S} \frac{r_s}{q_s}\Big(1 - \sup_{f\in F} q\cdot(u\circ f)\Big)\mu_s(dq) \\ &= \int_{\Delta S} \Big(1 - \sup_{f\in F} q\cdot(u\circ f)\Big)\nu_s(dq).\end{aligned}$$

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

$$Q_a := \Big\{q \in \Delta S \Big| \sup_{f\in Ga\overline{f}} q\cdot(u\circ f) \geq q\cdot\big(u\circ \underline{f}^s\big)\Big\},$$

and note that

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

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

$$\begin{aligned}&\int_{\Delta S}\Big(1 - \sup_{f\in F_a} q\cdot(u\circ f)\Big)\overline{\mu}(dq) \\ &\quad = \int_{Q_a}\big(a\big(1 - h(u\circ G, q)\big)\big)\overline{\mu}(dq) + \int_{Q_a^c} q_s\overline{\mu}(dq),\end{aligned}$$

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

$$\begin{aligned}&\int_{Q_a}\big(1 - h(u\circ G, q)\big)\overline{\mu}(dq) + \int_{Q_a^c}\frac{q_s}{a}\overline{\mu}(dq) \\ &\quad = \int_{Q_a}\big(1 - h(u\circ G, q)\big)\nu_s(dq) + \int_{Q_a^c}\frac{q_s}{a}\nu_s(dq).\end{aligned}$$

We will now take limits as $a \to 0$. By Lemma B.7, $q_s > 0$ $\overline{\mu}$-a.s., so by dominated convergence,

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

$$= \int_{\Delta S} \big(1 - h(u \circ G, q)\big) \lim_{a \to 0} 1_{\{q_s \geq a(1-h(u \circ G, q))\} \cap \{q_s > 0\}}(q) \overline{\mu}(dq)$$
$$= \int_{\Delta S} \big(1 - h(u \circ G, q)\big) 1_{\{q_s > 0\}}(q) \overline{\mu}(dq)$$
$$= \int_{\Delta S} \big(1 - h(u \circ G, q)\big) \overline{\mu}(dq).$$

For $q \in Q_a^c$,

$$11 - q_s = q \cdot \big(u \circ f^s\big) > 1 - a\big(1 - h(u \circ G, q)\big),$$
$$\frac{q_s}{a} < 1 - h(u \circ G, q) \leq 1,$$

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

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

By a symmetric argument for ν_s, taking limits as $a \to 0$ yields that

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

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

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

By the uniqueness properties of the subjective learning representation (Theorem 1 of DLST), $\overline{\mu} = \nu_s$ for all $s \in S_+$. Hence,

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

for all measurable $Q \subset \Delta S$ and $s \in S_+$.

Finally, for $s \notin S_+$, $\rho_s(\underline{f}^s, \overline{f}) = 1$ so $q_s = 0$ μ_s-a.s. By Lemma B.7, $q_s = 0$ 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, since $r_s = \int_{\Delta S} q_s \overline{\mu}(dq)$ for all $s \in S_+$ from above,

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

This implies that $\sum_{s \notin S_+} r_s = 0$, which contradicts that r has full support on S. Thus, $S_+ = S$, so μ is well-calibrated and (2) is true.

Now, suppose (2) is true so μ is well-calibrated. Note that ties of ρ_s and $\overline{\rho}$ coincide by definition. As above, let S_+ denote the set of states $s \in S$ such that $\rho_s(\underline{f}^s, \overline{f}) = 0$. 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 $Q_0 := \{q \in \Delta S | \sum_{s \notin S_+} q_s = 0\}$ yields

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

a contradiction. Thus, $S_+ = S$. Let $F \in \mathcal{K}_s$ and $s \in S$. Since $\rho_s(\underline{f}^s, \overline{f}) = 0$, by Lemmas A.6 and B.8 and the fact that μ is well-calibrated,

$$\begin{aligned} \int_{[0,r_s]} a \, dF_\rho^s(a) &= \int_{\Delta S} \frac{r_s}{q_s} \Big(1 - \sup_{f \in F} q \cdot (u \circ f) \Big) \mu_s(dq) \\ &= \int_{\Delta S} \frac{r_s}{q_s} \Big(1 - \sup_{f \in F} q \cdot (u \circ f) \Big) \frac{q_s}{r_s} \overline{\mu}(dq) \\ &= \int_{\Delta S} \Big(1 - \sup_{f \in F} q \cdot (u \circ f) \Big) \overline{\mu}(dq) = \int_{[0,1]} a \, dF_{\overline{\rho}}(a). \end{aligned}$$

Thus, $F_\rho^s =_m F_{\overline{\rho}}$ so (1) is true.

REFERENCES

AHN, D., AND T. SARVER (2013): "Preference for Flexibility and Random Choice," *Econometrica*, 81 (1), 341–361. [1989,1998]

ANSCOMBE, F., AND R. AUMANN (1963): "A Definition of Subjective Probability," *The Annals of Mathematical Statistics*, 34 (1), 199–205. [1985,1986,1988,1990,1992,2003]

BERGEMANN, D., AND S. MORRIS (2016): "Bayes Correlated Equilibrium and the Comparison of Information Structures in Games," *Theoretical Economics*, 11, 487–522. [1989]

BILLINGSLEY, P. (1986): *Probability and Measure* (Second Ed.). New York: John Wiley and Sons, Inc. [2011]

BLACKWELL, D. (1951): "Comparison of Experiments," in *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*. Berkeley and Los Angeles: University of California Press, 93–102. [1987,2000,2001,2020]

——— (1953): "Equivalent Comparisons of Experiments," *The Annals of Mathematical Statistics*, 24 (2), 265–272. [1987,2000,2001,2020]

BLOCK, H., AND J. MARSCHAK (1960): "Random Orderings and Stochastic Theories of Response," in *Contributions to Probability and Statistics*, ed. by I. Olkin. Stanford, CA: Stanford University Press, 97–132. [1986,1993]

CAPLIN, A., AND M. DEAN (2015): "Revealed Preference, Rational Inattention, and Costly Information Acquisition," *American Economic Review*, 105 (7), 2183–2203. [1988,2004]

CAPLIN, A., AND D. MARTIN (2015): "A Testable Theory of Imperfect Perception," *Economic Journal*, 125 (582), 184–202. [1988,2004,2006]

CHAMBERS, C., AND N. LAMBERT (2014): "Dynamically Eliciting Unobservable Information," Report. [1989]

CHIAPPORI, P., AND B. SALANIÉ (2000): "Testing for Asymmetric Information in Insurance Markets," *The Journal of Political Economy*, 108 (1), 56–78. [1987,2004]

DAWID, A. (1982): "The Well-Calibrated Bayesian," *Journal of the American Statistical Association*, 77 (379), 605–610. [2004]

DEKEL, E., B. LIPMAN, AND A. RUSTICHINI (2001): "Representing Preferences With a Unique Subjective State Space," *Econometrica*, 69 (4), 891–934. [1989]

DILLENBERGER, D., J. LLERAS, P. SADOWSKI, AND N. TAKEOKA (2014): "A Theory of Subjective Learning," *Journal of Economic Theory*, 153, 287–312. [1986]

ELLIS, A. (2012): "Foundations for Optimal Inattention," Report. [1988]

FALMAGNE, J. (1978): "A Representation Theorem for Finite Random Scale Systems," *Journal of Mathematical Psychology*, 18, 52–72. [1986]

FINKELSTEIN, A., AND K. MCGARRY (2006): "Multiple Dimensions of Private Information: Evidence From the Long-Term Care Insurance Market," *American Economic Review*, 96 (4), 938–958. [1984,1987]

FINKELSTEIN, A., E. LUTTMER, AND M. NOTOWIDIGDO (2009): "Approaches to Estimating the Health State Dependence of the Utility Function," *American Economic Review: Papers and Proceedings*, 99 (2), 116–121. [1993]

FUDENBERG, D., AND T. STRZALECKI (2015): "Dynamic Logit With Choice Aversion," *Econometrica*, 83 (2), 651–691. [1989]

FUDENBERG, D., P. STRACK, AND T. STRZALECKI (2015): "Stochastic Choice and Optimal Sequential Sampling," Report. [1988]

GILOVICH, T., R. VALLONE, AND A. TVERSKY (1985): "The Hot Hand in Basketball: On the Misperception of Random Sequences," *Cognitive Psychology*, 17, 295–314. [2003]

GRANT, S., A. KAJII, AND B. POLAK (1998): "Intrinsic Preference for Information," *Journal of Economic Theory*, 83, 233–259. [1989]

——— (2000): "Temporal Resolution of Uncertainty and Recursive Non-Expected Utility Models," *Econometrica*, 68 (2), 425–434. [1989]

GUL, F., AND W. PESENDORFER (2006): "Random Expected Utility," *Econometrica*, 74, 121–146. [1986,1988,1989,1993]

GUL, F., P. NATENZON, AND W. PESENDORFER (2014): "Random Choice as Behavioral Optimization," *Econometrica*, 82 (5), 1873–1912. [1986]

HENDREN, N. (2013): "Private Information and Insurance Rejections," *Econometrica*, 81 (5), 1713–1762. [1984,1987]

KAMENICA, E., AND M. GENTZKOW (2011): "Bayesian Persuasion," *American Economic Review*, 101 (6), 2590–2615. [1989]
KARNI, E. (2007): "Foundations of Bayesian Theory," *Journal of Economic Theory*, 132, 167–188. [1993]
KARNI, E., AND Z. SAFRA (2016): "A Theory of Stochastic Choice Under Uncertainty," *Journal of Mathematical Economics*, 63, 164–173. [1989]
KARNI, E., D. SCHMEIDLER, AND K. VIND (1983): "On State Dependent Preferences and Subjective Probabilities," *Econometrica*, 51 (4), 1021–1031. [1993]
KREPS, D. (1979): "Representation Theorem for 'Preference for Flexibility'," *Econometrica*, 47, 565–578. [1989]
KREPS, D., AND E. PORTEUS (1978): "Temporal Resolution of Uncertainty and Dynamic Choice Theory," *Econometrica*, 46 (1), 185–200. [1989]
LU, J. (2016): "Supplement to 'Random Choice and Private Information'," *Econometrica Supplemental Material*, 84, http://dx.doi.org/10.3982/ECTA12821. [1988]
MATĚJKA, F., AND A. MCKAY (2015): "Rational Inattention to Discrete Choices: A New Foundation for the Multinomial Logit Model," *American Economic Review*, 105 (1), 272–298. [1988]
MCFADDEN, D. (1981): "Econometric Models of Probabilistic Choice," in *Structural Analysis of Discrete Data With Econometric Applications*, ed. by C. Manski and D. McFadden. Cambridge, MA: MIT Press, Chapter 5. [1986,1996,1998]
MCFADDEN, D., AND M. RICHTER (1990): "Stochastic Rationality and Revealed Stochastic Preference," in *Preferences, Uncertainty and Optimality*, ed. by J. S. Chipman, D. McFadden, and M. Richter. Boulder, CO: Westview Press, Inc., 161–186. [1986]
MÜLLER, A., AND D. STOYAN (2002): *Comparison Methods for Stochastic Models and Risks*. Chichester: John Wiley and Sons, Inc. [2002]
NATENZON, P. (2013): "Random Choice and Learning," Report. [1988]
RABIN, M. (2002): "Inference by Believers in the Law of Small Numbers," *The Quarterly Journal of Economics*, 117 (3), 775–816. [2003]
RAYO, L., AND I. SEGAL (2010): "Optimal Information Disclosure," *The Journal of Political Economy*, 118 (5), 949–987. [1989]
READ, D., AND G. LOEWENSTEIN (1995): "Diversification Bias: Explaining the Discrepancy in Variety Seeking Between Combined and Separated Choices," *Journal of Experimental Psychology: Applied*, 1, 34–49. [2003]
SAITO, K. (2015): "Preferences for Flexibility and Randomization Under Uncertainty," *American Economic Review*, 105, 1246–1271. [1989]
SAVAGE, J. (1954): *The Foundations of Statistics*. New York: John Wiley and Sons, Inc. [1984,1986, 2003]
SCHNEIDER, R. (1993): *Convex Bodies: The Brunn–Minkowski Theory*. Cambridge: Cambridge University Press. [2019]
STROTZ, R. (1955): "Myopia and Inconsistency in Dynamic Utility Maximization," *The Review of Economic Studies*, 23 (3), 165–180. [2002]
TRAIN, K. (2009): *Discrete Choice Methods With Simulation*. Cambridge: Cambridge University Press. [1986]

Dept. of Economics, UCLA, Los Angeles, CA 90095, U.S.A.; jay@econ.ucla.edu.

Co-editor Itzhak Gilboa handled this manuscript.

Manuscript received September, 2014; final revision received June, 2016.