American Economic Review 2019, 109(9): 3192–3228
https://doi.org/10.1257/aer.20171534

Bayesian Identification: A Theory for State-Dependent Utilities†

By Jay Lu*

We provide a revealed preference methodology for identifying beliefs and utilities that can vary across states. A notion of comparative informativeness is introduced that is weaker than the standard Blackwell ranking. We show that beliefs and state-dependent utilities can be identified using stochastic choice from two informational treatments, where one is strictly more informative than another. Moreover, if the signal structure is known, then stochastic choice from a single treatment is enough for identification. These results illustrate novel identification methodologies unique to stochastic choice. Applications include identifying biases in job hiring, loan approvals, and medical advice. (*JEL* D11, D82, D83, J23, M51)

How can we untangle beliefs from tastes in everyday decision making? A hiring manager may reject a minority worker either because he believes the worker will perform poorly or because he undervalues the output of the minority worker due to prejudice. A loan officer may approve an applicant either because he believes the applicant will not default or because he is excessively discounting the cost of default. A physician may recommend a clinical trial to a patient either because he believes the patient will qualify or because he is receiving kickbacks from the pharmaceutical company sponsoring the trial.

In each of these cases, the precise motivation behind the agent's decisions may be hard to discern from the perspective of an outside observer. Revealed preference offers a useful methodology for identifying an agent's beliefs and preferences based on choice data. Following Savage (1954), subjective expected utility theory provides a behavioral foundation for separately identifying beliefs and utilities. Nevertheless, when utilities are state-dependent, i.e., they may vary across different states, a well-known indeterminacy arises: one can always scale utilities up and beliefs down in such a way so that they are consistent with observed choices. This presents a notable challenge for identification.

* Department of Economics, UCLA, Los Angeles, CA 90095 (email: jay@econ.ucla.edu). Jeff Ely was the coeditor for this article. I am grateful to Simon Board, Edi Karni, Yusufcan Masatlioglu, Moritz Meyer-ter-Vehn, Jawaad Noor, Marek Pycia, Tomasz Sadzik and Pablo Schenone, as well as seminar audiences at Bocconi, Berkeley, SWET (UC Riverside), Michigan, SAET (IMPA), the Caltech Choice Conference, Yale, RUD (LBS), the Hitotsubashi Institute, Harvard-MIT, HKU, and NUS for helpful comments and suggestions. I also thank the coeditor and three anonymous referees for many excellent comments. The author declares that he has no relevant or material financial interests that relate to the research described in this paper.

† Go to https://doi.org/10.1257/aer.20171534 to visit the article page for additional materials and author disclosure statement.

This paper introduces a new revealed preference methodology for identifying beliefs and utilities even when utilities are state-dependent. We make use of a notion of comparative informativeness that is weaker than the standard Blackwell ranking. Our main result shows that beliefs and utilities can be identified using stochastic choices from two treatments, where one is strictly more informative than another. Moreover, if the outside observer knows the agent's signal structure, then identification can be achieved using stochastic choice from a single treatment. These results illustrate how different informational treatments can provide new identification methodologies unique to stochastic choice.

In Section I, we describe an application of our identification strategy to the classic problem of job hiring. A hiring manager (the agent) conducts interviews to screen candidates from a certain minority demographic (e.g., race, gender, or sexual orientation). Workers who are hired receive either a "high" or "low" rating depending on their performance in the first month. Ratings are indicative of future worker output. The manager wants to hire workers who are likely to receive a high rating but the manager's utility for output could depend on the worker's rating, i.e., utilities are *state-dependent*. For example, suppose the firm has a policy whereby low-rated workers are reassigned to a different division (e.g., back office) while high-rated workers remain and interact extensively with the manager. In this case, a manager who harbors prejudice may have a lower marginal utility of output for high-rated workers who remain than for low-rated workers who are reassigned. An outside observer (the analyst) observes that the manager rejects more minority applicants than nonminority applicants. Since she does not observe interviews or worker performances, the analyst cannot be sure if applicants are rejected because the manager is *biased* and has a lower utility for high-rated minority workers or because he is *unbiased* but believes minority applicants will perform poorly. In the context of discrimination, it is unclear if this is taste based, as in Becker (1957), or statistical, as in Arrow (1971) and Phelps (1972).

Following recent trends, suppose the firm decides to incorporate pre-employment testing in the screening process to improve hiring decisions.[1] This provides the analyst with additional data for identification. In particular, the analyst can now observe choice data from two informational treatments: (i) hiring demand from interviews only, and (ii) hiring demand from interviews plus job testing. Identifying whether the manager is biased or not can now be obtained by comparing the hiring demands from both treatments; in other words, a biased and an unbiased manager cannot have the same hiring demand in *both* treatments. To see why, suppose both managers have the same hiring demand in the first treatment with interviews only. Since the biased manager has a lower marginal utility for the output of high-rated workers, his applicants must be of higher quality if his hiring demand is the same as that of the unbiased manager. In the second treatment, managers receive additional information about applicants from the job test. This additional information means that hiring in the second treatment would more closely reflect the true quality of workers. As a result, demand will be higher for the biased manager than the unbiased one in the second treatment.

[1] See Autor and Scarborough (2008) and Hoffman, Kahn, and Li (2018) for empirical studies on the impact of introducing pre-employment testing on hiring decisions.

In general, since information affects beliefs but not tastes, the addition of more information would have different effects on hiring demand depending on the bias of the manager.

In the context of discrimination, traditional approaches for identifying beliefs and tastes involve collecting data on worker performance or conducting audit or correspondence studies.[2] This paper complements those approaches by providing an alternate identification methodology for situations where data on worker performance is scarce and audit studies are costly.[3] Moreover, since the methodology is based on revealed preference, it can be used in conjunction with these other approaches to detect when the agent may have incorrect beliefs (e.g., the agent exhibits *erroneous* statistical discrimination). There are many other economic situations where this methodology could be applied. For loan approvals, the informational treatment could be the introduction of automated credit scoring. For medical advice, the informational treatment could be the adoption of better diagnostic technology.

Section II introduces the general framework. An agent ("he") chooses an action from a menu of possible actions. There is a finite set of states and each action corresponds to a state-contingent payoff. Before choosing an action, the agent receives information (e.g., interviews in the hiring example) that informs him about the state. We model information canonically as a distribution over all possible posterior beliefs and call it a *posterior distribution*. The agent also has a *utility* function over payoffs that may vary depending on the state.

An analyst ("she") is an outside observer who wishes to identify the agent's posterior distribution and utility. We consider the case where the agent's signal structure is unknown or *private*, i.e., the analyst is completely ignorant about the agent's posterior distribution. This informational asymmetry implies that from the analyst's perspective, the agent's choices appear to be stochastic, that is, they consist of choice probabilities for each action in a menu. We first show that the classic issue of indeterminacy with state-dependent utilities extends to stochastic choice; information and utility cannot be identified given all relevant stochastic choice data.

We then introduce a notion of comparative informativeness that will be central to our analysis. A posterior distribution is *more informative* than another if ex post payoffs under the first are a mean-preserving spread of those under the second. Moreover, if this dispersion is strict for all nonconstant payoffs, then we say it is *strictly more informative* than the other. Compared to the standard Blackwell (1951, 1953) ordering, this notion of informativeness is weaker; with two states, they are equivalent.

Our main result, Theorem 1, shows that information and utility are uniquely identified given stochastic choice from two informational treatments, where one is strictly more informative than another. Moreover, binary choice data will

[2] See Altonji and Pierret (2001) and Benson, Board, and Meyer-ter-Vehn (2018) for examples of the former and Goldin and Rouse (2000) and Bertrand and Mullainathan (2004) for examples of the latter.

[3] Data on worker performance is usually measured indirectly via proxy variables and requires data collection over a long time horizon. For some drawbacks of audit studies, see Heckman (1998).

suffice for identification. In the hiring example, we demonstrate how the analyst can directly identify the manager's bias using hiring demands from both informational treatments. We also consider the case when the agent's signal structure is known or *public*, i.e., he knows the conditional signal distributions for each state but is unable to observe or verify actual signal realizations. In this case, Theorem 2 shows that stochastic choice from a single treatment can pin down beliefs and utilities as long as signals are strictly informative.

In Section III, we compare identification under stochastic choice with identification under two other forms of choice data that are commonly studied: menu choice and conditional choice. We first show that the menu choice analog of Theorem 1 does not hold: without additional data, information and utility cannot be identified given menu choices from two informational treatments, where one is more informative than another. This indeterminacy remains even if the analyst knows the signal structure and has access to conditional choices, that is, the choices of the agent conditional on the realization of the signal. This echoes a result from Karni, Schmeidler, and Vind (1983) and illustrates the identification properties unique to stochastic choice.

Our theoretical results provide guidance on data collection for an analyst interested in identification. Depending on whether the signal structure is public or private, stochastic choice data from either one or two informational treatments can be used to pin down beliefs and utilities. These results on belief identification also have interesting implications for inference. Assuming the agent's signal structure is known, an analyst can use the agent's pre-signal choices to infer his post-signal conditional choices without taking a stance on the agent's beliefs. The same cannot be said for stochastic choice; in this case, beliefs are uniquely identified and thus essential for inferring the agent's stochastic choice.

Section IV provides a full characterization of stochastic choice data from two informational treatments, where one is strictly more informative than another. We first present standard axioms to ensure that the stochastic choice data from each individual treatment is consistent with subjective expected utility maximization. Our next three axioms relate choice data across the two informational treatments. *Taste consistency* and *belief consistency* ensure that the agent's utility and prior remain unchanged across both treatments. The last axiom, *strict informativeness* ensures that the posterior distributions from the two treatments can be compared via the strict informativeness ranking. Theorem 3 states that these conditions are necessary and sufficient for our representation.

The practicality of our identification methodology depends on the feasibility of stochastic choice data. In the applications we considered above (job hiring, loan approvals, medical advice, etc.), stochastic choice corresponds to an agent's repeated decisions across a large subject pool and is readily available in many cases. In other applications (e.g., choosing health insurance), stochastic choice corresponds to repeated decisions over long periods of time so data feasibility would naturally be an issue. Nevertheless, our results offer a preliminary benchmark for how to proceed in those situations as well. For example, if we interpret stochastic choice as corresponding to unobserved heterogeneity in a population of agents, then with large populations and independent signals, our methodology could still be used to identify state-dependent utilities to some extent.

A. *Related Literature*

There is a long literature addressing the identification shortcomings of subjective expected utility.[4, 5] Luce and Krantz (1971), Fishburn (1973), and Karni (2007) all use enlarged choice spaces that include conditional actions to model state-dependent utilities.[6] On the other hand, Karni, Schmeidler, and Vind (1983) and Karni and Schmeidler (2016) use preferences conditional on hypothetical lotteries in order to achieve identification. Both approaches use primitives that do not manifest themselves in material choice behavior.[7] Karni (1993) does use a traditional primitive in a model with a state-dependent mapping that translates to state-dependent preferences via a normalized state-independent reference utility. Drèze (1987), Drèze and Rustichini (1999) and Karni (2006) use the fact that the probability of states are affected by the agent's actions to identify state-dependent utilities but this approach is limited to instances where moral hazard is present.

Recent papers have employed menu choice to address this issue.[8] Sadowski (2013) and Schenone (2016) assume a normalized state-independent utility across a subset of outcomes in order to achieve identification. Krishna and Sadowski (2014) and Dillenberger, Krishna, and Sadowski (2017) make use of the recursive structure in an infinite-period model for identification. Karni and Safra (2016) consider an additional preference relation on hypothetical mental state-act lotteries. Ahn and Sarver (2013) study both menu and stochastic choice in order to identify beliefs and utilities over a subjective state space. Their model differs from ours in two ways. First, since they work in the lottery setup, their state space is subjective where each state *is* a utility realization. In contrast, beliefs in our model are over an objective state space so information and learning can be modeled explicitly.[9] Second, menu choice is unnecessary for identification in our setup and the methodology that we introduce would not apply in their setting.[10]

I. Application: Job Hiring

In this section, we present a simple example to illustrate the main identification methodology. A firm is looking to hire workers from a pool of applicants. All workers who are hired receive a rating based on their performance at the end of the first month. Ratings are indicative of future worker output. Workers who

[4] In a well-known correspondence between Savage and Aumann in January 1971 (see Drèze 1987), Savage responds to Aumann's critique by redefining the state space in a way so that state dependence disappears. This reconstruction requires some bold assumptions, featuring scenarios where "the lady dies medically and yet is restored in good health to her husband." Even Savage admits that "Just how to do that seems to be an art for which I can give no prescription and for which it is perhaps unreasonable to expect one…."

[5] While we focus on state-dependent utilities, Dillenberger, Postlewaite, and Rozen (2017) show that an alternate form of indeterminacy arises if we allow for payoff-dependent beliefs.

[6] Skiadas (1997) adopts a similar approach by considering preferences over act-event pairs.

[7] See Karni (1993) and Karni and Mongin (2000) for more detailed discussions.

[8] In Section III, we compare identification under menu choice with identification under stochastic choice.

[9] This allows for richer utility comparisons. For example, we can talk about an agent who has state-dependent utilities but state-independent preferences.

[10] In a different context, Masatlioglu, Nakajima, and Ozdenoren (2014) also study a primitive consisting of choices from two treatments, where the ex ante choice does not involve menus.

receive a "high" rating will generate net output H while those who receive a "low" rating will generate net output L.[11] Both H and L are random variables with known distributions, and we assume $H \geq 0$ is sufficiently high so workers who receive a high rating are worth retaining. We also refer to a worker's rating as his *state* $\in \{high, low\}$.

In the beginning, a hiring manager (agent) screens applicants by conducting interviews. Hiring decisions are based on estimates of future output. Suppose after an interview, the manager has belief q that the applicant will receive a high rating if hired and $1 - q$ that the applicant will receive a low rating if hired. The firm is risk neutral and would like to hire an applicant if

$$qE[H] + (1-q)E[L] \geq 0.$$

The manager however may have preferences that are misaligned with those of the firm. In particular, we allow for the possibility that the manager may have *state-dependent* utilities, where his marginal utility for output is different depending on whether the worker receives a high or low rating. Letting u_{high} and u_{low} be the manager's utility in the high and low states, respectively, the manager will hire an applicant if

$$qE[u_{high}(H)] + (1-q)E[u_{low}(L)] \geq 0, \tag{1}$$

where the utility of not hiring is normalized to 0. Note that classic state-independent utility is obtained if we assume $u_{high} = u_{low}$. Importantly, we will see how this standard assumption is not without loss of generality.

What could be the cause of state-dependent preferences in this example? Suppose the firm has a policy, whereby workers with low ratings are reassigned to a different division (e.g., back office) while those with high ratings remain and interact extensively with the manager. If the applicant pool consists only of workers from a minority demographic (e.g., race, gender, or sexual orientation), then a manager who harbors prejudice may have a lower marginal utility of output for high-rated workers who remain with him than for low-rated workers who are reassigned. This prejudice against high-rated minority workers who remain would correspond to classic taste-based discrimination, as in Becker (1957).

An outside observer (analyst) wants to identify the manager's preferences from observable hiring data. When utilities are state-dependent, this presents a well-known challenge. Since interviews are private so beliefs q are not observable, the analyst can never be sure whether hiring decisions are driven by the manager's preferences or his beliefs. In the context of discrimination, it means that when the analyst observes fewer minority applicants being hired, she cannot be sure if this discrimination is the result of taste-based or statistical reasons.

For a simple illustration, suppose that $u_{low}(x) = x$ but $u_{high}(x) = \beta x$ for some parameter β. Here, $\beta < 1$ captures the manager's *bias* of underweighting the marginal output of a high-rated worker who remains with the manager.

[11] Both H and L are net of total costs (e.g., wages).

When $\beta = 1$, the manager is unbiased and has incentives that are completely aligned with those of the firm. We will show that hiring data generated by a biased manager ($\beta < 1$) can also be generated by an unbiased manager ($\beta = 1$) facing workers of lower quality, i.e., lower likelihood of getting a high rating. Given inequality (1), the *hiring demand*, i.e., the proportion of applicants hired, for the biased manager is given by

$$\Pr\left\{\frac{q}{1-q} \geq \beta^{-1} z\right\}, \tag{2}$$

where, $z := -E[L]/E[H]$ is the cost-benefit ratio of hiring.[12] In other words, an applicant is hired if the odds $q/(1-q)$ that the worker will receive a high rating is greater than the cutoff $\beta^{-1}z$. Now, suppose the applicants for the unbiased manager ($\beta = 1$) are lower quality so all odds are scaled down by β. Since the unbiased manager uses the cutoff z, his hiring demand is given by

$$\Pr\left\{\frac{\beta q}{1-q} \geq z\right\}. \tag{3}$$

Expressions (2) and (3) are the same, so the analyst cannot distinguish between the biased and unbiased managers. While the biased manager uses a higher cutoff than the unbiased manager, his applicants are also of higher quality so both managers hire the same number of workers. This indeterminacy is robust even if the analyst observes variation in the cost-benefit ratio z.[13]

We now show how informational treatments can provide a way to distinguish between the two managers. Consider a second treatment where the manager conducts pre-employment job testing in addition to interviewing. Suppose test scores $\theta \in [0.5, 1.5]$ are independent of interviews and distributed according to densities $J_{high}(\theta) = \theta$ and $J_{low}(\theta) = 2 - \theta$ in the high and low states respectively. While the analyst knows that the manager conducts both interviews and the test in this second treatment, she does not know the type of test being administered nor the actual test results. All information is still private to the manager.

Hiring data in this second treatment will now be different between the two managers. First, the biased manager will hire only if the odds of a high rating is above his cutoff $\beta^{-1}z$. Conditional on his posterior q after the interview, Bayes' rule implies that the biased manager will hire if the test score θ satisfies

$$\frac{q}{1-q}\frac{\theta}{2-\theta} \geq \beta^{-1}z.$$

In other words, $\theta \geq \underline{\theta}(q)$ for some cutoff $\underline{\theta}(q)$.[14] The hiring demand for the biased manager is given by

$$E\left[\int_{\underline{\theta}(q)}^{1.5} \big(q\theta + (1-q)(2-\theta)\big)\, d\theta\right]. \tag{4}$$

[12] We assume $E[L] < 0$ since otherwise the manager would hire all applicants.
[13] For instance, if there is variation in wages or the production technology.
[14] Explicitly, $\underline{\theta}(q) := 2(1-q)z/(\beta q + (1-q)z)$.

The unbiased manager will hire if

$$\frac{\beta q}{1-q}\frac{\theta}{2-\theta} \geq z,$$

which is characterized by the same cutoff $\underline{\theta}(q)$ as the biased manager. The final hiring demand for the unbiased manager is given by

$$E\left[\int_{\underline{\theta}(q)}^{1.5}\left(\frac{\beta q}{\beta q+(1-q)}\theta+\frac{1-q}{\beta q+(1-q)}(2-\theta)\right)d\theta\right]. \tag{5}$$

Since $\beta < 1$, expression (5) is less than expression (4). Hiring demand for the unbiased manager is lower than that of the biased manager in this second treatment.

In the first treatment with interviews only, both managers have the same hiring demand. This is because while the biased manager has a higher cutoff, his applicants are also more likely to receive a high rating. In the second treatment, managers receive additional information about applicants from the job test. This additional information means that hiring in the second treatment would more closely reflect the true quality of workers; demand will be higher for the biased manager than the unbiased one (see Figure 1 for an illustration of a specific case). In general, since information affects beliefs but not tastes, the addition of more information would have different effects on hiring demand depending on the bias of the manager. In the context of discrimination, it means that the analyst can distinguish between taste-based and statistical reasons for discrimination.

The discussion above provides a simple illustration of identifying state-dependent utilities using choice data from two informational treatments. Theorem 1 shows that such exercises can be performed as long as one treatment is strictly more informative than another. The fact that test scores were independent of interviews was unimportant; identification is also possible if managers were required to interview applicants twice and interviews are not perfectly correlated. Moreover, the second treatment could involve strictly less information such as the mandatory expungement of criminal records.[15]

Other Methodologies.—The usefulness of informational treatments for identification depends on the data available to the analyst. When there is variation in a state-independent outside option (e.g., varying the payoff from not hiring), then the analyst can identify the agent's beliefs and utilities from existing revealed preference methodologies.[16] On the other hand, when the outside option is fixed as in the example above or there is no state-independent outside option (e.g., choice between promoting a worker or not), then identification is not possible without additional data (see Lemma 1). In this case, information treatments can be used to obtain identification. Note that the use of stochastic choice data (e.g., fraction of applicants hired) is important as non-stochastic choice data may not be sufficient for identification.[17]

[15] See Agan and Starr (2018) for a recent study on such "ban the box" policies.
[16] For example, see Sadowski (2013), Schenone (2016), and also Lu (2016).
[17] See Section III for identification under menu and conditional choice.

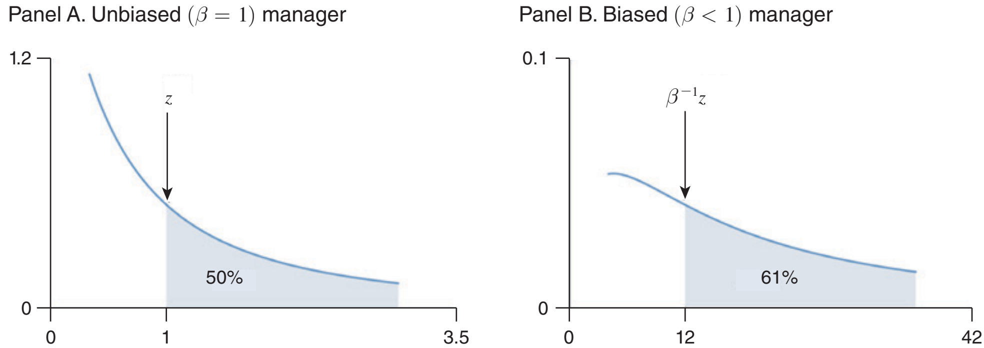

Figure 1. Distribution of Beliefs in Second Treatment

Notes: Panel A shows the distribution of odds of a high rating in the second treatment for the unbiased manager $(\beta = 1)$. The initial odds is 1 (prior belief of 50:50) and the cutoff is $z = 1$. Panel B shows the distribution of odds in the second treatment for the biased manager $(\beta = 1/12 < 1)$. The initial odds is higher at 12 but the cutoff is also higher at $\beta^{-1}z = 12$. Suppose there is no information in the first treatment and managers are initially indifferent between hiring and not hiring. Workers in the biased case however are 12 times more likely to receive a high rating. This gap is reflected in hiring in the second treatment: 50 percent versus 61 percent. Axes are normalized to highlight comparisons.

More generally, the methodology outlined here can also be used to complement existing approaches. For instance, suppose that in the hiring example, the analyst observes the realized performance ratings of workers. In this case, she can compare this data with the manager's elicited beliefs to determine whether the manager's beliefs are correct.[18]

Other Applications.—There are many other situations where this methodology could be applied. Consider loan approvals where the state is whether an applicant defaults or not. The analyst would like to identify whether a loan officer is approving more applicants because he believes his applicants are less likely to default or because he is excessively discounting the future cost of debt collection due to greater impatience. In this case, the informational treatment could be the introduction of automated credit scoring.

Finally, consider medical advice regarding clinical trials. The state is whether a patient qualifies for a clinical trial or not. A physician recommends patients to the trial without knowing for sure whether patients will qualify. The analyst would like to identify whether a physician is recommending more patients to the trial because he believes his patients are more likely to qualify or because the pharmaceutical company is providing kickbacks to the physician for recommending qualifying patients. In this case, the informational treatment could be the adoption of better diagnostic technology.[19]

[18] Our results apply as long as the manager's (possibly incorrect) prior is the average of his (possibly incorrect) posterior beliefs.

[19] Note that access to individual patient data would be a serious issue to the analyst due to privacy laws.

II. Bayesian Identification

A. *Model Setup*

We now present the general model. An agent ("he") faces a series of repeated decisions where he has to choose an action from a menu of possible actions. His payoff will ultimately depend on the realization of some state. In the motivating example (see Section I), the menu consists of two possible actions: hiring or not hiring an applicant. The state refers to a worker's rating in the review period: high or low.

Formally, let S denote a finite set of states, and let ΔS denote the set of all beliefs about S. We follow Anscombe and Aumann (1963) by modeling payoffs as risky prospects, i.e., lotteries in ΔX, where X is some finite set of outcomes.[20] Each *action* corresponds to a state-contingent payoff, i.e., a mapping $f : S \rightarrow \Delta X$. We can denote any action as a vector $f = (f_1, \ldots, f_n)$, where f_i is the payoff in state $s_i \in S$. For example, in the motivating example, s_1 and s_2 correspond to high and low ratings, respectively.[21] Hiring corresponds to the action $f = (H, L)$, where $H \in \Delta X$ and $L \in \Delta X$ are the distributions of worker output $x \in X$ in the high and low states, respectively. Since the firm receives nothing if the applicant is not hired, we can represent not hiring with the action $g = (0,0)$. A feasible (finite) set of actions A is called a *menu* and we let $\mathcal{A}$ denote the set of all menus.

An analyst ("she") has access to the agent's stochastic choice data, i.e., the choice frequency for each action in the menu. Let ΔA denote the set of all choice probabilities over actions. The agent's *stochastic choice* corresponds to a mapping $\rho : \mathcal{A} \rightarrow \Delta A$ such that ρ_F has full support in $F \in \mathcal{A}$. In other words, for any menu F and action $f \in F$, $\rho_F(f)$ is the probability that f is chosen in F. In the motivating example, if we let $F = \{f, g\}$ denote the menu consisting of the hiring and not hiring actions, then $\rho_F(f)$ is the hiring demand, i.e., the proportion of applicants hired. When there are only two actions, we employ the shorthand notation $\rho(f, g)$ to denote the choice probability of f over g.

The agent evaluates payoffs in each state according to a subjective expected utility. Given each state $s \in S$, let $u_s : \Delta X \rightarrow \mathbb{R}$ denote the agent's nonconstant von Neumann-Morgenstern (vNM) utility function in that state. A *utility* is thus a vector $u = (u_s)_{s \in S}$ of vNM utilities one for each state. We use the simplifying notation,

$$p \cdot (u \circ f) := \sum_{s \in S} p(s)\, u_s(f(s)),$$

to denote the state-dependent subjective expected utility of the agent for choosing the action $f \in A$ given his belief $p \in \Delta S$.

When information is private, the analyst is completely ignorant about the agent's signal structure. As a result, we model his signal structure canonically as

[20] The lottery structure is used primarily for within-state preference identification; it can be dispensed with if preferences are known and the only indeterminacy is utility comparisons across states as in the motivating example.

[21] Note that although the universal state space is the product of all workers' individual state spaces, since workers are evaluated independently and each worker's future output only depends on his rating, we can use $\{high, low\}$ as the canonical state space without loss.

a distribution μ over the belief space ΔS. We call μ a *posterior distribution.* The *prior* of μ is the average belief $\int_{\Delta S} q\, d\mu$. A belief $p \in \Delta S$ is *full-support* if it puts weight on all states.[22] A posterior distribution μ is *full-support* if it puts weight only on full-support beliefs. Let (μ, u) denote a posterior distribution μ with a full-support prior and utility u. We say ρ is represented by (μ, u) if it is consistent with the stochastic choice of an agent with a (possibly) state-dependent utility u and a posterior distribution μ.[23]

DEFINITION: *ρ is represented by (μ, u) if*

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

To illustrate this definition in the context of the motivating example, recall that the menu $F = \{f, g\}$ consists of hiring $f = (H, L)$ and not hiring $g = (0, 0)$. The manager's state-dependent utility is given by $u = (u_{high}, u_{low})$, where $u_{high}(x) = \beta x$ and $u_{low}(x) = x$. Recall that $\beta < 1$ captures the manager's bias of underweighting the marginal output of a high-rated worker who remains and interacts extensively with the manager. As a result, $u \circ f = (E[\beta H], E[L])$ and $u \circ g = (0, 0)$. Recall that $z := -E[L]/E[H]$ is the cost-benefit ratio of hiring. The manager's hiring demand is given by

$$D(z) := \rho_F(f) = \mu\big\{q \in [0,1] \,|\, q\beta - (1 - q)z \geq 0\big\}, \tag{6}$$

where μ is the posterior distribution for interviews. Equation (6) specifies the manager's hiring demand as a function of the cost-benefit ratio z and corresponds exactly to expressions (2) and (3). We say that ρ is represented by (μ, u).

A special case of the representation is of course when $\mu = \delta_p$ is degenerate. In this case, the stochastic choice ρ is deterministic, i.e., ρ only takes on values 0 or 1, and the representation reduces to the standard model of subjective expected utility with state-dependent utilities.

Given that stochastic choice consists of more than just ordinal choice data, one may posit that it may be possible to obtain identification. The following result answers in the negative; the classic problem of belief indeterminacy with state-dependent utilities model extends to stochastic choice.

LEMMA 1: *If ρ is represented by (μ, u), where μ is full-support, then for any full-support prior r, it is also represented by some (ν, v), where ν has prior r.*

PROOF:

For any $\beta \in \mathbb{R}^S_{++}$, define $\phi_\beta : \Delta S \to \Delta S$ such that

$$[\phi_\beta(q)](s) := \frac{q(s)\,\beta_s}{q \cdot \beta}. \tag{7}$$

[22] That is, $p(s) > 0$ for all $s \in S$.

[23] To deal with ties, we technically need an additional restriction ("regularity") on μ in order for this representation to be well-defined (see axiomatization in Section IV). Nevertheless, our main results hold regardless of whether μ is regular or not as long as we restrict the representation to menus without ties.

Since μ is full-support, we can choose β such that $\nu := \mu \circ \phi_\beta^{-1}$ has prior r. Define the utility v such that $v_s := \beta_s^{-1} u_s$ for all $s \in S$. Since ρ is represented by (μ, u), we have

$$\begin{aligned}\rho_F(f) &= \mu\{q \in \Delta S \,|\, q \cdot (u \circ f) \geq q \cdot (u \circ g) \ \forall g \in F\} \\ &= \mu\{q \in \Delta S \,|\, \phi_\beta(q) \cdot (v \circ f) \geq \phi_\beta(q) \cdot (v \circ g) \ \forall g \in F\} \\ &= \nu\{q \in \Delta S \,|\, q \cdot (v \circ f) \geq q \cdot (v \circ g) \ \forall g \in F\},\end{aligned}$$

so (ν, v) also represents ρ. ∎

In other words, for any stochastic choice that is represented by a full-support posterior distribution and state-dependent utility, there always exists an alternative posterior distribution and a state-dependent utility that represents the same stochastic choice. Moreover, this alternate posterior distribution can have any prior that shares the same support as the initial prior. In the context of the motivating example, Lemma 1 implies that the analyst cannot determine the bias β of the manager from hiring data. This is true even if she observes the entire demand curve $D(z)$.

Note that when the posterior distribution is not full support, Lemma 1 may not hold. For example, suppose μ is perfectly revealing about all states. By considering actions that are indicators for each state, the analyst can back out the agent's prior directly from the choice frequencies. Nevertheless, even in this case, the agent's state-dependent utility is still indeterminate; it can be arbitrarily scaled in every state. More generally, if posteriors have full-support on some subset of states, then the prior for those states cannot be identified. For instance, if μ corresponds to partitional information, then the priors about any two states that are in the same event of the partition are indeterminate. Whenever signals involve some amount of noise though, Lemma 1 applies.

B. *Comparative Informativeness*

We now show how the analyst can use stochastic choice data from two different informational treatments to obtain identification. First, we introduce a new notion of comparative informativeness. We say one signal is more informative than another if ex post payoffs under the first signal are a *mean-preserving spread* (*m.p.s.*) of those under the second.[24] For any payoff vector $w \in \mathbb{R}^S$, let μ^w denote the distribution of the ex post payoff $q \cdot w$ under the posterior distribution μ.

DEFINITION: *μ is more informative than ν if μ^w is an m.p.s. of ν^w for all $w \in \mathbb{R}^S$. Moreover, it is strictly more informative if $\mu^w \neq \nu^w$ whenever w is nonconstant.*

In other words, ex post payoffs are more dispersed if one signal is more informative than another. When there are only two states, this notion of informativeness is

[24] Formally, λ_1 is an m.p.s. of λ_2 if $\int_{\mathbb{R}} \varphi \, d\lambda_2 \geq \int_{\mathbb{R}} \varphi \, d\lambda_1$ for all concave φ.

equivalent to the commonly used Blackwell ordering. In general however, this comparison is weaker.[25] To see why, recall that one signal dominates another in the Blackwell ordering if posterior beliefs under the first are a mean-preserving spread of those under the second. Formally, a posterior distribution μ dominates ν in the Blackwell ordering if there exists a mean-preserving transition kernel[26] K such that for any measurable set of beliefs B,

$$\mu(B) = \int_{\Delta S} K(q,B)\, d\nu.$$

It is easy to see via Jensen's inequality that ex post payoffs under μ must be a mean-preserving spread of those under ν, so μ is more informative than ν.

Suppose there exists some set of positive ν measure such that $K(q,\cdot)$ has full-dimensional support. In this case, μ would be strictly more informative than ν. For instance, any μ with full-dimensional support is strictly more informative than δ_p, where p is its prior. Intuitively, being strictly more informative means that the additional information content is revealing about all states. In the motivating example, the addition of pre-employment testing provides strictly more information about applicants. To see this, recall that test scores $\theta \in [0.5, 1.5]$ are distributed according to $J_{high}(\theta) = \theta$ and $J_{low}(\theta) = 2 - \theta$ so they always provide some additional information about the applicant. If there are two rounds of interviews and the second round is not perfectly correlated with the first round, then this would also provide strictly more information.

An alternative characterization of the Blackwell ordering is that one signal dominates another in the Blackwell ordering if it provides higher ex ante payoffs (i.e., greater option value) for all menus. In contrast, our notion of informativeness only requires that one signal provides higher ex ante payoffs than another but only for *binary* menus. Lemma 3 in the Appendix formalizes this result. Regardless, it still implies that the two posterior distributions must have the same prior as required by Bayesian updating.

We now define stochastic choice data from two treatments, where one is more informative than another.

DEFINITION (Comparative Informativeness): (ρ_1, ρ_2) *is represented by* (μ_1, μ_2, u) *if*

(*i*) ρ_i *is represented by* (μ_i, u) *for* $i \in \{1,2\}$;

(*ii*) μ_2 *is more informative than* μ_1.

A comparative informativeness (CI) representation carries two implicit assumptions. First, signals from one treatment is more informative than another. This is satisfied if the agent is Bayesian and the second treatment involves receiving additional information. Second, utilities (possibly state-dependent) are

[25] It is equivalent to the linear convex ordering in statistics (see Lemma 3 in the Appendix). For an explicit example of how this is strictly weaker than the Blackwell ordering, see Elton and Hill (1992).

[26] That is, K is a mapping from ΔS to distributions on ΔS such that $\int_{\Delta S} r\, dK(q) = q$.

not affected by signals and are the same across the two treatments.[27] This is a reasonable assumption if the stochastic choice data is collected from the same agent and there is no reason to believe that tastes have changed across the two treatments. In Section IV, we show how to test these assumptions empirically.

We now present our main result. Although beliefs cannot be identified using stochastic choice data from a single treatment, they can be identified using stochastic choice data from two informational treatments, where one is strictly more informative than another. Notably, stochastic choice from all binary menus is sufficient for this exercise. This provides a foundation for beliefs given state-dependent utilities within the classic revealed preference methodology.

THEOREM 1: *Suppose* (ρ_1, ρ_2) *and* (τ_1, τ_2) *are represented by* (μ_1, μ_2, u) *and* (ν_1, ν_2, v)*, respectively. If* μ_2 *is strictly more informative than* μ_1*, then the following are equivalent:*

(*i*) $\rho_i(f, g) = \tau_i(f, g)$ *for all* f, g *and* $i \in \{1, 2\}$;

(*ii*) $(\mu_1, \mu_2) = (\nu_1, \nu_2)$ *and* $u = av + b$ *with* $a > 0$.

PROOF:

See Appendix Section B.

We now provide a sketch of the proof for two states. Suppose (ρ_1, ρ_2) is represented by (μ_1, μ_2, u). Consider constructing an alternate representation as follows. For any $\beta > 0$, define the adjusted belief $\phi(q)$ and utility v, where

$$\phi(q) := \frac{\beta q}{\beta q + (1 - q)},$$

$$v := (\beta^{-1} u_1, u_2).$$

By Lemma 1, ρ_i is also represented by $(\mu_i \circ \phi^{-1}, v)$ for $\{i \in 1, 2\}$.[28] Suppose that $\mu_1 \circ \phi^{-1}$ is also more informative than $\mu_2 \circ \phi^{-1}$. Since they must have the same prior,

$$\int_{\Delta S} \phi \, d\mu_1 = \int_{\Delta S} \phi \, d\mu_2.$$

Note that if $\beta < 1$, then ϕ is strictly convex so the left-hand side must be strictly greater than the right-hand side as μ_1 is strictly more informative than μ_2. The case for $\beta > 1$ is symmetric so the only possibility is $\beta = 1$ in which case ϕ is the identity and $v = u$. In other words, while we can find different priors that rationalize the stochastic choice from each treatment, strict informativeness means that

[27] This is similar in spirit but weaker than other assumptions in the literature on belief identification. See discussion in Section III.

[28] This covers all possible alternate representations (see the full proof in Appendix Section B).

we can only find one prior that can rationalize the stochastic choices from both treatments.[29]

To see how to apply Theorem 1 in practice, let us return to the motivating example. Recall equation (6) and note that the manager's demand $D_i(z)$ in treatment $i \in \{1,2\}$ can be written as

$$D_i(z) = \mu_i\left\{q \in [0,1] \,\middle|\, q \geq \frac{z}{\beta + z}\right\}.$$

Let F^{μ_i} be the c.d.f. of q under μ_i and $a = z/(\beta + z)$ so $F^{\mu_i}(a) = 1 - D_i(z)$. Integrating by parts, we can write the manager's prior p_i as

$$p_i = \int_0^1 q\, d\mu_i = 1 - \int_0^1 F^{\mu_i}(a)\, da.$$

Since $F^{\mu_i}(a) = 1 - D_i(z)$ and $a = z/(\beta + z)$, we have by a change of variables

$$(8) \qquad p_i = \int_0^1 D_i(z)\, da = \int_{\mathbb{R}_+} \frac{D_i(z)\beta}{(\beta + z)^2}\, dz.$$

If μ_2 is more informative than μ_1, then they must have the same prior, i.e., $p_1 = p_2$. Using equation (8), we thus have

$$(9) \qquad \int_{\mathbb{R}_+} \frac{D_1(z) - D_2(z)}{(\beta + z)^2}\, dz = 0.$$

Clearly, if $\mu_1 = \mu_2$, then $D_1 - D_2 = 0$ and β is indeterminate. However, as long as μ_2 is strictly more informative than μ_1, equation (9) allows us to solve for β directly from the manager's hiring demand.[30] Note that in this example, the value of the outside option (not hiring) is fixed and variation in the cost-benefit ratio z (e.g., variation in wages or the production technology) is enough to identify β.

When identification is possible, beliefs are more than a mere modeling device. Scaling utilities in different states has testable implications, allowing us to make counterfactual welfare statements such as "the agent's utility in one state is twice as large as that in another." Interestingly, the typical normalization assumed to obtain state-independent utilities under state-independent preferences is no longer vacuous but carries empirical implications.

We end this section with a couple of remarks. First, strict informativeness is important. Clearly, if $\mu_1 = \mu_2$, then Lemma 1 shows that identification is not possible. More generally, if the more informative treatment does not provide additional information that distinguishes between two states, then this will contribute

[29] Note that comparative informativeness naturally implies that the prior from both treatments must be the same. While necessary, this is not sufficient for identification.

[30] If no such β exists, then there does not exist any prior that is consistent with the manager's choices from both treatments. Appendix Section J provides a more general procedure for eliciting beliefs and utilities.

to another dimension of indeterminacy.[31] Section IV details the testable implications of strict informativeness. Second, preferences cannot be constant in any state. Here, as in classic models, null states continue to pose a problem. We leave the question of whether it is possible to enrich the data in a way that can accommodate null events and still allow for identification for future research. For both remarks, we can always consider a smaller state space where these issues do not arise.

C. *Public Signal Structures*

In this section, we consider the case where the analyst knows the agent's signal structure (i.e., the mapping from states to signals) but signal realizations are still private.[32] Here, we depart from the standard revealed preference methodology in considering non-choice data. To illustrate, consider the motivating example and suppose test scores are the manager's only source of information (no interviews). The departure is that the analyst now knows what kind of test is being used. For instance, suppose the type of test is decided by the Human Resources (HR) department or a third party (see Autor and Scarborough 2008) while the actual test is administered by the manager. Thus, although the analyst may know the manager's signal structure, test scores themselves are private and unverifiable.

Formally, let Θ denote some measurable space of signals and $\Delta\Theta$ denote the set of all signal distributions. A public signal structure is a vector of signal distributions $J = (J_s)_{s\in S}$, where $J_s \in \Delta\Theta$ is the distribution of signals conditional on the state $s \in S$. For instance, in the motivating example, $\Theta = [0.5, 1.5]$ is the range of test scores and θ and $2-\theta$ are the densities of scores for a worker with a high and low rating, respectively.

When is J consistent with some posterior distribution μ? Suppose P_θ is the posterior of the agent after observing the realization of the signal θ. Then μ should be the distribution of P_θ induced by the unconditional signal distribution $\sigma \in \Delta\Theta$. Moreover, P_θ and J should be related via Bayes' rule while the unconditional signal distribution σ should satisfy

$$\sigma = p \cdot J := \sum_s p(s) J_s,$$

where p is the prior of μ. We summarize this in the following definition.

DEFINITION: *J is consistent with μ if there is a $P : \Theta \to \Delta S$ such that $\mu = \sigma \circ P^{-1}$ and for all measurable $A \subset \Theta$ and $s \in S$,*

$$J_s(A) = \frac{\int_A P_\theta(s)\, d\sigma}{p(s)}, \tag{10}$$

where $\sigma = p \cdot J$ and p is the prior of μ.

[31] For example, let p be a full-support prior over three states $S = \{s_1, s_2, s_3\}$. Suppose μ involves learning about the event $\{s_1, s_2\}$ and $\nu = \delta_p$. Even though μ is more informative than ν, it is not strictly more informative so only $p(s_3)$ can be identified. On the other hand, if μ corresponds to full-information, then it is strictly more informative than ν.

[32] We thus make a distinction between a public *signal structure* and public *signals*.

If J is consistent with μ, then $\sigma \in \Delta\Theta$ is the distribution of signal realizations and $P : \Theta \rightarrow \Delta S$ is the agent's posterior mapping. Note that equation (10) relating J and P is exactly Bayes' rule. To illustrate, recall the motivating example, where J_{high} has density θ and J_{low} has density $2 - \theta$. Assume test scores are the manager's only source of information and his prior is $p = 0.5$. This means that scores are (unconditionally) distributed uniformly on $[0.5, 1.5]$. It is easy to check that $P_\theta(high) = 0.5\theta$ is the posterior mapping that satisfies Bayes' rule; in other words, a score of θ reveals that a worker will receive a high rating with probability 0.5θ. Since scores are uniformly distributed, the manager's posterior distribution μ is uniform on $[0.25, 0.75]$ and J is consistent with μ.

We now consider the case where the analyst has access to stochastic choice data and the signal structure is public. When is identification possible given a known signal structure? Of course, if J is completely uninformative, then this reduces to the standard subjective expected utility model without identification. We thus focus on the case where J is *strictly informative* in that J_s are all linearly independent.[33] Lemma 5 in the Appendix shows that this is equivalent to any consistent posterior distribution being strictly more informative than no information.

The following result shows that as long as the public signal structure is strictly informative, beliefs can be identified using stochastic choice from a *single* treatment. Furthermore, utilities can be identified for any two states s and t that are *imperfectly revealed*, that is, the supports of J_s and J_t have nonempty intersection.[34]

THEOREM 2: *Suppose ρ and τ are represented by (μ, u) and (ν, v), respectively, where J is consistent with μ and ν. If J is strictly informative, then the following are equivalent:*

(*i*) $\rho(f, g) = \tau(f, g)$ *for all* f, g;

(*ii*) $\mu = \nu$ *and* $(u_s, u_t) = a(v_s, v_t) + b$ *with* $a > 0$ *for all* s, t *imperfectly revealed.*

PROOF:

See Appendix Section D.

When agents with different priors use the same signal structure, their stochastic choices must be different. In particular, if two posterior distributions μ and ν are consistent with the same information structure, then they must be related as follows. Let p and r be their respective priors and $\beta_s = r(s)/p(s)$ for all $s \in S$. Define the measure μ_β such that for any measurable set of beliefs B,

$$\mu_\beta(B) = \int_B q \cdot \beta \, d\mu.$$

[33] That is, for any $w \in \mathbb{R}^S$, $w \cdot J = 0$ implies $w = 0$.

[34] In other words, the agent cannot perfectly distinguish between the two states.

If we let ϕ_β be defined as in the proof of Lemma 1, then $\nu = \mu_\beta \circ \phi_\beta^{-1}$.[35] The proof of Theorem 2 involves showing that this restriction is sufficient to pin down beliefs, that is, $\mu_\beta \circ \phi_\beta^{-1} = \mu \circ \phi_\beta^{-1}$ only when β is constant. While identification in Theorem 1 is due to the restriction from comparative informativeness, identification in Theorem 2 is due to the restriction from the known signal structure.

Note that if there are states that are perfectly revealed by the signal, then some utility indeterminacy remains. For example, if J is fully informative, then utility cannot be identified even though the prior can. As long as there is some noise (e.g., J_s all share the same support), then full identification of both beliefs and utilities is possible.

To see how to apply Theorem 2, let us return to the motivating example supposing test scores are the manager's sole source of information and the analyst knows the manager's signal structure J. Recall from equation (8) that the manager's prior p is related to his hiring demand $D(z)$ via

$$p = \int_{\mathbb{R}_+} \frac{D(z)\beta}{(\beta + z)^2} \, dz. \tag{11}$$

Given densities θ and $2 - \theta$ in the high and low states, respectively, the manager will hire an applicant if

$$\frac{p}{1-p} \frac{\theta}{2-\theta} \geq \beta^{-1} z,$$

or $\theta \geq \underline{\theta}(z) := 2(1-p)z/(\beta p + (1-p)z)$. Thus, hiring demand must satisfy

$$D(z) = \int_{\underline{\theta}(z)}^{1.5} \big(p\theta + (1-p)(2-\theta)\big) \, d\theta. \tag{12}$$

Equations (11) and (12) allow us to solve for β and p directly from the manager's hiring demand $D(z)$. Note that while we have assumed that the manager has correct beliefs, Theorem 2 may even apply in some cases where the manager's prior is incorrect.[36]

III. Comparisons with Menu and Conditional Choices

In this section, we compare identification under stochastic choice with identification under two other forms of choice data that are commonly studied: menu choice and conditional choice. Menu choice reflects the agent's prior before the realization of any signals. Conditional choice reflects the agent's posterior after the realization of a signal. Note that in both cases, at the moment of choice, the agent does not have any more information than the analyst. As a result, in contrast to stochastic choice, there is no informational asymmetry between the agent and the analyst.

[35] See Lemma 6 in the Appendix.

[36] For example, when the agent has an incorrect prior but also has incorrect beliefs about the signal structure so that the induced signal distribution is correct.

First, we model menu choice as a preference relation $\succeq$ over menus in $\mathcal{A}$. This reflects the agent's ex ante evaluation of different sets of feasible actions given his expectation about the informativeness of his signal. A natural representation for $\succeq$ is the following. Recall that $u \circ f \in \mathbb{R}^S$ denotes the utility vector for the action $f \in A$.

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

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

This is the subjective learning model of Dillenberger, Lleras, Sadowski, and Takeoka (2014) but allowing for state-dependent utilities.

Suppose the analyst observes menu choices from two informational treatments, where one is more informative than another. Formally, $(\succeq_1, \succeq_2)$ is represented by (μ_1, μ_2, u) if $\succeq_i$ is represented by (μ_i, u) for $i \in \{1,2\}$ and μ_2 is more informative than μ_1. Can the analyst then use $\succeq_1$ and $\succeq_2$ to identify the agent's beliefs and utilities as in the case of stochastic choice? The following result answers in the negative; the menu choice analog of Theorem 1 does not hold.

LEMMA 2: *If $(\succeq_1, \succeq_2)$ is represented by (μ_1, μ_2, u), then for any full-support prior r, it is also represented by some (ν_1, ν_2, v), where ν_1 and ν_2 have prior r.*

PROOF:

See Appendix Section E.

As in Lemma 1, the argument relies on constructing an alternate representation with prior r. As it turns out, this alternate representation corresponds exactly to an agent who uses the same signal structure but has prior r. Formally, set ν_i so that it has prior r and is consistent with the signal structure corresponding to μ_i for $i \in \{1,2\}$ over the canonical signal space $\Theta = \Delta S$. Define $\beta_s = r(s)/p(s) > 0$ so $\nu_i = \mu^i_\beta \circ \phi_\beta^{-1}$, where μ^i_β and ϕ_β are defined as in the discussion after Theorem 2. Letting $v_s := \beta_s^{-1} u_s$,

$$\begin{aligned}\int_{\Delta S} \sup_{f \in F} q \cdot (v \circ f)\, d\nu_i &= \int_{\Delta S} \sup_{f \in F} (\phi_\beta(q) \cdot (v \circ f))\, d\mu^i_\beta \\ &= \int_{\Delta S} \sup_{f \in F} \Big(\sum_s q(s) \beta_s v_s(f(s)) \Big)\, d\mu_i \\ &= \int_{\Delta S} \sup_{f \in F} q \cdot (u \circ f)\, d\mu_i,\end{aligned}$$

so $\succeq_i$ is represented by (ν_i, v). The rest of the proof of Lemma 2 involves checking that ν_2 is also more informative than ν_1.

In other words, menu choice naturally subsumes a relationship between posteriors beliefs and the prior that does not exist for stochastic choice. This is because indeterminacy with state-dependent utilities under menu choice is exactly the same as the indeterminacy of an agent facing the same signal structure but with an unknown prior. Fixing signal structures, the informativeness ranking holds regardless of the prior. This is why the restriction from informativeness has no bite under menu choice but does under stochastic choice. To see this formally, note the difference between the alternate distribution $\mu \circ \phi_{\beta}^{-1}$ under stochastic choice in Lemma 1 and the alternate distribution $\mu_{\beta} \circ \phi_{\beta}^{-1}$ under menu choice in Lemma 2. In this first case, the distribution of signals is unchanged while the posterior beliefs conditional on each signal are different. In the second case, both the distribution of signals and the posterior beliefs conditional on signals are different. Thus, identification is not possible with two treatments of menu choice but is with two treatments of stochastic choice, suggesting an inherent contrast between the two forms of choice data.[37]

What if the analyst also has access to conditional choice data? We model conditional choice as a preference relation $\succeq_{\theta}$ over actions in A for each signal realization θ in some signal space Θ. Let $\succeq_{\Theta} := (\succeq_{\theta})_{\theta \in \Theta}$ denote the family of conditional preferences. Suppose the analyst observes the following data: a menu preference $\succeq$, conditional preferences $\succeq_{\Theta}$ and the signal structure J. Note that even though the analyst knows the choices conditional on every signal realization, she does not know the distribution of signals nor the stochastic choice induced by the signals in contrast to Section IIC.

Can the analyst identify beliefs given all this data? The answer is no; Lemma 7 in the Appendix provides a formal proof. The reason is as follows. As Karni, Schmeidler, and Vind (1983) have shown, given a known signal structure, conditional preferences cannot identify beliefs. This is because the indeterminacy under conditional choice is exactly the same as the indeterminacy of an agent facing the same information but with an unknown prior. Since this indeterminacy is the same as that under menu choice, it must be that menu choice does not provide any additional identification power. All these forms of data are redundant, rendering identification impossible.

For an illustration, let us return to the motivating example supposing test scores are the manager's sole source of information. A biased manager with prior p will hire an applicant if

$$\frac{p}{1-p}\frac{\theta}{2-\theta} \geq \beta^{-1} z.$$

An unbiased manager with a lower prior $\phi = \frac{\beta p}{\beta p + (1-p)}$ will hire an applicant if

$$\frac{\beta p}{1-p}\frac{\theta}{2-\theta} = \frac{\phi}{1-\phi}\frac{\theta}{2-\theta} \geq z.$$

[37] In the special case when utilities are state independent, however, both menu and stochastic choice convey the same empirical content (see Lu 2016).

Table 1—Identification of State-Dependent Utilities

	Stochastic choice	Menu or conditional choice
Private signal structure (one strictly more informative)	2 treatments	More data needed for identification
Public signal structure (strictly informative)	1 treatment	

Thus, conditional preferences are the same for both managers. It is easy to check that menu preferences would also be the same by the same argument following Lemma 2. Thus, the analyst cannot distinguish between the biased and unbiased managers given menu and conditional preferences even if she knows the signal structure.

These results provide (guidelines for) data collection when identifying beliefs under state-dependent utilities. Table 1 summarizes our findings. With stochastic choice, beliefs can be identified using two informational treatments (where one is strictly more informative than another) or a single treatment if the signal structure is public (and information is strictly informative). In contrast, with menu and conditional choices, beliefs cannot be identified without additional data or assumptions. This highlights an intrinsic difference between stochastic choice data and menu and conditional choice data. Note that with a public signal structure and two treatments of stochastic choice, there is over identification; the analyst can then check for internal inconsistencies on the part of the agent.

This does not necessarily mean that it is impossible to identify beliefs given menu or conditional choices. For example, identification with menu choice is possible with additional assumptions (e.g., state independence for a subset of prizes).[38] Which type of choice data is more natural depends on the particular environment. In the applications we considered above (job hirings, loan approvals, medical advice, etc.), stochastic choice corresponds to repeated decisions across a large subject pool and is readily available in many cases. In applications involving consumption savings problems or portfolio choice, menu choice may be more readily accessible. Our results thus contribute to our understanding of identification given various data and modeling assumptions.

Policy Implications.—Suppose a policymaker knows the signal structure and observes an agent's ex ante choices before any information is revealed. She is confronted with two types of predictive questions: (i) which action will the agent choose given a signal realization? and (ii) what probability will the agent choose each action? Our results imply the first question can be answered but not the second. This is because in the case of the former, she can simply choose any arbitrary prior and use that to infer the agent's conditional preferences; our analysis guarantees that her predictions will be correct.[39] On the other hand, Theorem 1

[38] See Sadowski (2013) and Schenone (2016). This is similar in spirit to the assumption that signals only affect utilities via beliefs but stronger. For instance, Lemma 2 would not hold if we replaced it with this assumption.

[39] Suppose the predicted conditional preferences $\succeq_\Theta$ are based on some prior r while the true prior is in fact p. Note that we can create an alternate representation consistent with the true prior p and all the data $(\succeq, \succeq_\Theta, J)$. Since J and p uniquely determine $\succeq_\Theta$, it must coincide with the true conditional preferences.

implies that priors are uniquely identified given ex ante and stochastic choice so the policy maker cannot make arbitrary assumptions about the agent's prior. Thus, the second question cannot be answered without additional data.

Belief identification also has important welfare consequences. In Gilboa, Samuelson, and Schmeidler (2014) and Brunnermeier, Simsek, and Xiong (2014), Pareto arguments for trade are less appealing when they are motivated by differences in beliefs as opposed to differences in preferences. Our theory thus provides a revealed preference methodology that could prove useful for conducting such analysis for regulatory evaluations.

IV. Characterization

In this section, we provide testable implications of our model where one treatment is strictly more informative than another. First, as with any model of random utility, we need to address ties. Following Lu (2016), we model ties by relaxing the restriction that all choice probabilities have to be fully described. For example, if two actions f and g are tied, then the stochastic choice does not specify individual choice probabilities for either f or g. Formally, we model this as non-measurability and let ρ denote the corresponding outer measure without loss of generality.[40] With this definition, $\rho(f,g) = \rho(g,f) = 1$ whenever f and g are tied. Let $\mathcal{A}_0$ denote the set of menus without ties.

We now introduce standard conditions for random expected utility maximization. Endow $\mathcal{A}$ with the standard Hausdorff metric and ΔA with the topology of weak convergence. Also, let ext$F$ denote the extreme actions of F. For notational brevity, we omit universal quantifiers below.

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

AXIOM 1.2 (Linearity): *$\rho_F(f) = \rho_{aF+(1-a)g}(af + (1-a)g)$ for $a > 0$.*

AXIOM 1.3 (Continuity): *$\rho : \mathcal{A}_0 \rightarrow \Delta A$ is continuous.*

AXIOM 1.4 (Extremeness): *$\rho_F(extF) = 1$.*

Axioms 1.1–1.4 characterize any random expected utility model (Gul and Pesendorfer 2006). The next axiom deals with state-dependent utilities. For actions $f, g \in A$ and state $s \in S$, define $fsg \in A$ as the action that pays according to f in state s and g otherwise. Formally, $(fsg)(s) = f(s)$ and $(fsg)(t) = g(t)$ for all $t \neq s$. Also define $Fsg = \{fsg : f \in F\}$.

AXIOM 1.5 (Static Tastes): *$\rho_{Fsg}(fsg) \in \{0,1\}$.*

Axiom 1.5 ensures that signals do not affect the agent's preferences. In other words, choice over actions that differ only in a single state must be deterministic.

[40] For additional details, see Lu (2016).

It provides the empirical content for the fact that information does not affect tastes in the CI representation.[41] The next axiom ensures that the agent's preferences are not degenerate in every state; in other words, in every state, not all actions are tied.

AXIOM 1.6 (Non-Degeneracy): *For every* $s \in S$, $\rho_{Fsg}(fsg) < 1$ *for some* $f \in F$ *and* $g \in A$.

Collectively, we call the above axioms Axiom 1. Call a posterior distribution μ *regular* if μ^w is either degenerate or has no mass points for all $w \in \mathbb{R}^S$. This generalizes the related notion from Gul and Pesendorfer (2006) and is used to deal with ties in the random utility.[42] Call (μ, u) regular if μ is regular. The following result, which extends Lu (2016) to state-dependent utilities, shows that Axiom 1 characterizes regular representations.

PROPOSITION 1: ρ *satisfies Axiom* 1 *if and only if it is represented by a regular* (μ, u).

PROOF:

See Appendix Section G.

We now introduce the conditions that relate stochastic choice from the two informational treatments. The first condition below ensures that the agent's preferences are the same in both treatments.

AXIOM 2 (Taste Consistency): $\rho_{1,Fsg}(fsg) = \rho_{2,Fsg}(fsg)$.

Let $\underline{h} \in A$ denote some worst action, that is $\rho_i(f, \underline{h}) = 1$ for all $f \in A$, which is guaranteed by Axiom 1. Note that Axiom 2 means that treatment i does not matter. We can now define dominance as follows. An action f *dominates* another g if $\rho_i(fs\underline{h}, gs\underline{h}) = 1$ for all $s \in S$ and *strictly dominates* if $\rho_i(gs\underline{h}, fs\underline{h}) = 0$ for all $s \in S$ as well. Let $f \geq g$ and $f > g$ denote dominance and strict dominance ,respectively.

The next two conditions deal with the informational comparisons between the two treatments. First, we introduce a technical tool that will help our analysis. If $h > \underline{h}$, let $(\underline{h}, h)$ denote the set of actions f such that $h > f > \underline{h}$.

DEFINITION: *For* $h > \underline{h}$, *define for all* $f \in (\underline{h}, h)$,

$$f_i^h(\alpha) := \rho_i(f, \alpha h + (1 - \alpha)\underline{h}).$$

[41] One could scale both beliefs and utilities stochastically in a delicate manner so that even though preferences are not affected by signals, utilities may be. This could allow for a reinterpretation of the model where signals affect both beliefs and tastes. Nevertheless, in most applications where the agent's tastes are stable, this is a reasonable assumption. One could also circumvent this issue by introducing some additional discipline in the data (e.g., some independence of belief and taste shocks) or enriching the primitive (e.g., public signal structure).

[42] Gul and Pesendorfer's (2006) notion of regularity implies that the posterior distribution μ is always strictly informative while this definition of regularity does not.

The function f_i^h traces out the utility distribution of the action f with respect to h.[43] Lemma 8 in the Appendix shows that any action $h > \underline{h}$ induces a c.d.f. f_i^h on the unit interval. We call an action *calibrating* if the induced c.d.f.s have the same mean across both treatments.

DEFINITION: *h is calibrating if f_1^h and f_2^h have the same mean for all* $f \in (\underline{h}, h)$.

If the agent has the same prior across both treatments (i.e., he is Bayesian), then there must exist a calibrating action. This motivates our next axiom.

AXIOM 3 (Belief Consistency): *There exists a calibrating action.*

Calibrating actions are useful in that they yield the same utility to the agent in every state. In the case of state-independence for instance, constant acts are calibrating. As a result, they are useful for measuring and comparing the agent's information across the two treatments. In Appendix Section J, we introduce a procedure for eliciting a calibrating action which can then be used to extract beliefs and utilities. That analysis also suggests an alternate characterization of our model.

The final condition characterizes strict informativeness. We say two actions f and g are *incomparable* if neither $f \geq g$ nor $g \geq f$.

AXIOM 4 (Informativeness): *If h is calibrating, then f_2^h is an m.p.s. of f_1^h for all* $f \in (\underline{h}, h)$. *Moreover,* $f_1^h \neq f_2^h$ *whenever f and* $(1/2)\underline{h} + (1/2)h$ *are incomparable.*

We now state our main representation result. Call (μ_1, μ_2, u) regular if both μ_1 and μ_2 are regular.

THEOREM 3: (ρ_1, ρ_2) *satisfies Axioms* 1–4 *if and only if it is represented by a regular* (μ_1, μ_2, u)*, where μ_2 is strictly more informative than μ_1.*

PROOF.
See Appendix Section I.

For a sketch of the proof, note that Proposition 1 ensures the stochastic choice from each treatment can be represented by some posterior distribution and some state-dependent utility. Axiom 2 ensures that utilities across treatments are the same while Axiom 3 ensures that posterior distributions across treatments have the same prior. Finally, Axiom 4 is the behavioral characterization of the strictly more informative ordering. It relates dispersion in ex post payoffs with dispersion in stochastic choice.

As an illustration, consider the special case, where ρ_1 is deterministic or $\mu_1 = \delta_p$. Axiom 3 ensures that μ_2 has prior p which precisely characterizes

[43] It is the state-dependent version of test functions from Lu (2016).

whether the agent is Bayesian. On the other hand, Axiom 4 ensures that μ_2 is strictly informative so identification can be achieved.[44]

Appendix

A. *Characterizations of Informativeness*

We first show the equivalence of more informativeness, the linear convex stochastic order and higher ex ante payoffs for all binary menus.

LEMMA 3: *The following are equivalent:*

(*i*) μ *is* (*strictly*) *more informative than* ν;

(*ii*) *for all* $w \in \mathbb{R}^S$ *and convex* φ,

$$\int_{\Delta S} \varphi(q \cdot w)\, d\mu \;\geq\; \int_{\Delta S} \varphi(q \cdot w)\, d\nu \tag{13}$$

(*moreover*, *the inequality is strict whenever* φ *is strictly convex and* w *is not constant*);

(*iii*) *for all* $w_1, w_2 \in \mathbb{R}^S$,

$$\int_{\Delta S} \max_i (q \cdot w_i)\, d\mu \;\geq\; \int_{\Delta S} \max_i (q \cdot w_i)\, d\nu \tag{14}$$

(*moreover*, *for all nonconstant* w_1, *the inequality is strict for some constant* w_2).

PROOF:

We first consider the weak case and prove the equivalence of (i), (ii), and (iii). Note that the equivalence of (i) and (ii) follows immediately from a change of variables, so we will focus on their equivalence with (iii). First, note that (ii) implies (iii) trivially as $\max\{q \cdot w_1, q \cdot w_2\}$ is a convex function of q. Now suppose (iii) is true, so for all $\alpha, \lambda \in \mathbb{R}$ and $w \in \mathbb{R}^S$,

$$\int_{\Delta S} \max\{\alpha, q \cdot \lambda w\}\, d\mu \;\geq\; \int_{\Delta S} \max\{\alpha, q \cdot \lambda w\}\, d\nu.$$

Since $\varphi(x) = \max\{\alpha, \lambda x\}$ generates all convex functions, this proves (ii). Thus, all three are equivalent for the weak case.

We now deal with the strict case. We will show that (i) implies (ii) implies (iii) implies (i). First, suppose (i) holds with strictness, so μ^w dominates ν^w in

[44] If we dropped the second clause of Axiom 4, then μ_2 would only be weakly more informative than μ_1.

the convex order. From Blackwell (1951, 1953), this means that there exists a mean-preserving transition kernel K on $\mathbb{R}$ such that for any measurable $B \subset \mathbb{R}$,

$$\mu^w(B) = \int_{\mathbb{R}} K(x,B)\, d\nu^w.$$

Suppose there exists some strictly convex φ and nonconstant w such that inequality (13) holds with equality. Thus,

$$\int_{\mathbb{R}} \varphi\, d\nu^w = \int_{\mathbb{R}} \varphi\, d\mu^w = \int_{\mathbb{R}} \left(\int_{\mathbb{R}} \varphi\, dK(x) \right) d\nu^w.$$

This implies that $\varphi(x) = \int_{\mathbb{R}} \varphi\, dK(x)$ ν^w-a.s.. Since φ is strictly convex, this means that $K(x) = \delta_x$ ν^w-a.s. so $\mu^w = \nu^w$ yielding a contradiction. Thus, (ii) must hold with strictness. To see (iii), suppose there is some nonconstant w_1 such that inequality (14) holds with equality for all constant w_1. Thus, for all $\alpha, \lambda \in \mathbb{R}$,

$$\int_{\Delta S} \max\{q \cdot \lambda w_1, \alpha\}\, d\mu = \int_{\Delta S} \max\{q \cdot \lambda w_1, \alpha\} d\nu.$$

Since $\varphi(x) = \max\{\alpha, \lambda x\}$ generates all convex functions and w_1 is nonconstant, this contradicts the strictness of (ii). Thus, (iii) must hold with strictness. Finally, if (iii) holds with strictness, then $\mu^w \neq \nu^w$ for all nonconstant w proving (i) must hold with strictness. This completes the proof. ∎

B. *Proof of Theorem 1*

First, we show that if two stochastic choices agree over all binary choices, then their vNM utilities in each state must be affine transformations of each other.

LEMMA 4: *Suppose ρ and τ are represented by (μ, u) and (ν, v), respectively. If $\rho(f,g) = \tau(f,g)$ for all $f, g \in A$, then $u_s = \beta_s v_s + \alpha_s$ for $\beta_s > 0$.*

PROOF:

Fix some state $s \in S$ and let $u_s(p) \geq u_s(q)$ for some $p, q \in \Delta X$. Consider an action $f \in A$ such that $f(s) = p$ and $f(t) = q$ for $t \neq s$. Also, let $g \in A$ be such that $g(t) = q$ for all $t \in S$. Note that

$$\begin{aligned}
\rho(f,g) &= \mu\{q \in \Delta S \,|\, q \cdot (u \circ f) \geq q \cdot (u \circ g)\} \\
&= \mu\left\{ q \in \Delta S \,\middle|\, q(s)\, u_s(p) + \sum_{t \neq s} q(t)\, u_t(q) \geq q(s)\, u_s(q) + \sum_{t \neq s} q(t)\, u_t(q) \right\} \\
&= \mu\{r \in \Delta S \,|\, q(s)\, u_s(p) \geq q(s)\, u_s(q)\} = 1.
\end{aligned}$$

Since $\tau(f,g) = \rho(f,g)$, this implies that

$$\nu\{r \in \Delta S | q(s)\, v_s(p) \geq q(s)\, v_s(q)\} = 1.$$

By the full-support prior assumption, we know that $q(s) > 0$ with some strictly positive ν-probability. Thus, $v_s(p) \geq v_s(q)$, so by symmetric reasoning, we have $u_s(p) \geq u_s(q)$ if and only if $v_s(p) \geq v_s(q)$ for all $p, q \in \Delta X$. This implies that $u_s = \beta_s v_s + \alpha_s$ for all $s \in S$. Since u_s is nonconstant, this means that $\beta_s > 0$ for all $s \in S$. ∎

We now prove Theorem 1. Suppose (ρ_1, ρ_2) and (τ_1, τ_2) are represented by (μ_1, μ_2, u) and (ν_1, ν_2, v), respectively, and μ_2 is strictly more informative than μ_1. To show necessity, note that if $(\mu_1, \mu_2, u) = (\nu_1, \nu_2, av + b)$ for $a > 0$, then $\rho_i(f, g) = \tau_i(f, g)$ for all $f, g \in A$ and $i \in \{1, 2\}$ immediately from the representation. To show sufficiency, suppose that $\rho_i(f, g) = \tau_i(f, g)$ for all $f, g \in A$ and $i \in \{1, 2\}$. We will show that this implies that $(\mu_1, \mu_2, u) = (\nu_1, \nu_2, av + b)$ for $a > 0$.

By Lemma 4, we know that $u_s = \beta_s v_s + \alpha_s$ for $\beta_s > 0$. We will now show that β is constant. Since $\beta_s > 0$ for all $s \in S$, we can define ϕ_β as in equation (7) from the proof of Lemma 1, so by the same argument, ρ_i is represented by $(\mu_i \circ \phi_\beta^{-1}, v)$. Thus,

$$(\mu_i \circ \phi_\beta^{-1})\{q \in \Delta S \,|\, q \cdot (v \circ f) \geq q \cdot (v \circ g)\}$$

$$= \rho_i(f, g) = \tau_i(f, g) = \nu_i\{q \in \Delta S \,|\, q \cdot (v \circ f) \geq q \cdot (v \circ g)\}.$$

Since this is true for all $f, g \in A$, by the Cramér-Wold theorem, $\mu_i \circ \phi_\beta^{-1} = \nu_i$ for $i \in \{1, 2\}$.

Since ν_2 is more informative than ν_1, they must have the same prior. Thus,

$$\int_{\Delta S} \phi_\beta d\mu_1 = \int_{\Delta S} q \, d\nu_1 = \int_{\Delta S} q \, d\nu_2 = \int_{\Delta S} \phi_\beta \, d\mu_2.$$

This implies that for every $t \in S$,

$$\int_{\Delta S} \frac{q(t)}{\sum_s q(s) \beta_s} d\mu_1 = \int_{\Delta S} \frac{q(t)}{\sum_s q(s) \beta_s} d\mu_2.$$

Summing over all $t \in S$, we obtain

$$\int_{\Delta S} (q \cdot \beta)^{-1} d\mu_1 = \int_{\Delta S} (q \cdot \beta)^{-1} d\mu_2.$$

Since the inverse function is strictly convex and μ_2 is strictly more informative than μ_1, Lemma 3 implies that β must be constant. Thus, we have $\phi_\beta(q) = q$ so $\mu_i = \nu_i$ for $i \in \{1, 2\}$ as desired. ∎

C. *Public Signal Structure*

This section provides a couple results for the case when the signal structure is public. The first shows that a signal structure J is linearly independent if and only if any consistent posterior distribution is strictly more informative than no information.

LEMMA 5: *Let μ be consistent with J and have a full-support prior p. Then the following are equivalent:*

(*i*) *J_s are linearly independent*;

(*ii*) *μ^w is degenerate if and only if $w \in \mathbb{R}^S$ is constant*;

(*iii*) *μ is strictly more informative than δ_p.*

PROOF:

Since J is consistent with μ, let $\sigma \in \Delta\Theta$ be the unconditional signal distribution and $P : \Theta \to \Delta S$ be the posterior mapping. We will show that (i) implies (ii) implies (iii) implies (i). First, suppose J_s are linearly independent. Note that if $w \in \mathbb{R}^S$ is constant, then clearly μ^w is degenerate. Now, consider some w such that μ^w is degenerate so $q \cdot w = \lambda$ μ-a.s.. Let $\alpha := w - \lambda$ so $q \cdot \alpha = 0$ μ-a.s. or by a change of variables, $P_\theta \cdot \alpha = 0$ σ-a.s.. By Bayes' rule from equation (10), this means that for any measurable $R \subset \Theta$,

$$\sum_s \alpha_s p(s) J_s(R) = \sum_s \alpha_s \int_R P_\theta(s)\, d\sigma = \int_R (P_\theta \cdot \alpha)\, d\sigma = 0.$$

Since J_s are linearly independent and $p(s) > 0$ for all $s \in S$, this implies that $\alpha = 0$ or $w = \lambda$ is constant proving (ii).

Now, suppose (ii) is true so by Jensen's inequality,

$$\int_{\Delta S} \varphi(q \cdot w)\, d\mu \geq \varphi(p \cdot w) = \int_{\Delta S} \varphi(q \cdot w)\, d\delta_p$$

for all $w \in \mathbb{R}^S$ and convex φ. Moreover, if w is nonconstant, then μ^w is not degenerate proving (iii).

Finally, suppose (iii) is true so μ is strictly more informative than δ_p. Suppose $\alpha \cdot J = 0$ for some $\alpha \in \mathbb{R}^S$. We will show that $\alpha = 0$. Define $w \in \mathbb{R}^S$ such that $w_s = \alpha_s / p(s)$. By Bayes' rule from equation (10), for any measurable $R \subset \Theta$,

$$0 = \sum_s p(s)\, w_s J_s(R) = \sum_s w_s \int_R P_\theta(s)\, d\sigma = \int_R (w \cdot P_\theta)\, d\sigma.$$

Since this is true for all measurable $R \subset \Theta$, it must be that $P_\theta \cdot w = 0$ σ-a.s. or by a change of variables, $q \cdot w = 0$ μ-a.s.. Since μ is strictly more informative than δ_p, this means that w must be constant so $w = 0$. This implies $\alpha = 0$ proving (i). ■

The next result is very useful and characterizes the relationship between any two posterior distributions that are consistent with the same signal structure.[45]

[45] For a characterization of posterior beliefs, see Câmara and Alonso (2016).

LEMMA 6: *Suppose μ and ν have full-support priors p and r, respectively, and are both consistent with J. Then for any measurable $\varphi : \Delta S \to \mathbb{R}$,*

$$\int_{\Delta S} \varphi(q)\, d\nu = \int_{\Delta S} \varphi(\phi(q))(q \cdot \lambda)\, d\mu,$$

where $\lambda \in \mathbb{R}^S$ is such that $\lambda_s = r(s)/p(s)$ and $\phi : \Delta S \to \Delta S$ is such that for all $s \in S$,

$$[\phi(q)](s) = \frac{\lambda_s q(s)}{q \cdot \lambda}.$$

PROOF:

Let σ be the signal distribution and P the posterior mapping for μ and J and likewise for η and Q in regards to ν and J. Since $\eta = r \cdot J$, by Bayes' rule from equation (10), we have for any measurable $R \subset \Theta$,

$$\eta(R) = \sum_s r(s) J_s(R) = \sum_s \frac{r(s)}{p(s)} \int_R P_\theta(s)\, d\sigma = \int_R (P_\theta \cdot \lambda)\, d\sigma.$$

This implies that for any measurable $\psi : \Theta \to \mathbb{R}$,

$$\int_\Theta \psi(\theta)\, d\eta = \int_\Theta \psi(\theta)(P_\theta \cdot \lambda)\, d\sigma. \tag{15}$$

By Bayes' rule again, we have for any measurable $R \subset \Theta$ and $s \in S$,

$$\frac{r(s)}{p(s)} \int_R P_\theta(s)\, d\sigma = r(s) J_s(R) = \int_R Q_\theta(s)\, d\eta = \int_R Q_\theta(s)(P_\theta \cdot \lambda)\, d\sigma,$$

where the last equation follows from equation (15) above. Since this is true for all measurable $R \subset \Theta$, it must be that σ-a.s.,

$$\lambda_s P_\theta(s) = Q_\theta(s)(P_\theta \cdot \lambda).$$

Since $P_\theta \cdot \lambda > 0$, we have $Q_\theta = \phi(P_\theta)$ σ-a.s.. Returning to equation (15), this means that for any measurable $\varphi : \Delta S \to \mathbb{R}$,

$$\int_{\Delta S} \varphi(q)\, d\nu = \int_\Theta \varphi(Q_\theta)\, d\eta = \int_\Theta \varphi(\phi(P_\theta))(P_\theta \cdot \lambda)\, d\sigma = \int_{\Delta S} \varphi(\phi(q))(q \cdot \lambda)\, d\mu,$$

as desired. ∎

D. *Proof of Theorem 2*

Let ρ and τ be represented by (μ, u) and (ν, v), respectively, and J be consistent with both μ and ν. We first show necessity. Suppose $\mu = \nu$ and $(u_s, u_t) = a(v_s, v_t) + b$ for some $a > 0$ and any two imperfectly revealed states $s, t \in S$. Partition $S = S_1 \cup \cdots \cup S_m$ such that any $s \in S_i$ and $t \in S_j$ are perfectly revealed for all $i \neq j$. Let $B_i = \Delta S_i$ so $\mu(B_1 \cup \cdots \cup B_m) = 1$. Note that for $q \in B_i$ and any $f \in A$,

$$q \cdot (u \circ f) = \sum_{s \in S_i} q(s)\, u_s(f(s)) = a\Big(\sum_{s \in S_i} q(s)\, v_s(f(s))\Big) + b = a(q \cdot (v \circ f)) + b.$$

Thus, $q \cdot (u \circ f) \geq q \cdot (u \circ g)$ if and only if $q \cdot (v \circ f) \geq q \cdot (v \circ g)$ for all $q \in B_1 \cup \cdots \cup B_m$. Since $\mu = \nu$, this means that $\rho(f,g) = \tau(f,g)$ for all $f,g \in A$. This proves necessity.

We now prove sufficiency. Suppose $\rho(f,g) = \tau(f,g)$ for all $f,g \in A$. By Lemma 4, we know that $u_s = \beta_s v_s + \alpha_s$ for $\beta_s > 0$. Since $\beta_s > 0$ for all $s \in S$, we can define ϕ_β as in equation (7) from the proof of Lemma 1. By the same argument as in the proof of Theorem 1, ρ is represented by $\left(\mu \circ \phi_\beta^{-1}, v\right)$ and by the Cramér-Wold theorem, $\mu \circ \phi_\beta^{-1} = \nu$.

Let p and r be the priors of μ and ν, respectively. Define $\lambda \in \mathbb{R}^S$ such that $\lambda_s = r(s)/p(s) > 0$ for all $s \in S$ and ϕ_λ as in Lemma 6. Since J is consistent with ν, this means that for any measurable function $\varphi : \Delta S \to \mathbb{R}$,

$$\int_{\Delta S} \varphi(\phi_\beta(q))\, d\mu = \int_{\Delta S} \varphi(q)\, d\nu = \int_{\Delta S} \varphi(\phi_\lambda(q))(q \cdot \lambda)\, d\mu.$$

Let $\gamma \in \mathbb{R}^S$ be such that $\gamma_s = \lambda_s/\beta_s$ for all $s \in S$, and by a slight abuse of notation, let $\beta^{-1} \in \mathbb{R}^S$ denote $\left(\beta^{-1}\right)_s = 1/\beta_s$. Note that $\phi_{\beta^{-1}} \circ \phi_\beta$ is the identity mapping and $\phi_{\beta^{-1}} \circ \phi_\lambda = \phi_\gamma$. Thus, for any measurable function φ,

$$\int_{\Delta S} (\varphi \circ \phi_{\beta^{-1}})(\phi_\beta(q))\, d\mu = \int_{\Delta S} (\varphi \circ \phi_{\beta^{-1}})(\phi_\lambda(q))(q \cdot \lambda)\, d\mu,$$

$$\int_{\Delta S} \varphi(q)\, d\mu = \int_{\Delta S} \varphi(\phi_\gamma(q))(q \cdot \lambda)\, d\mu. \tag{16}$$

Partition $S = S_1 \cup \cdots \cup S_m$ such that γ_s is the same for all $s \in S_i$. Let $B_i = \Delta S_i$ and note that $\phi_\gamma(q) = q$ for all $q \in B_i$. Moreover, $\phi_\gamma(q) \in B_i$ if and only if $q \in B_i$. From equation (16), we have

$$\int_{B_i} \varphi(q)\, d\mu = \int_{B_i} \varphi(q)(q \cdot \lambda)\, d\mu$$

for all measurable φ. Thus, by the Radon-Nikodym theorem, $q \cdot \lambda = 1$ for all $q \in B_i$ μ-a.s..

Let $B := B_1 \cup \cdots \cup B_m$ and $\gamma^k \in \mathbb{R}^S$ denote $\left(\gamma^k\right)_s = \left(\gamma_s\right)^k$. Consider some compact set $C \subset B^c = \Delta S \backslash B$. Applying equation (16) iteratively, we have

$$\begin{aligned}
\mu(C) &= \int_{\Delta S} \mathbf{1}_C(\phi_\gamma(q))(q \cdot \lambda)\, d\mu \\
&= \int_{\Delta S} \mathbf{1}_C(\phi_{\gamma^k}(q))(\phi_{\gamma^{k-1}}(q) \cdot \lambda) \cdots (q \cdot \lambda)\, d\mu \\
&= \int_{B^c} \mathbf{1}_C(\phi_{\gamma^k}(q))(\phi_{\gamma^{k-1}}(q) \cdot \lambda) \cdots (q \cdot \lambda)\, d\mu,
\end{aligned}$$

where the last equality follows from the fact that $\phi_\gamma(q) = q \notin C$ and $q \cdot \lambda = 1$ for all $q \in B$ μ-a.s.. Now, for any $q \in B^c$, there is some $k > 0$ such that $\phi_{\gamma^k}(q) \notin C$. Thus, $\lim_k \mathbf{1}_C(\phi_{\gamma^k}(q)) = 0$ for all $q \in B^c$. By dominated convergence, $\mu(C) = 0$. Since this is true for all compact $C \subset B^c$, $\mu(B^c) = 0$.

Thus, $\mu(B) = 1$ so $q \cdot \lambda = 1$ μ-a.s.. Since J_s are linearly independent, Lemma 5 implies that λ is constant. By the definition of λ, this means that $r = p$ so $\nu = \mu$ by Lemma 6. Finally, suppose s and t are imperfectly revealed, so there must exist some S_i such that $s, t \in S_i$. This implies that $\gamma_s = \gamma_t$ and since λ is constant, $\beta_s = \beta_t$. Thus (u_s, u_t) is an affine transformation of (v_s, v_t) as desired. ∎

E. *Proof of Lemma 2*

Let $(\succeq_1, \succeq_2)$ be represented by (μ_1, μ_2, u) and let p be the prior of μ_i for $i \in \{1,2\}$. Note that we can always let $\Theta = \Delta S$ be the canonical signal space and J_i be the information structure corresponding to μ_i. Let ν_i be the posterior distribution consistent with J_i for prior r. Define $\lambda \in \mathbb{R}^S$, where $\lambda_s = r(s)/p(s) > 0$ and the utility v, where $v_s := \lambda_s^{-1} u_s$ for all $s \in S$. Set ϕ_λ as in Lemma 6 so

$$\begin{aligned}\int_{\Delta S} \sup_{f \in F} q \cdot (v \circ f)\, d\nu_i &= \int_{\Delta S} \sup_{f \in F} (\phi_\lambda(q) \cdot (v \circ f))(q \cdot \lambda)\, d\mu_i \\ &= \int_{\Delta S} \sup_{f \in F} \Big(\sum_s q(s) \lambda_s v_s(f(s)) \Big)\, d\mu_i \\ &= \int_{\Delta S} \sup_{f \in F} q \cdot (u \circ f)\, d\mu_i.\end{aligned}$$

Thus, $\succeq_i$ is represented by (ν_i, v).

Finally, we show that ν_2 is more informative than ν_1. For any $w_1, w_2 \in \mathbb{R}^S$, let $\tilde{w}_1, \tilde{w}_2 \in \mathbb{R}^S$ be such that $\tilde{w}_j(s) = w_j(s)\beta_s$ for all $s \in S$. By the same argument as above, we have for $j \in \{1,2\}$,

$$\int_{\Delta S} \max_j (q \cdot w_j)\, d\nu_i = \int_{\Delta S} \max_j (q \cdot \tilde{w}_j)\, d\mu_i.$$

Since μ_2 is more informative than μ_1, Lemma 3 implies that

$$\int_{\Delta S} \max_j (q \cdot w_j)\, d\nu_2 \geq \int_{\Delta S} \max_j (q \cdot w_j)\, d\nu_1,$$

so ν_2 is also more informative than ν_1. ∎

F. *Conditional Choice*

In this section, we formally show that even if the analyst knows the menu preference $\succeq$, conditional preferences $\succeq_\Theta$ and the public signal structure J, she still cannot identify beliefs. This generalizes a lemma from Karni, Schmeidler and Vind (1983) to include menu choice. We say $(\succeq, \succeq_\Theta, J)$ is represented by (μ, u), if J is consistent with μ, the menu preference has a subjective learning representation consistent with μ, and conditional preferences have subjective expected utility representations consistent with J and μ.

DEFINITION: $(\succeq, \succeq_\Theta, J)$ *is represented by* (μ, u) *if*

(*i*) *J is consistent with* μ;

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

(*iii*) $\succeq_\theta$ *is represented by* $P_\theta \cdot (u \circ f)$ *σ-a.s., where* $\sigma \in \Delta\Theta$ *is the signal distribution and* $P : \Theta \rightarrow \Delta S$ *is the posterior mapping*.

LEMMA 7: *If* $(\succeq, \succeq_\Theta, J)$ *is represented by* (μ, u), *then for any full-support prior r, it is also represented by some* (ν, v), *where* ν *has prior r.*

PROOF:

Let $(\succeq, \succeq_\Theta, J)$ be represented by (μ, u), where μ has prior p. Let ν be the posterior distribution consistent with J for prior r and define the utility v such that $v_s := \lambda_s^{-1} u_s$ for all $s \in S$, where $\lambda_s = r(s)/p(s) > 0$. By Lemma 2, $\succeq$ is also represented by (ν, v).

We show that ν is consistent with the conditional preferences. Let σ be the signal distribution and P the posterior mapping for μ and J and likewise for η and Q in regards to ν and J. Recall from the proof of Lemma 6 that σ-a.s.,

$$Q_\theta(s) = \frac{\lambda_s P_\theta(s)}{\lambda \cdot P_\theta}.$$

We thus have σ-a.s. $f \succeq_\theta g$ if and only if $P_\theta \cdot (u \circ f) \geq P_\theta \cdot (u \circ g)$ if and only if

$$\sum_s P_\theta(s) \lambda_s v_s(f(s)) \geq \sum_s P_\theta(s) \lambda_s v_s(g(s)),$$

$$Q_\theta \cdot (v \circ f) \geq Q_\theta \cdot (v \circ g),$$

as $\lambda \cdot P_\theta > 0$. Note that by equation (15) in the proof of Lemma 6, σ and η are mutually absolutely continuous. Thus, $\succeq_\theta$ is represented by $P_\theta \cdot (u \circ f)$ η-a.s. which concludes the proof. ∎

G. *Proof of Proposition 1*

We first prove sufficiency. Let U denote the space of all state-dependent utilities $v = (v_s)_{s \in S}$. From Lu (2016), Axioms 1.1–1.4 imply that there exists a measure π on $\Delta S \times U$ such that

$$\rho_F(f) = \pi\{(q, v) \in \Delta S \times U \,|\, q \cdot (v \circ f) \geq q \cdot (v \circ g) \text{ for all } g \in F\}.$$

Moreover, π satisfies the regularity property in that $q \cdot (v \circ f) = q \cdot (v \circ g)$ with π-measure zero or one. Note that Axiom 1.5 implies that for every $s \in S$,

$$\pi\{(q, v) \in \Delta S \times U \,|\, q(s) v_s(f(s)) \geq q(s) v_s(g(s))\} = \rho(fsh, gsh) \in \{0, 1\}.$$

Note that if $q(s) = 0$ μ-a.s., then $\rho(fsh, gsh) = 1$ contradicting Axiom 1.6. By the regularity of π, this implies that $q(s) > 0$ μ-a.s.. Thus, $v_s(f(s)) \geq v_s(g(s))$ π-measure zero or one for all $f, g \in A$. By Lu (2016), this means that π has some degenerate distribution on some nonconstant $u_s \in \mathbb{R}^X$. Since this is true for all $s \in S$, this implies that ρ is represented by (μ, u), where μ is the marginal of π on ΔS and $u = (u_s)_{s\in S}$. Note that since $q(s) > 0$ μ-a.s., the prior of μ has full support. Necessity is straightforward. ∎

H. *Proof that f^h Is a c.d.f.*

In this section, we show that the induced function f^h is a c.d.f. This is an extension of a result from Lu (2016) to the case with state dependence. We prove the following.

LEMMA 8: *Suppose ρ is represented by (μ, u). Let $h > \underline{h}$ and $f \in (\underline{h}, h)$. Then for measurable $\varphi : [0,1] \to \mathbb{R}$,*

$$\int_0^1 \varphi \, df^h = \int_{\Delta S} \varphi\left(\frac{q \cdot (u \circ h) - q \cdot (u \circ f)}{q \cdot (u \circ h) - q \cdot (u \circ \underline{h})}\right) d\mu.$$

PROOF:

Fix some action f and define the function ξ such that

$$\xi(q) := \frac{q \cdot (u \circ h) - q \cdot (u \circ f)}{q \cdot (u \circ h) - q \cdot (u \circ \underline{h})}.$$

Note that since $h > \underline{h}$, $q \cdot (u \circ h) > q \cdot (u \circ \underline{h})$ μ-a.s. so ξ is well defined. Moreover, $f \in (\underline{h}, h)$ so ξ has range $[0,1]$. Let $\mu^\xi := \mu \circ \xi^{-1}$ denote the image measure of ξ on $[0,1]$. By a standard change of variables, for any measurable $\varphi : [0,1] \to \mathbb{R}$,

$$\int_0^1 \varphi \, d\mu^\xi = \int_{\Delta S} (\varphi \circ \xi) \, d\mu.$$

We prove the lemma by showing that the c.d.f. of μ^ξ is exactly f^h. Now,

$$\begin{aligned}
\mu^\xi[0,a] &= \mu\left\{q \in \Delta S \,|\, a \geq \xi(q) \geq 0\right\} \\
&= \mu\left\{q \in \Delta S \,|\, q \cdot (u \circ f) \geq aq \cdot (u \circ \underline{h}) + (1-a) q \cdot (u \circ h)\right\} \\
&= \rho(f, h^a) = f^h(a)
\end{aligned}$$

for all $a \in [0,1]$. Note that $\mu^\xi[0,1] = 1 = f^h(1)$ so f^h is the c.d.f. of μ^ξ as desired. ∎

I. *Proof of Theorem 3*

We first prove sufficiency. From Proposition 1, we know that ρ_1 and ρ_2 are represented by regular (ν_1, v) and $(\tilde{\nu}, \tilde{v})$. Moreover, Axiom 1.6 implies that $q(s) > 0$ ν_1-a.s. and $\tilde{\nu}$-a.s. for all $s \in S$. By Axiom 2, we have that $v_s(f(s)) \geq v_s(g(s))$ if and only if $\rho_1(fsh, gsh) = 1$ if and only if $\rho_2(fsh, gsh) = 1$ if and only if $\tilde{v}_s(f(s)) \geq \tilde{v}_s(g(s))$. Thus, $w_s = \beta_s \tilde{v}_s + \alpha_s$ for $\beta_s > 0$ so if we define ϕ_β as in the proof of Lemma 1, then by the same argument as before, ρ_2 is represented by (ν_2, v), where $\nu_2 := \tilde{\nu} \circ \phi_\beta^{-1}$. Moreover, if we let $\bar{h}$ and $\underline{h}$ be the best and worst actions, then without loss of generality, we can normalize v such that $v_s(\bar{h}(s)) = 1$ and $v_s(\underline{h}(s)) = 0$ for all $s \in S$ which is possible as v_s are all nonconstant.

Thus, ρ_i is represented by (ν_i, v). By Axiom 3, let h be a calibrating action. Define $\gamma \in \mathbb{R}^S$ such that $\gamma_s := v_s(h(s)) > 0$ for all $s \in S$ as $h > \underline{h}$ by definition. Define u such that $u_s = \gamma_s^{-1} v_s$ for all $s \in S$. If we let $\mu_i := \nu_i \circ \phi_\gamma^{-1}$, then ρ_i is represented by (μ_i, u).

Finally, we show that μ_2 is strictly more informative than μ_1. Since $u \circ h = \mathbf{1}$, by Lemma 8, we have for any φ,

$$\int_0^1 \varphi \, df_i^h = \int_{\Delta S} \varphi(1 - q \cdot (u \circ f)) \, d\mu_i.$$

Now, for any $w \in \mathbb{R}^S$, we can find some $\lambda > 0$, $k \in \mathbb{R}$, and $f \in A$ such that

$$\lambda w + k\mathbf{1} = \mathbf{1} - (u \circ f).$$

Axiom 4 implies that for any convex φ,

$$\int_{\Delta S} \varphi(\lambda q \cdot w + k) \, d\mu_1 = \int_0^1 \varphi \, df_1^h \leq \int_0^1 \varphi \, df_2^h = \int_{\Delta S} \varphi(\lambda q \cdot w + k) \, d\mu_2.$$

Since $\varphi(x)$ is convex if and only if $\varphi(\lambda x + k)$ is convex, this means that μ_2 is more informative than μ_1. To see strict informativeness, suppose $w \in \mathbb{R}^S$ is nonconstant. We can then find $\lambda > 0$ and $k \in \mathbb{R}$ such that $\lambda w + k\mathbf{1} = \mathbf{1} - (u \circ f)$, where f and $(1/2)\underline{h} + (1/2)h$ are incomparable. Axiom 4 then delivers strict informativeness.

We now show necessity. Suppose (ρ_1, ρ_2) is represented by (μ_1, μ_2, u), where μ_2 is strictly more informative than μ_1. Note that Axiom 1 follows from Proposition 1 and Axiom 2 is trivial. To see Axiom 3, set $h \in A$ such that $u(h(s)) = 1$ for all $s \in S$. Let p be the prior of both μ_1 and μ_2 and note that from Lemma 8,

$$\int_0^1 a \, df_i^h = 1 - \int_{\Delta S} q \cdot (u \circ f) \, d\mu_i = 1 - p \cdot (u \circ f).$$

Since this is true for all $f \in (\underline{h}, h)$, h is a calibrating action.

Finally, we show Axiom 4. Suppose h is a calibrating action so again by Lemma 8,

$$\int_{\Delta S} \frac{q \cdot (u \circ f)}{q \cdot (u \circ h)} \, d\mu_1 = \int_{\Delta S} \frac{q \cdot (u \circ f)}{q \cdot (u \circ h)} \, d\mu_2$$

for all $f \in (\underline{h}, h)$. Summing up over all the indicator actions for each state, we have

$$\int_{\Delta S} (q \cdot (u \circ h))^{-1} \, d\mu_1 = \int_{\Delta S} (q \cdot (u \circ h))^{-1} \, d\mu_2.$$

Since μ_2 is strictly more informative than μ_1, this means $u \circ h$ must be constant. Axiom 4 then follows from Lemma 8. ∎

J. *Elicitation*

In this section, we introduce an elicitation procedure for calibrating actions. For each $h \in A$, let B_h denote the set of acts $f \in (\underline{h}, h)$ such that f_1^h has a strictly higher mean than f_2^h. We can consider the linear extension of this set $\bar{B}_h$, where $f \in \bar{B}_h$ if $\alpha f + (1 - \alpha)\underline{h} \in B_h$ for $\alpha > 0$. We now construct a sequence of actions $(h_k)_{k \in \mathbb{N}}$ as follows. First, choose some initial $h_0 \in A_0 := A$. Now, recursively define

$$A_{k+1} := \bar{B}_{h_k} \cap A_k$$

and let $h_{k+1} \in A_{k+1}$. If no such h_{k+1} exists, then set $h_{k+1} = h_k$. Given any sequence h_k, $A_{k+1} \subset A_k$ is a monotonically decreasing sequence of sets. Hence, by the monotone convergence theorem, it will always converge and we let A^* denote its limit. We say a sequence $h_k \to h^*$ converges *generically* if A^* has less than full dimension.[46] In other words, the sequence of actions are chosen such that A^* is reduced to a lower-dimensional set. Note that one can always choose h_{k+1} far from the boundaries of each A_{k+1} so that it will converge generically. We can always use this procedure to elicit the calibrating action.

PROPOSITION 2: *Suppose (ρ_1, ρ_2) is represented by (μ_1, μ_2, u), where μ_2 is strictly more informative than μ_1. If $h_k \to h^*$ generically, then h^* is calibrating.*

PROOF:

By Lemma 8, note that for every $h > \underline{h}$ and $f \in (\underline{h}, h)$,

$$\int_0^1 a \, df_i^h = 1 - \int_{\Delta S} \frac{q \cdot (u \circ f)}{q \cdot (u \circ h)} \, d\mu_i.$$

For every $h \in A$, define $w^h \in \mathbb{R}^S$, where

$$w^h := \int_{\Delta S} q (q \cdot (u \circ h))^{-1} \, d\mu_2 - \int_{\Delta S} q (q \cdot (u \circ h)^{-1} \, d\mu_1.$$

Thus, f_1^h has a strictly higher mean than f_2^h if and only if $w^h \cdot (u \circ f) > 0$. Note that by dominated convergence, w^h is continuous in h. Also, note that $w^h \cdot (u \circ h) = 0$ and

$$A^* := \bigcap_k \left\{ f \in A \,|\, w^{h_k} \cdot (u \circ f) > 0 \right\}.$$

[46] That is, the dimension of $u(A^*)$ is less than the dimension of $u(A)$.

Thus, A^* corresponds to an open convex cone in $\mathbb{R}^S$.

Suppose $h_k \to h^*$ generically. We will show that $u \circ h^*$ is constant. Suppose otherwise and let $\bar{h}$ be an action such that $u \circ \bar{h} = \mathbf{1}$. Since $h_k \to h^*$, $u \circ h_k \to u \circ h^*$ so we can assume without loss of generality that $u \circ h_k$ is also not constant for all k. Thus, by Lemma 3,

$$w^{h_k} \cdot (u \circ \bar{h}) = \int_{\Delta S} q \cdot (u \circ h_k)^{-1} d\mu_2 - \int_{\Delta S} q \cdot (u \circ h_k)^{-1} d\mu_1 > 0,$$

as μ_2 is strictly more informative than μ_1. Since this is true for all h_k, this means that $\bar{h} \in A^*$. Suppose $w^{h^*} \cdot (u \circ \bar{h}) < 0$. Since $w^{h^*} \cdot (u \circ h^*) = 0$,

$$w^{h^*} \cdot \left[\frac{1}{2}(u \circ \bar{h}) + \frac{1}{2}(u \circ h^*)\right] < 0,$$

which means that $(1/2)\bar{h} + (1/2)h^* \notin A^*$. Since $h^*, \bar{h} \in A^*$ and A^* is convex, this is a contradiction. The case for $w^{h^*} \cdot (u \circ \bar{h}) > 0$ is symmetric so we have

$$0 = w^{h^*} \cdot (u \circ \bar{h}) = \int_{\Delta S} (u \circ h^*)^{-1} d\mu_2 - \int_{\Delta S} (u \circ h^*)^{-1} d\mu_1.$$

By strict informativeness, $u \circ h^*$ must be constant, yielding a contradiction. Thus, $u \circ h^*$ is constant so h^* is calibrating as desired. ∎

Proposition 2 not only provides a procedure for eliciting a calibrating action if one exists, it also suggest an alternate characterization of our model. For example, suppose Axioms 1 and 2 are satisfied. One can then always construct a sequence h_k that converges to h^* generically. If h^* is calibrating, then Axiom 3 is satisfied and we can then test Axiom 4. On the other hand, if h^* is not calibrating (e.g., it does not strictly dominate $\underline{h}$), then no representation can exist. From a computational perspective, this procedure is less demanding than testing Axiom 3 over all actions and provides a more efficient characterization.

REFERENCES

Agan, Amanada, and Sonja Starr. 2018. "Ban the Box, Criminal Records, and Racial Discrimination: A Field Experiment." *Quarterly Journal of Economics* 133 (1): 191–235.

Ahn, David S., and Todd Sarver. 2013. "Preference for Flexibility and Random Choice." *Econometrica* 81 (1): 341–61.

Alonso, Ricardo, and Odilon Câmara. 2016. "Bayesian Persuasion with Heterogeneous Priors." *Journal of Economic Theory* 165: 672–706.

Altonji, Joseph G., and Charles R. Pierret. 2001. "Employer Learning and Statistical Discrimination." *Quarterly Journal of Economics* 116 (1): 313–50.

Anscombe, Francis J., and Robert Aumann. 1963. "A Definition of Subjective Probability." *The Annals of Mathematical Statistics* 34 (1): 199–205.

Arrow, Kenneth. 1971. *Some Models of Racial Discrimination in the Labor Market.* Santa Monica: RAND Corporation.

Autor, David H., and David Scarborough. 2008. "Does Job Testing Harm Minority Workers? Evidence from Retail Establishments." *Quarterly Journal of Economics* 123 (1): 219–77.

Becker, Gary S. 1957. *The Economics of Discrimination.* Chicago: University of Chicago Press.

Benson, Alan, Simon Board, and Moritz Meyer-ter-Vehn. 2018. "Discrimination in Hiring: Evidence from Retail Sales." Unpublished.

Bertrand, Marianne, and Sendhil Mullainathan. 2004. "Are Emily and Greg More Employable Than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination." *American Economic Review* 94 (4): 991–1013.
Blackwell, David. 1951. "Comparison of Experiments." 93–102. Paper presented at the Second Berkeley Symposium on Mathematical Statistics and Probability, Berkeley, CA.
Blackwell, David. 1953. "Equivalent Comparisons of Experiments." *The Annals of Mathematical Statistics* 24 (2): 265–72.
Brunnermeier, Markus K., Alp Simsek, and Wei Xiong. 2014. "A Welfare Criterion of Models with Distorted Beliefs." *Quarterly Journal of Economics* 129 (4): 1753–97.
Dillenberger, David, R. Vijay Krishna, and Philipp Sadowski. 2017. "Subjective Information Choice Processes." Unpublished.
Dillenberger, David, Juan Sebastián Lleras, Philipp Sadowski, and Norio Takeoka. 2014. "A Theory of Subjective Learning." *Journal of Economic Theory* 153: 287–312.
Dillenberger, David, Andrew Postlewaite, and Kareen Rozen. 2017. "Optimism and Pessimism with Expected Utility." *Journal of the European Economic Association* 15 (5): 1158–75.
Drèze, Jacques H. 1987. "Decision Theory with Moral Hazard and State-Dependent Preferences." In *Essays on Economic Decisions Under Uncertainty, 23–89*. Cambridge, UK: Cambridge University Press.
Drèze, Jacques H., and Aldo Rustichini. 1999. "Moral Hazard and Conditional Preferences." *Journal of Mathematical Economics* 31 (2): 159-81.
Elton, J., and T. P. Hill. 1992. "Fusions of a Probability Distribution." *The Annals of Probability* 20 (1): 421–54.
Fishburn, Peter C. 1973. "A Mixture-Set Axiomatization of Conditional Subjective Expected Utility." *Econometrica* 41 (1): 1–25.
Gilboa, Itzhak, Larry Samuelson, and David Schmeidler. 2014. "No-Betting-Pareto Dominance." *Econometrica* 82 (4): 1405–42.
Goldin, Claudia, and Cecilia Rouse. 2000. "Orchestrating Impartiality: The Impact of 'Blind' Auditions on Female Musicians." *American Economic Review* 90 (4): 715–41.
Gul, Faruk, and Wolfgang Pesendorfer. 2006. "Random Expected Utility." *Econometrica* 74 (1): 121–46.
Heckman, James J. 1998. "Detecting Discrimination." *Journal of Economic Perspectives* 12 (2): 101–16.
Hoffman, Mitchell, Lisa B. Kahn, and Danielle Li. 2018. "Discretion in Hiring." *Quarterly Journal of Economics* 133 (2): 765–800.
Karni, Edi. 1993. "Notes and Comments: A Definition of Subjective Probabilities with State-Dependent Preferences." *Econometrica* 61 (1): 187–98.
Karni, Edi. 2006. "Subjective Expected Utility Theory without States of the World." *Journal of Mathematical Economics* 42 (3): 325–42.
Karni, Edi. 2007. "Foundations of Bayesian Theory." *Journal of Economic Theory* 132 (1): 167–88.
Karni, Edi, and Philippe Mongin. 2000. "On the Determination of Subjective Probability by Choices." *Management Science* 46 (2): 233–48.
Karni, Edi, and Zvi Safra. 2016. "A Theory of Stochastic Choice under Uncertainty." *Journal of Mathematical Economics* 63: 164–73.
Karni, Edi, and David Schmeidler. 2016. "An Expected Utility Theory for State-Dependent Preferences." *Theory and Decision* 81 (4): 467–78.
Karni, Edi, David Schmeidler, and Kard Vind. 1983. "On State Dependent Preferences and Subjective Probabilities." *Econometrica* 51 (4): 1021–31.
Krishna, R. Vijay, and Philipp Sadowski. 2014. "Dynamic Preference for Flexibility." *Econometrica* 82 (2): 655–703.
Lu, Jay. 2016. "Random Choice and Private Information." *Econometrica* 84 (6): 1983–2027.
Luce, R. Duncan, and David H. Krantz. 1971. "Conditional Expected Utility." *Econometrica* 39 (2): 253–71.
Masatlioglu, Yusufcan, Daisuke Nakajima, and Emre Ozdenoren. 2014. "Revealed Willpower." Unpublished.
Phelps, Edmund S. 1972. "The Statistical Theory of Racism and Sexism." *American Economic Review* 62 (4): 659–61.
Sadowski, Philipp. 2013. "Contingent Preference for Flexibility: Eliciting Beliefs from Behavior." *Theoretical Economics* 8 (2): 503–34.
Savage, Leonard J. 1954. *The Foundations of Statistics.* Hoboken: John Wiley & Sons, Inc.
Schenone, Pablo. 2016. "Identifying Subjective Beliefs in Subjective State Space Models." *Games and Economic Behavior* 95: 59–72.
Skiadas, Costis. 1997. "Subjective Probability under Additive Aggregation of Conditional Preferences." *Journal of Economic Theory* 76 (2): 242–71.