A Bayesian Theory of State-Dependent Utilities*

Jay Lu†

August 2016

Abstract

We provide a foundation for beliefs within the classic revealed preference methodology that allows for state-dependent utilities. Suppose an agent is Bayesian and signals affect beliefs but not tastes. Under these assumptions, an analyst who only observes the agent's pre-signal preferences and post-signal random choice can uniquely identify the agent's prior, signal structure and state-dependent utilities.

* I am grateful to Edi Karni, Yusufcan Masatlioglu, Jawaad Noor, Marek Pycia, Tomasz Sadzik and Pablo Schenone, as well as seminar audiences at Bocconi, Berkeley, SWET (UC Riverside), Michigan and SAET (IMPA) for helpful comments and suggestions.

† Department of Economics, UCLA; jay@econ.ucla.edu

1 Introduction

One of the classic objectives of decision theory was to provide a choice-theoretic foundation for the identification of beliefs and tastes. Following a long tradition of seminal works by Ramsey (1931), de Finetti (1937), Savage (1954) and Anscombe and Aumann (1963), subjective expected utility emerged as a successful and widely-used model of decision-making in economics. Nevertheless, in environments where utilities may be state-dependent, the model fails to provide a theory that successfully distinguishes beliefs and tastes. For instance, one could always scale utilities up and beliefs down in such a way so that the model remains unchanged. As Karni (2007) puts it, "Ultimately, however, this quest [of identifying beliefs and utilities] failed to achieve its goal."

To illustrate the relevance of this problem for everyday economics, consider the following application. A lender (the agent) is faced with a batch of loan applications and has to decide whether or not to approve each applicant. Each applicant can either be good or bad, but this is unknown to the lender. The lender observes some signal (e.g. a credit report) before making a decision. A regulator (the analyst) does not observe the lender's signals but does observe his approval rate, i.e., the proportion of applicants who are approved. The regulator wants to identify the lender's prior and signal structure. However, the fact that the lender may receive hidden payoffs (e.g. kickbacks) for approving loans complicates this problem. For instance, a lender with more pessimistic beliefs about applicants but receives a kickback for approving loans may exhibit the same approval rate as an optimistic lender who receives no kickbacks. This is analogous to the classic problem of identifying beliefs given state-dependent utilities.

Now, suppose there is an identical group of applicants whose credit reports were lost for some exogenous reason. Call this the pre-signal group. If the regulator also observes the lender's approval decisions for the pre-signal group, then he can now differentiate between a lender who is receiving kickbacks versus one who is not. Despite exhibiting the same post-signal approval rate, a lender who receives kickbacks may exhibit a stronger pre-signal preference to approve. This is because Bayesian updating provides just enough discipline across pre-signal and post-signal choices so that beliefs and state-dependent utilities can all be uniquely identified. This methodology for identification can be applied in many other economic settings, such as college admissions, job applications[1], etc.

[1] In the case of hiring, the pre-signal group could correspond to a treatment involving expunged criminal

In this paper, we use Bayesian updating to provide a choice-theoretic foundation for beliefs and state-dependent utilities in the classic revealed preference methodology. Formally, the analyst observes two types of choice data: a pre-signal preference relation and a post-signal random choice. The post-signal choices are random to the analyst precisely because he does not observe the agent's signals. We show that if the agent is Bayesian and information only affects his beliefs, then the analyst can identify the agent's beliefs, information and state-dependent utilities. In other words, beliefs convey testable *empirical meaning* despite utilities being state-dependent. Our theory thus delivers a choice-theoretic foundation for a separation of beliefs and tastes.

In Section 4, we present our general model. Following the setup of Anscombe and Aumann (1963), we consider an objective state space where each choice option is a state-contingent payoff called an act. The pre-signal choice data is a standard preference relation over acts. The post-signal choice data is a random choice function (RCF), that is, a choice distribution over acts for every menu of acts. A joint pre- and post-signal choice data set has a *Bayesian representation* if there is a state-dependent utility function u and a distribution of beliefs μ that rationalizes both the pre-signal preference relation and the post-signal RCF. Theorem 1 says that μ and u are unique in any Bayesian representation. In other words, beliefs and state-dependent utilities can all be completely identified.

Given that random choice already contains probabilistic data, one may think that identification of information would be relatively straightforward. After all, if the agent's signals are perfectly informative, then the analyst can essentially observe the prior directly from the choice frequencies. However, this reasoning misses two important points. First, in most economic applications, signals are not perfectly informative and involve noise. In this case, identification of beliefs and utilities based on random choice data alone is impossible (see Lemma 1). Our result shows what additional data is necessary in order to achieve this identification. Second, even when signals are perfectly informative and the agent's prior can be identified, his state-dependent utilities remain indeterminate as they can still be scaled arbitrarily. By incorporating pre-signal choices however, these can be identified. Note that the standard normalization usually assumed to obtain state-independent utilities given state-independent preferences is now something that can be empirically tested.

In Section 5, we consider the empirical content of beliefs in the context of different types of data. Lemma 1 shows that random choice by itself is not enough to identify beliefs.

records or "blind" auditions (see Goldin and Rouse (2000)).

Lemma 2 shows that preferences over menus is also not enough to identify beliefs. In other words, beliefs lack empirical meaning in the context of either of these data sets individually. What if the analyst knows the agent's true signal structure, i.e. the conditional signal distributions? Lemma 3 generalizes a result in Karni, Schmeidler and Vind (1983) and shows that even if the analyst knows the agent's signal structure, pre-signal menu preferences and conditional preferences for each signal realization, he still cannot pin down the agent's beliefs. This implies that an analyst who knows an agent's signal structure can simply pick any belief-utility pair consistent with the agent's pre-signal preferences and use that to infer the agent's conditional preferences. The same is not true if the analyst wanted to infer the agent's random choice. Lemma 4 shows that random choice along with the signal structure is enough to pin down beliefs. Our findings, summarized in Figure 1 of Section 5, are thus important for informing data collection on the part of the analyst.

In Section 6, we provide a full axiomatic characterization of the Bayesian representation. We first present the standard axioms for the deterministic (random) subjective expected utility representation of the pre-signal preference relation (post-signal RCF). Next, we introduce two new axioms specific to the Bayesian representation that links pre- and post-signal choice data. First, given a RCF ρ and a fixed act h, let $\succeq_{\rho,h}$ denote the induced preference relation where $f \succeq_{\rho,h} g$ if the expected probability that f is chosen over a uniformly drawn mixture between h and the worst act is greater than the corresponding expected probability for g. *Bayesian Consistency* says there there exists an act h^* such that $\succeq_{\rho,h^*}$ is exactly the pre-signal preference relation. We call such an act *calibrating*. In other words, average choice probabilities under the calibrating act agree exactly with the pre-signal preferences. Our second axiom, *Static Tastes*, says that random choice over acts that agree over all but one state must match the pre-signal preference relation. This ensures that all the randomness in post-signal choice is being driven only by information and not tastes. Theorem 2 says that the standard axioms along with these two new ones are necessary and sufficient for a Bayesian representation.

Our axiomatization is useful for another purpose: it provides a methodology for the analyst to elicit the agent's beliefs and utilities directly from choice data. Consider the following iterative procedure. First, fix an act h_0 and consider the set of acts that are indifferent to it according to the pre-signal preference relation. Now, choose some act in the set of acts that are $\succeq_{\rho,h_0}$-better than h_0 and call it h_1. By iterating this procedure and choosing these acts efficiently, one can obtain a sequence of acts h_i that converges. We then show that if the limit of a convergent sequence is some non-degenerate h^*, then h^* must be calibrating. Otherwise, Bayesian Consistency is violated. This procedure thus provides a revealed-preference test of whether an agent is Bayesian and if so, how to elicit his beliefs and utilities from choice data.

Our main methodology relies on random choice data that is readily available in the examples we have considered so far (loan approvals, college admissions, job applications, etc.). However, in other examples (e.g. health insurance) where a consumer's choices are infrequent and occur between long periods of time, the availability of this random choice data is an issue. In an ideal world, one could interpret the random choice as corresponding to the decisions of a consumer every year across many years. Assuming that tastes do not vary too much, then the identification methodology we introduce can still shed some light on beliefs and state-dependent utilities in these cases. Of course, this requires additional assumptions by the modeler and would be an interesting avenue for future research.

The theoretical results developed in this paper also have important policy implications. For example, in job hirings, if a regulator wants to implement some form of affirmative action via kickbacks, then the effect of the policy would be different depending on how informative job interviews or resumes are. 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 tastes. Our theory thus provides a methodology for identifying beliefs and tastes that is useful for conducting such welfare analyses for regulatory evaluations.

2 Related Literature

There is a long literature addressing the identification shortcomings of subjective expected utility.[2] Luce and Krantz (1971), Fishburn (1973) and Karni (2007) all use enlarged choice spaces that include conditional acts to model state-dependent utilities.[3] 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.[4] Karni (1993) does

[2] While we focus on state-dependent utilities, Dillenberger, Postlewaite and Rozen (2014b) show that an alternate form of indeterminacy arises if we allow for payoff-dependent beliefs.

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

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

use a traditional primitive in a model where a state-dependent mapping on payoffs 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 likelihoods of states are affected by the agent's actions to identify state-dependent utilities but this approach is limited to instances where such "moral hazard" is present.

Recent papers have employed menu choice to address this issue. Sadowski (2013) and Schenone (2016) assume a normalized state-independent utility across a subset of prizes in order to achieve identification. Krishna and Sadowski (2014) make use of the recursive structure in an infinite-period model while Karni and Safra (2016) consider an additional preference relation on hypothetical mental state-act lotteries. Ahn and Sarver (2013) study both menu and random choice in order to identify beliefs and utilities over a subjective state space. Their model is different 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 updating and information can be modeled explicitly. Separating utilities from the state space also allows for richer comparisons of utilities across states.[5] Second, in our model, the ex-ante (pre-signal) data consists only of choice over singletons and not menus.[6] In fact, the methodology that we use in order to achieve identification would not work in their setting so our approach is inherently different than theirs.[7] Thus, the Anscombe and Aumann (1963) setup provides just enough additional structure so that Bayes' rule has bite and identification is possible.

3 Motivating Example: Loan Approvals

We first present a motivating example to illustrate the problem and the identification methodology. Consider a lender (the agent) faced with a batch of long-term loan applications. For each applicant, there are two possible states: good (s_1) and bad (s_2). In the good state, the applicant's project is successful and the loan is repaid, while in the bad state, the project is

[5] For example, we can talk about an agent who has state-dependent utilities but state-independent preferences in our model.

[6] Although in a different context, Masatlioglu, Nakajima and Ozdenoren (2014) also study a primitive consisting of an ex-ante preference over *singletons* and ex-post choice.

[7] To see this, consider the following example. Consider two prizes and let μ and ν be two distributions over utilities (i.e. subjective states) in $\mathbb{R}^2_+$. Suppose μ is uniform over all utilities on the line $u_1 + u_2 = 1$ while ν is uniform over all utilities on the circle $u_1^2 + u_2^2 = 1$. Even though μ and ν are different, they both generate the same ex-ante preference relation over *singleton* lotteries and ex-post random choice.

unsuccessful and the applicant defaults. There is a fixed signal structure (e.g. credit reports) that informs the lender about the likelihood of default for each applicant. For simplicity, assume that there are two signals, a high credit score (θ_1) and a low credit score (θ_2), where a high score suggests a higher chance of project success. If the loan is approved, then the lender receives a payoff of 1 in the good state or a payoff of -1 in the bad state. Otherwise, the lender receives a normalized payoff of 0 in either state. The lender observes the signal for each applicant and then decides whether or not to approve the loan.[8] Moreover, he makes these decisions before any state is realized so his only source of information for each applicant is his signal.

A regulator (the analyst) is interested in identifying the lender's prior and signal structure. Unknown to the regulator however, the lender may be receiving additional payoffs (e.g. kickbacks) for approving loans. For instance, if the lender receives a kickback of $\frac{1}{2}$ for approving a loan, then his final payoff would be $\frac{3}{2}$ in the good state or $-\frac{1}{2}$ in the bad state. This introduces an identification problem for the regulator analogous to the classic issue with state-dependent utilities. Suppose the regulator observes the lender's approval rate of $\frac{1}{2}$, that is, half of all applicants are approved. Based on the approval rate, the regulator cannot distinguish between a lender who has pessimistic beliefs about applicants but receives a kickback versus one who has optimistic beliefs about applicants but receives no kickbacks.

To be explicit, consider a lender who receives no kickbacks. Suppose the lender has prior $p = \frac{1}{2}$ and a completely symmetric signal distribution $L(\theta_i | s_i) = \frac{3}{4}$ for $i \in \{1, 2\}$. The lender's posterior beliefs that the project will be a success given the high and low credit scores respectively are

$$(q_1, q_2) = \left(\frac{3}{4}, \frac{1}{4}\right)$$

Hence, he will approve whenever

$$\begin{aligned} q_i + (1 - q_i)(-1) &\geq 0 \\ q_i &\geq \frac{1}{2} \end{aligned}$$

or whenever he gets a high score. The unconditional probability of a high score is

$$pL(\theta_1 | s_1) + (1 - p) L(\theta_1 | s_2) = \frac{1}{2}$$

[8] We assume that the lender considers each applicant individually (e.g. due to risk-neutrality, time-separable preferences or fairness considerations).

which is consistent with the observed approval rate of $\frac{1}{2}$.

Now, consider a lender who receives a kickback of $\frac{1}{2}$ for approving every loan. Suppose the lender has prior $\tilde{p} = \frac{3}{10}$ and signal distributions $\tilde{L}(\theta_1 | s_1) = \frac{5}{6}$ and $\tilde{L}(\theta_2 | s_2) = \frac{9}{14}$. The lender's posterior beliefs that the project will be a success given the high and low credit scores respectively are

$$(\tilde{q}_1, \tilde{q}_2) = \left(\frac{1}{2}, \frac{1}{10}\right)$$

Hence, he will approve whenever

$$\begin{aligned} \tilde{q}_i \left(\frac{3}{2}\right) + (1 - \tilde{q}_i)\left(-\frac{1}{2}\right) &\geq 0 \\ \tilde{q}_i &\geq \frac{1}{4} \end{aligned}$$

or again when he gets a high score. The unconditional probability of a high score is

$$\tilde{p}\tilde{L}(\theta_1 | s_1) + (1 - \tilde{p})\tilde{L}(\theta_1 | s_2) = \frac{1}{2}$$

which is again consistent with the observed approval rate of $\frac{1}{2}$.

Since both lenders exhibit an approval rate of $\frac{1}{2}$, the regulator is unable to distinguish between the two based on approval rates. In fact, this problem is even more severe in that for *any* payoff scheme, the regulator will not be able to distinguish between the two. This is because within the expected utility framework, the identification of priors is intrinsically linked with the identification of utilities; beliefs are properly defined only when they are jointly specified along with utilities. As in this example, the possibility of kickbacks adds an extra dimension of indeterminacy that prevents full identification of beliefs.

Now, suppose there is an identical group of applicants whose credit reports were lost for some exogenous reason. Call this the pre-signal group. How would the two lenders treat applicants from this group?[9] Note that

$$p + (1 - p)(-1) = 0$$
$$\tilde{p}\left(\frac{3}{2}\right) + (1 - \tilde{p})\left(-\frac{1}{2}\right) = \frac{1}{10} > 0$$

Hence, the lender not receiving kickbacks will be indifferent between approving or not while

[9] One could also interpret this as the lender's "initial recommendation" before receiving any information about applicants.

the lender receiving kickbacks will strictly prefer approving. In other words, fixing the approval rate for post-signal applicants, the lender who is receiving kickbacks exhibits a *stronger* preference for approving pre-signal applicants. By looking at both approval rates *and* pre-signal preferences, the regulator can now distinguish between the two. We provide a formal statement of this result in Theorem 1.

In this example, we have assumed that the lender updates via Bayes' rule and has correct beliefs about the distribution of signals. In general though, the only assumption required is that the lender's beliefs about his subjective signals are correct. In other words, even if the lender is reading signals erroneously or has incorrect beliefs, our results apply as long as his subjective prior is the average of his subjective posteriors.[10] This is the exact restriction of Bayesian updating in the context of subjective beliefs.

4 General Model

We now present the general model. Let S be a finite state space and X be a finite set of prizes. The set of all beliefs is ΔS while payoffs are lotteries in ΔX. Following Anscombe and Aumann (1963), we model choice options as *acts*, that is, mappings $f : S \to \Delta X$. Let H be the set of all acts. Call a finite set of acts a *menu* and let $\mathcal{K}$ be the set of all menus.

First, the pre-signal choice data of an agent (e.g. the lender) is a preference relation $\succeq$ over acts. Given each state $s \in S$, let $u_s : \Delta X \to \mathbb{R}$ denote a von Neumann-Morgenstern (vNM) utility in that state. We call a vector of non-constant vNM utilities a *utility function* $u = (u_s)_{s \in S}$. For $p \in \Delta S$ and $f \in H$, we use the simplifying notation

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

Let (p, u) denote a non-degenerate belief $p \in \operatorname{int}(\Delta S)$ and utility u. The following is the classic subjective expected utility model.

Definition. $\succeq$ is *represented by* (p, u) if $f \succeq g$ iff $p \cdot (u \circ f) \geq p \cdot (u \circ g)$.

We now specify the post-signal choice data of the agent. Let ΔH be the set of all probability measures on H. The post-signal choice data is a mapping $\rho : \mathcal{K} \to \Delta H$ such that

[10] In practice, if the regulator has the choice of whether to elicit the pre-signal or post-signal data first, then eliciting the latter first has the advantage that a lender may learn the true distribution of signals.

$\rho_F(F) = 1$ for all $F \in \mathcal{K}$. Given an act $f \in F$, $\rho_F(f)$ is the probability that f is chosen in the menu F. When $F = \{f, g\}$ is binary, we use the shorthand notation $\rho(f, g) = \rho_F(f)$. For example, if f and g correspond to approving and not approving a loan respectfully, then $\rho(f, g)$ is exactly the approval rate. We call ρ a *random choice function (RCF)*.

We model signal structures in the canonical sense where a *signal distribution* is simply a distribution on ΔS. We say a signal distribution μ is *generic* if signals are full-dimensional and informative about all states.

Definition. μ is *generic* if $\mu\{q \in \Delta S \mid q \cdot w = 0\} < 1$ for all $w \neq 0$.

Let (μ, u) denote a generic signal distribution μ and utility function u. We say ρ is represented by (μ, u) if it is consistent with the probabilistic choice of an agent with a (possibly) state-dependent utility function u and a signal distribution μ.[11]

Definition. ρ is *represented by* (μ, u) if

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

The representations above for both the pre-signal and post-signal primitives are not unique. We now define the joint primitive $(\succeq, \rho)$ and specify when it is consistent with Bayesian updating.

Definition (Bayesian Representation)**.** $(\succeq, \rho)$ is *represented by* (μ, u) if $\succeq$ is represented by (p, u) and ρ is represented by (μ, u) where

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

The Bayesian representation carries two implicit assumptions. First, the agent updates via Bayes' rule with respect to his subjective signals. This is captured by the martingale property for beliefs that relate p and μ. Second, signals only affect beliefs and not the (possibly) state-dependent utilities. This is reasonable if we assume tastes are somewhat stable while beliefs are updated with the arrival of new information. As we will show later in the axiomatization characterization, both assumptions yield empirical predictions.

[11] Technically, we need an additional restriction (regularity) on μ in order for this representation to be well-defined (see axiomatization in Section 6). Nevertheless, our main result holds regardless of whether μ is regular or not as long as restrict the representation to menus that do not have ties. For instance, full information corresponds to a μ that is generic but not regular and our main result applies.

We now present our main result. Although the individual representations for $\succeq$ and ρ are not unique, the Bayesian representation for the joint primitive $(\succeq, \rho)$ is. By looking at both pre-signal *and* post-signal choice data, the analyst can uniquely identify the agent's beliefs, information structure and state-dependent utilities. This provides a foundation for beliefs within the classic revealed preference methodology.

Theorem 1. *Suppose $(\succeq, \rho)$ and $(\trianglerighteq, \tau)$ are represented by (μ, u) and (ν, v) respectively. Then the following are equivalent*

(1) *$f \succeq g$ iff $f \trianglerighteq g$ and $\rho(f,g) = \tau(f,g)$*

(2) $(\succeq, \rho) = (\trianglerighteq, \tau)$

(3) *$(\mu, u) = (\nu, \alpha v + \beta)$ for $\alpha > 0$ and $\beta \in \mathbb{R}^S$*

Proof. See Appendix A.1. □

For a brief sketch of the proof, suppose there are only two states s_1 and s_2 as in the motivating example. Let $(\succeq, \rho)$ be represented by (μ, u). Consider constructing an alternate Bayesian representation of $(\succeq, \rho)$ as follows. For any $\lambda > 0$, define the adjusted utility v_λ and the adjusted belief $\phi_\lambda(q)$ where

$$v_\lambda := \left(\frac{u_1}{\lambda},\ u_2\right)$$
$$\phi_\lambda(q) := \frac{q\lambda}{q\lambda + (1-q)}$$

Note that whenever $q \cdot (u \circ f) \geq q \cdot (u \circ g)$, $\phi_\lambda(q) \cdot (v_\lambda \circ f) \geq \phi_\lambda(q) \cdot (v_\lambda \circ g)$. Thus, since ρ is represented by (μ, u), it must also be represented by $\left(\mu \circ \phi_\lambda^{-1}, v_\lambda\right)$. By the same argument, since $\succeq$ is represented by (p, u), it is also represented by $(\phi_\lambda(p), v_\lambda)$. The final step is to check whether Bayes' rule holds for any of these alternate representations. First, consider $\lambda > 1$ in which case ϕ_λ is strictly concave. By Jensen's,

$$\phi_\lambda(p) = \phi_\lambda\left(\int_{[0,1]} q\ \mu(dq)\right) > \int_{[0,1]} \phi_\lambda(q)\,\mu(dq)$$

so $\lambda > 1$ is inconsistent with a Bayesian representation. The case for $\lambda < 1$ is symmetric. Thus, the unique Bayesian representation is when $\lambda = 1$. Intuitively, Bayes' rule provides just enough of a restriction on beliefs so that the effects of beliefs and utilities can be properly disentangled and identified.

Note that beliefs are now more than merely a convenient modeling device but carry real empirical meaning. For instance, the standard normalization that we usually assume in order to obtain state-independent utilities under state-independent preferences is no longer empirically vacuous. The fact that scaling utilities have empirical meaning now implies that we can also make counterfactual welfare implications for the agent. Statements such as "the agent's utility is twice as much in one state than another" are now empirically testable.

Two remarks are in order. First, the genericity assumption is important. For example, if the signal is completely uninformative in terms of distinguishing states s_1 and s_2, then there will be an extra dimension of indeterminacy. In terms of the random choice function, this indeterminacy will manifest itself as ties within a certain subset of acts. As long as there is some randomness in choice for all non-degenerate acts[12] however, genericity will be satisfied and Theorem 1 applies.

The other restriction is that preferences must be non-constant. Hence, as in the classic models, null states continue to pose a problem. The question remains whether it is possible to enrich the primitive in a manner that accommodates null events and still allow for identification. We leave this as an open question for future research. Note that for both of these remarks, we can redefine the state space in such a way where these issues do not arise and we can achieve identification on a smaller state space.

5 The Empirical Content of Beliefs and Tastes

In this section, we compare the empirical content of our model with those using other different primitives. This is a useful exercise because it allows us to distinguish between models in which beliefs actually have *empirical meaning* (such as in our Bayesian representation) versus those where beliefs cannot be uniquely identified and hence are ill-specified from the perspective of strict revealed preference methodology.[13] This informs the analyst about the types of questions where beliefs actually matter and what data to collect if he is interested

[12] One can of course define a Bayesian representation without a generic μ, in which case this would be the observable implication of genericity.

[13] This is very much in the operationalist spirit of Savage (1954): "If the state of mind in question is not capable of manifesting itself in some sort of extraverbal behavior, it is extraneous to our main interest. If, on the other hand, it manifests itself through more material behavior, that should, at least in principle, imply the possibility of testing whether a person holds one event to be more probable than another, by some behavior expressing, and giving meaning to, his judgement." Of course, Savage may not be as stringent as this passage suggests (see discussion in Karni and Mongin (2000)).

in identifying beliefs. In what follows, we present four models with different primitives; in the first three, beliefs do not have empirical meaning while in the last one, they do.

First, suppose the analyst only observes the post-signal random choice. We show that this is insufficient for identifying beliefs. We say a signal distribution μ *has prior* p if its average belief is p.

Lemma 1. *If ρ is represented by (μ, u), then for any $r \in int(\Delta S)$, it is also represented by some (ν, v) where ν has prior r.*

Proof. Fix some $\alpha \in \mathbb{R}^S$ such that $\alpha_s > 0$ for all $s \in S$, and define $\phi : \Delta S \to \Delta S$ such that

$$(\phi \circ q)(s) := \frac{q(s)\,\alpha_s}{q \cdot \alpha}$$

Define utility v such that $v_s := \frac{u_s}{\alpha_s}$ for all $s \in S$. Suppose ρ is represented by (μ, u) so

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

where $\nu := \mu \circ \phi^{-1}$. Hence, (ν, v) also represents ρ. Since we are free to choose α, we can set ϕ so that ν has prior r. $\square$

In other words, for any RCF that is represented by a signal distribution and state-dependent utility, we can always find an alternative signal distribution and state-dependent utility that represent the same RCF. Moreover, this alternate signal distribution can have any prior as long as it shares the same support as the initial prior. It is precisely in this sense that beliefs have no empirical meaning when the data only consists of random choice.

What if we were to consider preferences over menus? Define the pre-signal menu preference relation over $\mathcal{K}$ as follows.

Definition. $\succeq$ is *represented by* (μ, u) if it is represented by

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

Note that this is exactly the subjective learning model of Dillenberger, Sadowski, Lleras, Takeoka (2014a) but allowing for state-dependent utilities. Over singleton acts, this agrees exactly with the state-dependent subjective expected utility representation from before. The following lemma shows that beliefs cannot be identified using menu choice in this model.

Lemma 2. *If $\succeq$ is represented by (μ, u), then for any $r \in int(\Delta S)$, it is also represented by some (ν, v) where ν has prior r.*

Proof. Fix some $\alpha \in \mathbb{R}^S$ such that $\alpha_s > 0$ for all $s \in S$, and define ϕ and v as in Lemma 1 above. Suppose $\succeq$ is represented by (μ, u) so

$$\begin{aligned} V(F) &= \int_{\Delta S} \sup_{f \in F} \sum_s q(s)\, u_s(f(s))\, \mu(dq) \\ &= \int_{\Delta S} \sup_{f \in F} (q \cdot \alpha) \sum_s (\phi \circ q)(s)\, v_s(f(s))\, \mu(dq) \\ &= \int_{\Delta S} \left(\sup_{f \in F} \phi(q) \cdot (v \circ f) \right) (q \cdot \alpha)\, \mu(dq) \end{aligned}$$

Define the measure $\hat{\mu}(A) := \frac{\int_A (q \cdot \alpha) \mu(dq)}{\int_{\Delta S} (q \cdot \alpha) \mu(dq)}$ and $\nu := \hat{\mu} \circ \phi^{-1}$, then

$$\int_{\Delta S} \sup_{f \in F} q \cdot (v \circ f)\, \nu(dq) = \int_{\Delta S} \sup_{f \in F} \phi(q) \cdot (v \circ f)\, \hat{\mu}(dq) = \frac{V(F)}{p \cdot \alpha}$$

where $p := \int_{\Delta S} q \; \mu(dq)$. Again (ν, v) represents $\succeq$ and we are free to choose α such that ν has prior r. □

As in Lemma 1, alternate signal distributions can essentially have any prior. However, the transformation $\mu \circ \phi^{-1}$ in Lemma 1 is different than the transformation $\hat{\mu} \circ \phi^{-1}$ in Lemma 2. In other words, the transformation used to obtain an equivalent representation under random choice is different from that used to obtain an equivalent representation under menu choice. Since our pre-signal preference relation $\succeq$ over acts is a special case of the menu choice primitive, the same applies when comparing the pre-signal and post-signal choice data from Theorem 1. Given either set of data, one can freely change μ and u to obtain the same representation. Once restricted to both data sets however, we lose the flexibility to freely change μ and u *and* still satisfy the martingale property from Bayes' rule. This property provides just enough of a restriction so that we can identify beliefs.

In light of Theorem 1, random choice must convey more empirical content than menu choice. Otherwise, one could achieve identification simply with menu choice since menu choice already includes the pre-signal preference relation over acts as a special case. Note that in the special case when utilities are state-independent, both menu and random choice convey the same empirical content (see Lu (2015)).

Now, suppose the analyst actually knows the signal structure of the agent. Can the analyst always achieve identification? Consider a pre-signal preference $\succeq$ over menus and a family of post-signal preference relations $\succeq_\theta$ one for each signal realization $\theta \in \Theta$. Denote this family of preference relations by $\succeq_\Theta := (\succeq_\theta)_{\theta\in\Theta}$. We model the signal structure as a vector of conditional signal distributions L where $L(s)$ is a distribution on Θ conditional on $s \in S$. Suppose the analyst knows L. We say an event occurs L-a.s. if it occurs $\sum_s p(s) L(s)$-a.s. for any non-degenerate $p \in \Delta S$.

Definition. $(\succeq, \succeq_\Theta, L)$ is *represented by* (μ, u) if there are $p \in \Delta S$, $Q : \Theta \to \Delta S$ and a measure ξ on Θ where

(1) $\mu = \xi \circ Q^{-1}$ where $\xi(d\theta) Q(\theta)(s) = p(s) L(s, d\theta)$

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

(3) $\succeq_\theta$ is represented by $(Q(\theta), u)$ L-a.s..

The following result generalizes the concluding lemma in Karni, Schmeidler and Vind (1983) to choice over menus with arbitrary signal spaces. Beliefs cannot be identified using conditional preference relations even when the pre-signal preference relation over menus is known.

Lemma 3. *If $(\succeq, \succeq_\Theta, L)$ is represented by (μ, u), then for any $r \in int(\Delta S)$, it is also represented by some (ν, v) where ν has prior r.*

Proof. Fix some $\alpha \in \mathbb{R}^S$ such that $\alpha_s > 0$ for all $s \in S$, and define ϕ and v as in Lemma 1. Suppose $(\succeq, \succeq_\Theta, L)$ is represented by (μ, u) so $p = \int_{\Delta S} q\, \mu(dq)$. Let $r := \phi(p)$ and define the measure π on Θ and $J : \Theta \to \Delta S$ such that for all $s \in S$,

$$\pi(d\theta) J(\theta)(s) = r(s) L(s, d\theta)$$

Let $\nu := \pi \circ J^{-1}$ and note that

$$\int_{\Delta S} q(s) \nu(dq) = \int_\Theta J(\theta)(s) \pi(d\theta) = \int_\Theta r(s) L(s, d\theta) = r(s)$$

so $\int_{\Delta S} q\ \nu(dq) = r = \phi(p)$. From the proof of Lemma 2 above, we have that

$$\begin{aligned} V(F) &= \int_{\Delta S} \left(\sup_{f \in F} \phi(q) \cdot (v \circ f) \right) (q \cdot \alpha)\, \mu(dq) \\ &= \int_{\Theta} \left(\sup_{f \in F} \phi(Q(\theta)) \cdot (v \circ f) \right) (Q(\theta) \cdot \alpha)\, \xi(d\theta) \\ &= \int_{\Theta} \left(\sup_{f \in F} J(\theta) \cdot (v \circ f) \right) \pi(d\theta) = \int_{\Delta S} \left(\sup_{f \in F} q \cdot (v \circ f) \right) \nu(dq) \end{aligned}$$

Thus, $\succeq$ is represented by both (μ, u) and (ν, v). Moreover, for each $\theta \in \Theta$, $f \succeq_\theta g$ iff

$$\begin{aligned} Q(\theta) \cdot (u \circ f) &\geq Q(\theta) \cdot (u \circ g) \\ \sum_s p(s)\, L(s, d\theta)\, u_s(f(s)) &\geq \sum_s p(s)\, L(s, d\theta)\, u_s(g(s)) \\ \sum_s r(s)\, L(s, d\theta)\, v_s(f(s)) &\geq \sum_s r(s)\, L(s, d\theta)\, v_s(g(s)) \\ J(\theta) \cdot (v \circ f) &\geq J(\theta) \cdot (v \circ g) \end{aligned}$$

Hence (ν, v) also represents $(\succeq, \succeq_\Theta, L)$ and we can choose α such that ν has prior r. □

Note that in this case, the transformation used to obtain an equivalent representation for the pre-signal preference relation $\succeq$ is exactly the same as that used to obtain an equivalent representation for all the conditional preference relations $\succeq_\theta$. Hence, beliefs cannot be uniquely identified. The following demonstrates this in the motivating example.

Motivating Example (continued). Consider the loan approval application from Section 3. Suppose now that the regulator observes the signal structure $L(\theta_i | s_i) = \frac{3}{4}$ for $i \in \{1, 2\}$. Recall that the lender not receiving kickbacks has prior $p = \frac{1}{2}$ and posteriors $(q_1, q_2) = \left(\frac{3}{4}, \frac{1}{4}\right)$. Now, consider a lender who receives kickbacks but has prior $\hat{p} = \frac{1}{4}$ and note that

$$\hat{p}\left(\frac{3}{2}\right) + (1 - \hat{p})\left(-\frac{1}{2}\right) = 0$$

Thus, for applicants in the pre-signal group, this lender is also indifferent between approving or not. The lender's posterior beliefs given the high and low credit scores respectively are

$$(\hat{q}_1, \hat{q}_2) = \left(\frac{1}{2}, \frac{1}{10}\right)$$

Note that for $i \in \{1,2\}$,

$$\hat{q}_i = \frac{q_i \left(\frac{1}{3}\right)}{q_i \left(\frac{1}{3}\right) + (1 - q_i)}$$

This means that $q_i + (1 - q_i)(-1) \geq 0$ iff $\hat{q}_i \left(\frac{3}{2}\right) + (1 - \hat{q}_i)\left(-\frac{1}{2}\right) \geq 0$ so both lenders have the same post-signal conditional preferences. In other words, even though the regulator knows the pre-signal preferences, the conditional preferences and the signal structure, he still can not distinguish between a lender with prior $p = \frac{1}{2}$ and not receiving kickbacks versus a lender prior $\hat{p} = \frac{1}{4}$ and receiving kickbacks.

The above results have the following implication. Consider an analyst faced with the following two questions: (1) what is the probability that the agent will choose a specific option, and (2) which option will the agent choose conditional on a specific signal realization. Theorem 1 and Lemma 3 imply that identifying priors is important for answering the first question but not the second. What this means is that in the case of the latter, the analyst can simply choose any prior consistent with the agent's pre-signal preferences and use that to infer what the agent's conditional preferences will be. In particular, this means that an analyst who knows the agent's signal structure and his pre-signal preferences is able to answer question (2) but not (1).

Finally, suppose that the analyst also observes the agent's random choice as well.

Definition. (ρ, L) is *represented by* (μ, u) if there are $p \in \Delta S$, $Q : \Theta \to \Delta S$ and a measure ξ on Θ where

(1) $\mu = \xi \circ Q^{-1}$ where $\xi(d\theta)\, Q(\theta)(s) = p(s)\, L(s, d\theta)$

(2) ρ is represented by (μ, u)

Given both the signal structure and random choice, the analyst can now pin down both beliefs and utilities.

Lemma 4. *Suppose (ρ, L) and (τ, L) are represented by (μ, u) and (ν, v) respectively. Then $\rho = \tau$ iff $(\mu, u) = (\nu, \alpha v + \beta)$ for $\alpha > 0$ and $\beta \in \mathbb{R}^S$.*

Proof. See Appendix A.2. □

Lemma 4 implies that two agents with different priors but faced with the same signal structure will exhibit different random choice. Hence, an analyst can use random choice to differentiate two agents with different priors *provided that* both agents are receiving the same signals.

Figure 1 below summarizes our results. It shows how random choice in junction with either the pre-signal preference or the signal structure contains just enough data for identification. Moreover, Lemma 3 suggests that from the analyst's perspective, the data contained in the agent's random choice is richer than simply knowing his signal structure. Note that the unshaded regions represents areas where even if the analyst knew all the relevant data, he would still not be able to identify beliefs and tastes. Hence, beliefs have no empirical meaning in those regions. Finally, note that if the analyst knows the pre-signal preferences, post-signal random choice and also the signal structure, then there is over-identification and the analyst can check for internal inconsistencies on the part of the agent.

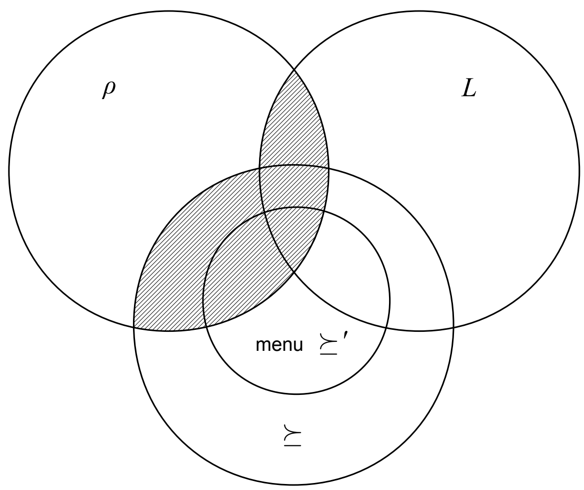

Figure 1: Empirical Content of Beliefs and Tastes. This figure shows the different combinations of data: ρ is post-signal random choice, L is the signal structure, $\succeq$ is the pre-signal preference over acts and $\succeq'$ is the pre-signal preference over menus. Shaded regions correspond to combinations of data where beliefs and utilities can be uniquely identified.

6 Characterization and Elicitation of Beliefs and Tastes

In this section, we provide an axiomatic characterization of our representation. This is important for two reasons. First, it allows us to test the choice-theoretic implications of our Bayesian representation. Second, since beliefs have meaning in our model, the axioms provide insights on how to actually elicit these beliefs (along with utilities) from choice data.

First, recall the axioms for a subjective expected utility representation under the Anscombe-Aumann model. Collectively, we call the axioms below Axiom 1.

Axiom 1.1. $\succeq$ *is a preference relation.*

Axiom 1.2. (Independence) $f \succeq g$ *implies* $af + (1-a)h \succ ag + (1-a)h$ *for* $a \in (0,1)$.

Axiom 1.3. (Continuity 1) $f \succ h \succ g$ *implies there are* $\{a, b\} \subset (0,1)$ *such that* $af + (1-a)g \succ h \succ bf + (1-b)g$.

Axiom 1.4. (Non-degeneracy) *For all* $s \in S$, $f_s \succ g_s$ *for some* f *and* g.

Proposition 1. $\succeq$ *satisfies Axiom 1 iff it is represented by* (p, u).

Proof. See Proposition 7.4 of Kreps (1988). □

We now introduce the standard axioms for the post-signal RCF. Let $\mathcal{K}_0$ denote the set of menus that do not include ties. Let $\text{ext}F$ denote the extreme acts of the menu $F \in \mathcal{K}$. Collectively, we call the axioms below Axiom 2.

Axiom 2.1. (Monotonicity) $G \subset F$ *implies* $\rho_G(f) \geq \rho_F(f)$.

Axiom 2.2. (Linearity) $\rho_F(f) = \rho_{aF+(1-a)g}(af + (1-a)g)$ *for* $a \in (0,1)$.

Axiom 2.3. (Continuity 2) $\rho : \mathcal{K}_0 \to \Delta H$ *is continuous.*[14]

Axiom 2.4. (Extremeness) $\rho_F(\text{ext}F) = 1$.

We say μ is *regular* if all $w \neq 0$, the μ-measure that $q \cdot w = 0$ is either zero or one.[15]

Proposition 2. ρ *satisfies Axiom 2 iff it is represented by a regular* (μ, u).

Proof. See Lu (2015). □

We now introduce the axiom connecting $\succeq$ with ρ that is necessary for Bayesian updating. Let $\underline{h} \in H$ be some worst act as guaranteed by the axioms.

Definition. Fix some $h \in H$. Then for any $f \in H$ such that $\rho(h, f) = 1$, define

$$f_{\rho,h}(a) := \rho\left(f,\ a\underline{h} + (1-a)h\right)$$

[14] We endow $\mathcal{K}_0$ with the Hausdorff metric and ΔH with the topology of weak convergence.
[15] More formally, for all $w \neq 0$, $\mu\{q \in \Delta S \mid q \cdot w = 0\} \in \{0, 1\}$.

Each $h \in H$ induces a preference relation $\succeq_{\rho,h}$ as follows. First, for $\rho(h, f) = \rho(h, g) = 1$, we have $f \succeq_{\rho,h} g$ if the integral of $f_{\rho,h}$ is greater than the integral of $g_{\rho,h}$. Since $\succeq_{\rho,h}$ is linear, we can then extend $\succeq_{\rho,h}$ to all acts. We say an act h^* is *calibrating* if its induced preference relation $\succeq_{\rho,h^*}$ coincides exactly with $\succeq$.

Definition. h^* is *calibrating* if $\succeq_{\rho,h^*} = \succeq$

As it turns out, Bayesian consistency requires there to exist a calibrating act. How do we find such an act? First, we define some notation. For each $h \in H$, let

$$B_{\rho,h} := \{ f \in H \mid f \succeq_{\rho,h} h \}$$

be the set of acts that are $\succeq_{\rho,h}$-better than h. We construct a sequence of acts $(h_i)_{i\in\mathbb{N}}$ as follows. First, choose some initial $h_0 \in H$. Let H_0 be the set of acts that are $\succeq$-indifferent to h_0. Now, recursively define

$$H_{i+1} := B_{\rho,h_i} \cap H_i$$

and let $h_{i+1} \in H_{i+1}$. If no such h_{i+1} exists, then set $h_{i+1} = h_i$. Given any sequence h_i, $H_{i+1} \subset H_i$ is a monotonically decreasing sequence of sets. Hence, by the monotone convergence theorem, it will always converge and we let H^* denote its limit. We say a sequence h_i is *diminishing* if $\dim(H^*) < \dim(H_0)$. In other words, the acts are chosen such that H^* is reduced to a lower-dimensional set. As long as each h_{i+1} is chosen far from the boundaries of each H_{i+1}, the resulting sequence will always be diminishing. The result below shows that we can always use this procedure to elicit the calibrating act.

Proposition 3. *Let $\succeq$ and ρ be represented by (p, u) and (μ, u) respectively. Then the following are equivalent*

(1) *There exists a calibrating act h^**

(2) *Any diminishing h_i converges to a non-degenerate h^**

(3) *$(\succeq, \rho)$ has a Bayesian representation*

Proof. See Appendix A.3. □

Proposition 3 not only provides a procedure for eliciting a calibrating act if one exists, it also shows the behavioral implications on the primitive $(\succeq, \rho)$. We can now define the axiom for Bayesian consistency as follows.

Axiom 3. (Bayesian Consistency) *There exists a calibrating act.*

Finally, the last condition we need is to require that taste preferences are not affected by information.

Axiom 4. (Static Tastes) *Suppose for all $g \in F$, $f \succeq g$ and $f(s) \neq g(s)$ at most one $s \in S$. Then $f \in F$ implies $\rho_F(f) = 1$.*

We now finally ready to state our representation result. We say a Bayesian representation (μ, u) is regular if the signal distribution μ is regular.

Theorem 2. *$(\succeq, \rho)$ satisfies Axioms 1-4 iff it has a regular Bayesian representation.*

Proof. Follows directly from Propositions 1-3. □

As we mentioned before, the Bayesian representation assumes that (1) the agent updates via Bayes' rule and (2) signals only affect beliefs not tastes. Bayesian Consistency (BC) addresses the first assumption by providing a testable restriction for when Bayes' rule is satisfied. Put differently, whenever BC holds, the analyst can rationalize the agent's behavior "as if" he is using Bayes' rule.

The second assumption is addressed by Static Tastes (ST). If ST is violated, then signals must be directly affecting tastes. Hence, ST provides the empirical content for the second assumption. However, the argument here is slightly more subtle; even if ST is satisfied, by scaling both beliefs *and* utilities stochastically in a careful way, it still may be possible to reinterpret the model as one where signals affect both beliefs and tastes. Hence, implicit in the Bayesian representation is the assumption that all the randomness in choice is attributed to changes in beliefs rather than tastes. This assumption is both useful and reasonable. It is useful because it provides a methodology to deal with choice data when agents have state-dependent utilities while the standard subjective expected models are silent on what to do. It is reasonable because in many applications, such as the lender example, one would expect credit scores to not directly impact the lender's tastes. Thus, similar to normalizing utilities in the classic Anscombe-Aumann model under state-independent preferences, the normalization here allows us to essentially define beliefs as the stochastic component of decision-making with information. This exactly captures what Karni, Schmeidler and Vind

(1983) describe as "the distinction between the transitory nature of probabilistic beliefs and the unchanging nature of tastes".[16]

[16] In fact, this assumption is relatively weak as even with rich primitives, identification is not always possible under this assumption (see Lemmas 1-3). Alternatively, one could introduce some discipline in the joint distribution of belief and taste shocks (e.g. independence) or enrich the primitive (e.g. the analyst observes choice data for two agents receiving the same signal) and still obtain identification.

References

Ahn, D. and T. Sarver (2013): "Preference for Flexibility and Random Choice," *Econometrica*, 81, 341–361.

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

Brunnermeier, M., A. Simsek, and W. Xiong (2014): "A Welfare Criterion of Models with Distorted Beliefs," *Quarterly Journal of Economics*, 129, 1711–1752.

de Finetti, B. (1937): "La Prévision: Ses Lois Logic, Ses Sources Subjectives," *Annals de I'Institute Henri Poincaré*, 7, 1–68.

Dillenberger, D., J. Lleras, P. Sadowski, and N. Takeoka (2014a): "A Theory of Subjective Learning," *Journal of Economic Theory*, 153, 287–312.

Dillenberger, D., A. Postlewaite, and K. Rozen (2014b): "Optimism and Pessimism with Expected Utility," Mimeo.

Drèze, J. (1987): "Decision Theory with Moral Hazard and State-Dependent Preferences," in *Essays on Economic Decisions Under Uncertainty*, Cambridge University Press.

Drèze, J. and A. Rustichini (1999): "Moral Hazard and Conditional Preferences," *Journal of Mathematical Economics*, 31.

Fishburn, P. (1973): "A Mixture-Set Axiomatization of Conditional Subjective Expected Utility," *Econometrica*, 41, 1–25.

Gilboa, I., L. Samuelson, and D. Schmeidler (2014): "No-Betting-Pareto Dominance," *Econometrica*, 82, 1405–1442.

Goldin, C. and C. Rouse (2000): "Orchestrating Impartiality: The Impact of "Blind" Auditions on Female Musicians," *American Economic Review*, 90, 715–741.

Karni, E. (1993): "A Definition of Subjective Probabilities with State-dependent preferences," *Econometrica*, 61, 187–198.

——— (2006): "Subjective expected utility theory without states of the world," *Journal of Mathematical Economics*, 42, 325–342.

——— (2007): "Foundations of Bayesian Theory," *Journal of Economic Theory*, 132, 167–188.

Karni, E. and P. Mongin (2000): "On the Determination of Subjective Probability by Choices," *Management Science*, 46, 233–248.

Karni, E. and Z. Safra (2016): "A Theory of Stochastic Choice Under Uncertainty," *Journal of Mathematical Economics*, 63, 164–173.

Karni, E. and D. Schmeidler (2016): "An Expected Utility Theory for State-Dependent Preferences," *Theory and Decision*, forthcoming.

Karni, E., D. Schmeidler, and K. Vind (1983): "On State Dependent Preferences and Subjective Probabilities," *Econometrica*, 51, 1021–1031.

Kreps, D. (1988): *Notes on the Theory of Choice*, Westview Press.

Krishna, R. and P. Sadowski (2014): "Dynamic Preference for Flexibility," *Econometrica*, 82, 655–703.

Lu, J. (2015): "Random Choice and Private Information," Mimeo.

Luce, R. and D. Krantz (1971): "Conditional Expected Utility," *Econometrica*, 39, 253–271.

Masatlioglu, Y., D. Nakajima, and E. Ozdenoren (2014): "Revealed Willpower," Mimeo.

Ramsey, F. (1931): "Truth and Probability," in *The Foundations of Mathematics and Other Logical Essays*, Routledge and Kegan Paul Ltd.

Sadowski, P. (2013): "Contingent Preference for Flexibility: Eliciting Beliefs from Behavior," *Theoretical Economics*, 8, 503–534.

Savage, J. (1954): *The Foundations of Statistics*, John Wiley and Sons, Inc.

Schenone, P. (2016): "Identifying Subjective Beliefs in Subjective State Space Models," *Games and Economic Behavior*, 95, 59–72.

Skiadas, C. (1997): "Subjective Probability under Additive Aggregation of Conditional Preferences," *Journal of Economic Theory*, 76, 242–271.

A Appendix: Proofs

A.1 Proof of Theorem 1

We prove the main identification theorem. First, consider the following technical lemma.

Lemma 5. *If $\succeq$ is represented by (p,u) and (r,v), then for all $s \in S$, $u_s = \alpha_s v_s + \beta_s$ for $\alpha_s > 0$.*

Proof. For $f \in H$ and $s \in S$, define $f_{s,x} \in H$ such that $f_{s,x}(s') = f(s')$ for all $s' \neq s$ and $f_{s,x}(s) = x \in \Delta X$. Now, $f_{s,x} \succeq f_{x,y}$ iff $u_s(x) \geq u_s(y)$ iff $v_s(x) \geq v_s(y)$. Thus, $u_s = \alpha_s v_s + \beta_s$ for $\alpha_s > 0$ by the uniqueness properties of expected utility. □

Let $(\succeq, \rho)$ and $(\trianglerighteq, \tau)$ are represented by (μ, u) and (ν, v) respectively. We will prove that the following are equivalent:

(1) $f \succeq g$ iff $f \trianglerighteq g$ and $\rho(f,g) = \tau(f,g)$

(2) $(\succeq, \rho) = (\trianglerighteq, \tau)$

(3) $(\mu, u) = (\nu, \alpha v + \beta)$ for $\alpha > 0$ and $\beta \in \mathbb{R}^S$

We will show that (3) implies (2) implies (1) implies (3). If (3) is true, then

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

so $(\succeq, \rho) = (\trianglerighteq, \tau)$ follows immediately and (2) is true. That (2) implies (1) is trivial. Thus, what remains is to show that (1) implies (3).

Suppose (1) is true so $\succeq \; = \; \trianglerighteq$ is represented by (p,u) and (r,v). Hence, $f \succeq g$ iff $p \cdot (u \circ f) \geq p \cdot (u \circ g)$ iff $r \cdot (v \circ f) \geq r \cdot (v \circ g)$ where $p = \int_{\Delta S} q \; \mu(dq)$ and $r = \int_{\Delta S} q \; \nu(dq)$. By Lemma 5, for all $s \in S$, $u_s = \alpha_s v_s + \beta_s$ for $\alpha_s > 0$. Define $\phi : \Delta S \to \Delta S$ such that

$$\phi(q)(s) := \frac{q(s)\,\alpha_s}{q \cdot \alpha}$$

so

$$\begin{aligned} p \cdot (u \circ f) = \sum_s p(s)\, u_s(f(s)) &= \sum_s p(s)\,\alpha_s v_s(f(s)) + \sum_s p(s)\,\beta_s \\ &= (p \cdot \alpha) \sum_s \phi(p)(s)\, v_s(f(s)) + p \cdot \beta \\ &= (p \cdot \alpha)\left[\phi(p) \cdot (v \circ f)\right] + p \cdot \beta \end{aligned}$$

Thus, $f \succeq g$ iff $r \cdot (v \circ f) \geq r \cdot (v \circ g)$ iff $\phi(p) \cdot (v \circ f) \geq \phi(p) \cdot (v \circ g)$.

For each $s \in S$, let $\{\underline{x}_s, \bar{x}_s\} \subset \Delta X$ be such that $\underline{v}_s = v_s(\underline{x}_s) \leq v_s(y) \leq v_s(\bar{x}_s) = \bar{v}_s$ for all $y \in \Delta X$. Note that $\underline{v}_s < \bar{v}_s$ for all $s \in S$ due to non-constantness. Define $\{\underline{f}, \bar{f}\} \subset H$ such that $\underline{f}(s) = \underline{x}_s$ and $\bar{f}(s) = \bar{x}_s$ for all $s \in S$. Hence for all $f \in H$, there is some $a \in [0,1]$ such that $f \sim \underline{f} a \bar{f}$. Now,

$$\phi(p) \cdot (v \circ f) = a\phi(p) \cdot \left(v \circ \underline{f}\right) + (1-a)\phi(p) \cdot \left(v \circ \bar{f}\right)$$
$$a = \frac{\phi(p) \cdot \left(v \circ \bar{f} - v \circ f\right)}{\phi(p) \cdot \left(v \circ \bar{f} - v \circ \underline{f}\right)}$$

so by symmetric argument, for all $f \in H$,

$$\frac{\phi(p) \cdot \left(v \circ \bar{f} - v \circ f\right)}{\phi(p) \cdot \left(v \circ \bar{f} - v \circ \underline{f}\right)} = \frac{r \cdot \left(v \circ \bar{f} - v \circ f\right)}{r \cdot \left(v \circ \bar{f} - v \circ \underline{f}\right)}$$

If we let $k := \frac{\phi(p) \cdot \left(v \circ \bar{f} - v \circ \underline{f}\right)}{r \cdot \left(v \circ \bar{f} - v \circ \underline{f}\right)}$, then $(\phi(p) - kr) \cdot \left(v \circ \bar{f} - v \circ f\right) = 0$ for all $f \in H$. Hence, $\phi(p) = kr$ so $\phi(p) = r$.

Note that

$$\begin{aligned}
\rho(f,g) &= \mu\{q \in \Delta S \mid q \cdot (u \circ f) \geq q \cdot (u \circ g)\} \\
&= \mu\{q \in \Delta S \mid \phi(q) \cdot (v \circ f) \geq \phi(q) \cdot (v \circ g)\} \\
&= \mu_\phi\{q \in \Delta S \mid q \cdot (v \circ f) \geq q \cdot (v \circ g)\} \\
&= \tau(f,g) = \nu\{q \in \Delta S \mid q \cdot (v \circ f) \geq q \cdot (v \circ g) \ \ \forall g \in F\}
\end{aligned}$$

where $\mu_\phi := \mu \circ \phi^{-1}$. Now, consider $f^* \in H$ such that $v_s(f^*_s) - v_s(\underline{x}_s) = v_{s'}(f^*_{s'}) - v_{s'}(\underline{x}_{s'}) := \varepsilon > 0$ for all $\{s, s'\} \subset S$. Note that this is possible as no v_s is constant. Hence,

$$\begin{aligned}
\rho\left(f, f^* a \underline{f}\right) &= \mu_\phi\left\{q \in \Delta S \;\middle|\; q \cdot (v \circ f) \geq aq \cdot (v \circ f^*) + (1-a)\, q \cdot v \circ \underline{f}\right\} \\
&= \mu_\phi\left\{q \in \Delta S \;\middle|\; q \cdot \left(v \circ f - v \circ \underline{f}\right) \geq aq \cdot \varepsilon \mathbf{1}\right\} \\
&= \mu_\phi\left\{q \in \Delta S \;\middle|\; q \cdot \frac{1}{\varepsilon}\left(v \circ f - v \circ \underline{f}\right) \geq a\right\} \\
&= \nu\left\{q \in \Delta S \;\middle|\; q \cdot \frac{1}{\varepsilon}\left(v \circ f - v \circ \underline{f}\right) \geq a\right\}
\end{aligned}$$

Consider $H^* \subset H$ such that $f \in H^*$ iff $v_s(f^*_s) \geq v_s(f_s) \geq v_s(\underline{x}_s)$ for all $s \in S$. Hence,

$0 \leq \frac{1}{\varepsilon}\left(v\left(f_s\right) - v\left(\underline{x}_s\right)\right) \leq 1$ for all $f \in H^*$ so

$$\mu_\phi \left\{q \in \Delta S \;\mid q \cdot \xi \geq a\right\} = \nu \left\{q \in \Delta S \;\mid q \cdot \xi \geq a\right\}$$

for all $\xi \in [0,1]^S$ and $a \in [0,1]$. By the Cramér–Wold Theorem, $\mu_\phi = \nu$. Hence

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

Given the definition of ϕ, this means that for all $s \in S$,

$$\begin{aligned}\frac{\int_{\Delta S} q\left(s\right) \mu\left(dq\right) \alpha_s}{\int_{\Delta S} q \;\mu\left(dq\right) \cdot \alpha} &= \int_{\Delta S} \left(\frac{q\left(s\right) \alpha_s}{q \cdot \alpha}\right) \mu\left(dq\right) \\ \frac{\int_{\Delta S} q\left(s\right) \mu\left(dq\right)}{\int_{\Delta S} q \;\mu\left(dq\right) \cdot \alpha} &= \int_{\Delta S} \left(\frac{q\left(s\right)}{q \cdot \alpha}\right) \mu\left(dq\right)\end{aligned}$$

where the second equality follows from the fact that $\alpha_s > 0$ for all $s \in S$. Summing up over all states, we have

$$\begin{aligned}\sum_s \frac{\int_{\Delta S} q\left(s\right) \mu\left(dq\right)}{\int_{\Delta S} q \cdot \alpha \;\mu\left(dq\right)} &= \sum_s \int_{\Delta S} \left(\frac{q\left(s\right)}{q \cdot \alpha}\right) \mu\left(dq\right) \\ \left(\int_{\Delta S} q \cdot \alpha \;\mu\left(dq\right)\right)^{-1} &= \int_{\Delta S} \left(q \cdot \alpha\right)^{-1} \mu\left(dq\right)\end{aligned}$$

By Jensen's, it must be that $q \cdot \alpha = \lambda$ μ-a.s. for some $\lambda > 0$.

Now, we can find some ε small enough such that

$$\varepsilon \frac{\alpha}{\lambda} = u \circ f - u \circ \underline{f}$$

for some $f \in H$ and $f^* \in H$ be such that $u_s\left(f^*_s\right) - u_s\left(\underline{x}_s\right) = u_{s'}\left(f^*_{s'}\right) - u_{s'}\left(\underline{x}_{s'}\right) := \varepsilon > 0$ for all $\{s, s'\} \subset S$. Now, note that μ-a.s.

$$q \cdot \left(u \circ f - u \circ \underline{f}\right) = q \cdot \left(\frac{\varepsilon}{\lambda} \alpha\right) = \varepsilon = q \cdot \varepsilon \mathbf{1} = q \cdot \left(u \circ f^* - u \circ \underline{f}\right)$$

Hence, $q \cdot (u \circ f) = q \cdot (u \circ f^*)$ μ-a.s. so $u \circ f = u \circ f^*$ by genericity. Thus,

$$\frac{\varepsilon}{\lambda} \alpha = u \circ f^* - u \circ \underline{f} = \varepsilon \mathbf{1}$$

so $\alpha = \lambda \mathbf{1}$. Hence, $\phi\left(q\right) = q$ so $p = r$ and $\mu = \nu$. Hence $(\mu, u) = (\nu, \alpha v + \beta)$ for $\alpha > 0$ and

$\beta \in \mathbb{R}^S$ proving (3).

A.2 Proof of Lemma 4

Necessity is trivial so we prove sufficiency. Let $\xi(d\theta) Q(\theta)(s) = p(s) L(s, d\theta)$ and $\pi(d\theta) J(\theta)(s) = r(s) L(s, d\theta)$ where $\mu = \xi \circ Q^{-1}$ and $\nu = \pi \circ J^{-1}$. Since $\rho = \tau$, by Lemma 5, we can assume that $u_s = \alpha_s v_s + \beta_s$ for $\alpha_s > 0$. For any $a \in \mathbb{R}^S_+$, define $\phi_a : \Delta S \to \Delta S$ such that

$$\phi_a(q)(s) := \frac{q(s) a_s}{q \cdot a}$$

Now,

$$\begin{aligned}
\rho_F(f) &= \mu \{ q \in \Delta S \mid q \cdot (u \circ f) \geq q \cdot (u \circ g) \ \forall g \in F \} \\
&= \mu \left\{ q \in \Delta S \;\middle|\; \sum_s q(s) (\alpha_s v_s (f(s))) \geq \sum_s q(s) (\alpha_s v_s (f(s))) \ \forall g \in F \right\} \\
&= \mu \{ q \in \Delta S \mid \phi_\alpha(q) \cdot (v \circ f) \geq \phi_\alpha(q) \cdot (v \circ g) \ \forall g \in F \} \\
&= \nu \{ q \in \Delta S \mid q \cdot (v \circ f) \geq q \cdot (v \circ g) \ \forall g \in F \}
\end{aligned}$$

where the last equality follows from the fact that $\rho = \tau$. By Cramer-Wold, this implies that $\mu \circ \phi_\alpha^{-1} = \nu$.

Next, note that

$$\pi(d\theta) = \xi(d\theta) \sum_s r_s \frac{Q(\theta)(s)}{p_s}$$

so

$$\xi(d\theta) \sum_s r_s \frac{Q(\theta)(s)}{p_s} J(\theta)(s) = \pi(d\theta) J(\theta)(s) = r(s) L(s, d\theta) = \xi(d\theta) r_s \frac{Q(\theta)(s)}{p_s}$$

Hence, $J(\theta) = \phi_{\frac{r}{p}}(Q(\theta))$ ξ-a.s. where we take $\frac{r}{p}$ to represent the vector with coordinates $\frac{r_s}{p_s}$ for all $s \in S$. Since $\mu \circ \phi_\alpha^{-1} = \nu$, we have that for any measurable $\hat{\psi} : \Delta S \to \mathbb{R}$,

$$\begin{aligned}
\int_{\Delta S} \hat{\psi}(\phi_\alpha(q)) \mu(dq) = \int_{\Delta S} \hat{\psi}(q) \nu(dq) &= \int_\Theta \hat{\psi}(J(\theta)) \pi(d\theta) \\
&= \int_\Theta \hat{\psi}\left(\phi_{\frac{r}{p}}(Q(\theta))\right) \sum_s r_s \frac{Q(\theta)(s)}{p_s} \xi(d\theta) \\
&= \int_{\Delta S} \hat{\psi}\left(\phi_{\frac{r}{p}}(q)\right) \sum_s q(s) \frac{r_s}{p_s} \mu(dq)
\end{aligned}$$

Letting $\hat{\psi} = \psi \circ \phi_{\frac{1}{\alpha}}$ for any measurable ψ yields

$$\int_{\Delta S} \psi(q)\,\mu(dq) = \int_{\Delta S} \psi\left(\phi_{\frac{1}{\alpha}}\left(\phi_{\frac{r}{p}}(q)\right)\right) \sum_s q(s)\,\frac{r_s}{p_s}\mu(dq)$$
$$= \int_{\Delta S} (q \cdot b)\,\psi(\phi_a(q))\,\mu(dq)$$

where $\{a, b\} \subset \mathbb{R}_+^S$ are such that $a_s = \frac{r_s}{p_s \alpha_s}$ and $b_s = \frac{r_s}{p_s}$ for all $s \in S$.

Now, note that

$$p = \int_{\Delta S} q\ \mu(dq) = \int_{\Delta S} (q \cdot b)\,\phi_a(q)\,\mu(dq) = \int_{\Delta S} \psi_1(q)\,\mu(dq)$$

where $\psi_1(q) := (q \cdot b)\,\phi_a(q)$. Now

$$(q \cdot b)\,\psi_1(\phi_a(q)) = (q \cdot b)\,(\phi_a(q) \cdot b)\,\phi_a(\phi_a(q))$$
$$= (q \cdot b)\,(\phi_a(q) \cdot b)\,\phi_{a^2}(q)$$

If we define $\psi_k := (q \cdot b)\,\psi_{k-1}(\phi_a(q))$, then

$$\psi_k(q) = (q \cdot b)\,(\phi_a(q) \cdot b) \cdots (\phi_{a^k}(q) \cdot b)\,\phi_{a^{k+1}}(q)$$

Moreover,

$$\int_{\Delta S} \psi_k(q)\,\mu(dq) = \int_{\Delta S} (q \cdot b)\,\psi_k(\phi_a(q))\,\mu(dq) = \int_{\Delta S} (q \cdot b)\,\psi_{k+1}(\phi_a(q))\,\mu(dq)$$

Hence, for all k,

$$p = \int_{\Delta S} \psi_k(q)\,\mu(dq) = \int_{\Delta S} \phi_{a^{k+1}}(q)\,(q \cdot b)\,(\phi_a(q) \cdot b) \cdots (\phi_{a^k}(q) \cdot b)\,\mu(dq)$$
$$= \int_{\Delta S} \phi_{a^{k+1}}(q)\,\mu_k(dq)$$

where $\mu_k(dq) := \mu(dq)\,(q \cdot b)\,(\phi_a(q) \cdot b) \cdots (\phi_{a^k}(q) \cdot b)$. Since $\mu(\text{int}(\Delta S)) = 1$, it must be that $p_s \in (0, 1)$ for all $s \in S$. Moreover, the fact that μ is generic implies that $a = \mathbf{1}$. This implies that

$$\mu(dq) = (q \cdot b)\,\phi_a(q)\,\mu(dq) = (q \cdot b)\,\mu(dq)$$

so $b = \mathbf{1}$. Hence, for all $s \in S$,

$$\frac{r_s}{p_s \alpha_s} = \frac{r_s}{p_s} = 1$$

This implies $r = p$ and $\alpha = \mathbf{1}$ as desired.

A.3 Proof of Proposition 3

Let $\succeq$ and ρ be represented by (r, v) and (ν, v) respectively. For each $s \in S$, let $\{\underline{x}_s, \bar{x}_s\} \subset \Delta X$ be such that $\underline{v}_s = v_s(\underline{x}_s) \le v_s(y) \le v_s(\bar{x}_s) = \bar{v}_s$ for all $y \in \Delta X$. Note that $\underline{v}_s < \bar{v}_s$ for all $s \in S$ due to Axiom 1.4. Moreover, by Theorem 1, we can normalize $\underline{v}_s = 0$ for all $s \in S$ without loss of generality. We will prove that the following are equivalent:

(1) There exists a calibrating act h^*

(2) Any diminishing h_i converges to a non-degenerate h^*

(3) $(\succeq, \rho)$ has a Bayesian representation

We will show that (2) implies (1) implies (3) implies (2). First suppose (2) is true and let h_i be an diminishing sequence. Hence, h_i converges to some non-degenerate h^* and $h^* \in H^* = \lim_i H_i$. For each $h_i \in H$, define $w_i \in \mathbb{R}^S_+$ such that

$$w_i := \int_{\Delta S} \frac{q}{q \cdot (u \circ h_i)} \nu(dq)$$

so $B_{\rho,h_i} = \{f \in H \mid w_i \cdot (v \circ f) \ge 1\}$. Let $\hat{w}_i$ be the projection of w_i on $v(H_0)$. We show that $\hat{w}_i \to 0$. Suppose otherwise and $\hat{w}_i \to \hat{w}^* \neq 0$. Hence, by continuity, we can find some h_j such that $w_j \cdot (u \circ h^*) < w_j \cdot (u \circ h_j)$ implying $h^* \notin H^*$ a contradiction. Hence, $\hat{w}^* = 0$ so $w^* \cdot (u \circ f) = w^* \cdot (u \circ g)$ for all $f \sim g$. This implies h^* is a calibrating act proving (1).

Now, suppose (1) is true. Hence, we can find some $h^* \in H$ such that $r \cdot (v \circ f) \ge r \cdot (v \circ g)$ iff

$$\int_{\Delta S} \frac{q \cdot (v \circ f)}{q \cdot (v \circ h^*)} \nu(dq) \ge \int_{\Delta S} \frac{q \cdot (v \circ g)}{q \cdot (v \circ h^*)} \nu(dq)$$

If we let $f \sim bh^* + (1-b)\underline{h}$ where $\underline{h}$ is the worst act, then we have

$$\int_{\Delta S} \frac{q \cdot (v \circ f)}{q \cdot (v \circ h^*)} \nu(dq) = b = \frac{r \cdot (v \circ f)}{r \cdot (r \circ h^*)}$$

Now, let $\alpha_s := v(h^*_s)$ for all $s \in S$ and define $\phi : \Delta S \to \Delta S$ such that

$$\phi(q)(s) := \frac{q(s)\alpha_s}{q \cdot \alpha}$$

Let $p := \phi(r)$, $\mu := \nu \circ \phi^{-1}$ and $u_s = \frac{v_s}{\alpha_s}$. Note that $q \cdot (v \circ f) \geq q \cdot (v \circ g)$ iff

$$\sum_s q_s v_s(f_s) \geq \sum_s q_s v_s(g_s)$$

$$\sum_s q_s \alpha_s u_s(f_s) \geq \sum_s q_s u_s(g_s)$$

$$\phi(q) \cdot (u \circ f) \geq \phi(q) \cdot (u \circ g)$$

Hence, ρ is also represented by (μ, u) and $\succeq$ by (p, u). Now,

$$\frac{r \cdot (v \circ f)}{r \cdot (r \circ h)} = \frac{\sum_s r_s \alpha_s \frac{v_s(f_s)}{\alpha_s}}{\sum_s r_s \alpha_s \frac{v_s(h_s^*)}{\alpha_s}} = \phi(r) \cdot (u \circ f) = p \cdot (u \circ f)$$

and

$$\int_{\Delta S} \frac{q \cdot (v \circ f)}{q \cdot (v \circ h^*)} \nu(dq) = \int_{\Delta S} \frac{\sum_s q_s \alpha_s \frac{v_s(f_s)}{\alpha_s}}{\sum_s q_s \alpha_s \frac{v_s(h_s^*)}{\alpha_s}} \nu(dq) = \int_{\Delta S} q \cdot (u \circ f) \mu(dq)$$

so

$$\int_{\Delta S} q \cdot (u \circ f) \mu(dq) = p \cdot (u \circ f)$$

Hence, $p = \int_{\Delta S} q \ \mu(dq)$ so (3) is true.

Finally, we show that (3) implies (2). Suppose $(\succeq, \rho)$ is represented by (μ, u). Normalize $\underline{u}_s = 0$ for all $s \in S$. Choose any $h_0 \in H$ and let $\lambda := p \cdot (u \circ h_0)$. Let $h^* \in H$ be such that $u_s(h_s^*) = \lambda \in (0, 1)$ for all $s \in S$. Hence, $h_0 \sim h^*$. Now, note that for each $h_i \sim h_0$, $h_i \sim h^*$ and by Jensen's,

$$\int_{\Delta S} \frac{q \cdot (u \circ h^*)}{q \cdot (u \circ h_i)} \mu(dq) = \int_{\Delta S} \frac{\lambda}{q \cdot (u \circ h_i)} \mu(dq) > \frac{\lambda}{\int_{\Delta S} q \cdot (u \circ h_i) \mu(dq)} = \frac{p \cdot (u \circ h^*)}{p \cdot (u \circ h_i)} = 1$$

Hence, $h^* \succ_{\rho, h_i} h_i$ for all h_i. This implies that if h_i is an diminishing sequence, then $h^* \in H^*$ proving (2).