Penn Institute for Economic Research
Department of Economics
University of Pennsylvania
3718 Locust Walk
Philadelphia, PA 19104-6297
pier@econ.upenn.edu
http://economics.sas.upenn.edu/pier

PIER Working Paper 11-042

"Models of Subjective Learning"

by

David Dillenberger and Philipp Sadowski

http://ssrn.com/abstract=1971259

Models of Subjective Learning*

David Dillenberger[†] Philipp Sadowski[‡]

December 2011

Abstract

We study a decision maker who faces a dynamic decision problem in which the process of information arrival is subjective. By studying preferences over menus of acts, we derive a sequence of utility representations that captures the decision maker's uncertainty about the beliefs he will hold when choosing from a menu. In the most general model of second-order beliefs, we characterize a notion of "more preference for flexibility" via a subjective analogue of Blackwell's (1951, 1953) comparisons of experiments. We proceed to analyze a model in which signals are subsets of the state space. The corresponding representation enables us to compare the behavior of two decision makers who expect to learn differently, even if they do not agree on their prior beliefs. The class of information systems that can support such a representation generalizes the notion of modeling information as a partition of the state space. We apply the model to study a decision maker who anticipates subjective uncertainty to be resolved gradually over time. We derive a representation that uniquely identifies both the filtration, which is the timing of information arrival with the sequence of partitions it induces, and the decision maker's prior beliefs.

Key words: Resolution of uncertainty, second-order beliefs, preference for flexibility, valuing binary bets more, generalized partition.

1. Introduction

The study of dynamic models of decision making under uncertainty when a flow of information on future risks is expected over time is central in all fields of economics. For example, investors decide when to invest and how much to invest based on what they expect to learn about the distribution of future cash flows. The concepts of value of information and value of flexibility (option value) quantify the positive effects of relying on more precise information structures.[1]

*First version September 2011. We thank David Ahn, Brendan Daley, Haluk Ergin, Itzhak Gilboa, Faruk Gul, Peter Landry, Wolfgang Pesendorfer, and Roee Teper for useful advice.

[†]Department of Economics, University of Pennsylvania. E-mail: ddill@sas.upenn.edu

[‡]Department of Economics, Duke University. E-mail: p.sadowski@duke.edu

[1]For a comprehensive survey of the theoretical literature, see Gollier (2001, chapters 24 and 25).

A standard dynamic decision problem has three components: the first is a set of states of the world that capture all relevant aspects of the decision environment. The second component is a set of feasible intermediate actions, each of which determines the payoff for any realized state. The third component is a description of the uncertainty that the decision maker faces; it consists of an information structure, which is the set of possible signals about the states that are expected to arrive over time, and the joint distribution of signals and states.

In many situations, the analyst may be confident in his understanding of the relevant state space and the relevant set of actions. He may, however, find it too restrictive to assume that he is aware of all the relevant uncertainties that people face. People may have access to private data which is unforeseen by others; they may interpret data in an idiosyncratic way; or they may be selective in the data they observe, for example by focusing their attention on specific signals. We collectively refer to those situations as "subjective learning". A natural question is whether we can rely on only the first two components above, and infer an individual's subjective process of learning solely from his observed choice behavior. If the answer is in the affirmative, we ask whether we can compare the behavior of individuals who expect to learn differently, and how such comparisons relate to the comparative statics for incremental increases in informativeness when learning is objective. These questions will be the subject of our analysis.

We consider an objective state space. Actions correspond to acts, that is, functions from states to outcomes, and preferences are defined over sets (or menus) of acts. The interpretation is that the decision maker (henceforth DM) initially chooses among menus and subsequently chooses an alternative from the menu. If the ultimate choice of an act takes place in the future, then the DM may expect information to arrive prior to this choice. Analyzing today's preferences over future choice situations (menus of acts rather than the acts themselves) allows us to capture the effect of the information the DM expects to learn via his value for flexibility. The preference relation over menus of acts is thus the only primitive of the model, leaving the uncertainty that the DM faces, as well as his ultimate choice of an act, unmodeled.

As a concrete example, consider a DM who is currently renting an apartment on a month-to-month lease and deliberates about buying a condominium at a nonnegotiable price. While the physical properties of the condominium are easily assessed, its value also depends on circumstances which are not characteristics of the condominium itself, about which the DM is uncertain, and which we refer to as states of the world. These can be, for example, the location of public schools or the demographic distribution of people across different neighborhoods. Availability of the condominium can be guaranteed for thirty days. Buying the condominium today saves the DM one month's rent. Delaying the purchase decision by one month allows him to conduct market research – gathering and interpreting formal and informal information about the state of the world – which enables him to make a more informed decision. The choice between buying today and delaying the purchase decision can be thought as a choice between a degenerate menu, where the DM purchases the condominium and saves the monthly rent, and the menu that contains the options to buy or not.

Section 2 outlines the most general model that captures subjective learning: the DM acts as if he has beliefs over the possible posterior distributions over the state space that he might face at the time of choosing from the menu. The model is parameterized by a probability measure on the collection of all possible posterior distributions. This probability measure, which we refer to as a second-order belief, is uniquely identified from choice behavior. We use this representation (first derived in Takeoka (2004)) to compare preference for flexibility among decision makers. We say that DM1 has more preference for flexibility than DM2 if whenever DM1 prefers to commit to a particular action rather than to maintain multiple options, so does DM2. We show that DM1 has more preference for flexibility than DM2 if and only if DM2's distribution of first-order beliefs is a mean-preserving spread of DM1's. This result is analogous to Blackwell's (1951, 1953) comparisons of experiments (in terms of their information content) in a domain where probabilities are objective and comparisons are made with respect to the accuracy of information systems. To rephrase our results in the language of Blackwell, DM1 has more preference for flexibility than DM2 if and only if DM2 would be weakly better off if he could rely on the information system induced by the subjective beliefs of DM1. In the condominium example above, we can consider two individuals who agree on their current evaluation of the condominium. Then one DM is willing to pay a larger fee (for example, a higher additional monthly rent) than the other DM to delay the decision whether or not to purchase the condominium if and only if he expects to be better informed by the end of the month.

The most general model does not allow the comparison of two individuals in terms of the information they expect to learn, unless they agree on their prior beliefs, because information may be tacit, that is, it cannot be described in terms of the objective state space. A describable signal is an element of the power set of the objective state space. The model outlined in Section 3 considers learning from describable signals. The DM has beliefs about which information set he might be in at the time he chooses from the menu. For any information set, he calculates his posterior beliefs by excluding all states that are not in that set and applying Bayes' law with respect to the remaining states. We characterize the class of information systems that admit such a representation as a natural generalization of a set partition. The requirement on information systems turns out to be closely related to the notion of a balanced collection of weights, as introduced by Shapley (1967) in the context of cooperative games. This representation allows us to compare the behavior of two individuals who expect to learn different amounts of information, without requiring that they share the same initial beliefs. Their behavior differs in the value they derive from the availability of binary bets as intermediate actions; roughly speaking, DM1 "values binary bets more" than DM2 if for any two states, he is willing to pay more in order to have the option to bet on one state versus the other. In this case, DM1 expects to receive more information than DM2, in the sense that given the true state of the world, he is more likely to be able to rule out any other state (i.e. to be in an information set that contains the true state but not the other state.)

Lastly, reconsider the condominium example, and assume that the availability of the condominium is not guaranteed, but rather the agent is given the right of first refusal in case another offer arrives within the next thirty days. In this situation, DM's information set at any point in this time interval may become the relevant one for his purchase decision. In Section 4 we provide a representation, which suggests that the DM behaves as if he has in mind a filtration, indexed by continuous time. Both the filtration, which is the timing of information arrival with the sequence of partitions it induces, and the DM's prior beliefs are uniquely determined from choice behavior. In this context, DM1 values binary bets more than DM2 if and only if he expects to learn earlier in the sense that his filtration is finer at any given point in time. DM1 has more preference for flexibility than DM2 if and only if they also share the same prior beliefs.

Several papers have explored the idea of subjective learning. As mentioned earlier, Takeoka (2004) derives the most general model of second-order beliefs. We show that even this general setting allows very intuitive comparative statics. Hyogo (2007) derives a representation that features second-order beliefs on a richer domain, where the DM simultaneously chooses a menu of acts and takes an action that might influence the (subjective) process of information arrival. Dillenberger, Lleras, and Sadowski (2011) study a model in which the information system is partitional, that is, signals correspond to information sets that do not intersect. Since a partition is a special case of a generalized partition, the model is a special case of the one outlined in Section 3. Learning by partition can also be viewed as a special case of the model in Section 4, where the DM does not expect to learn gradually over time, that is, he forms his final beliefs at time zero, right after he chose a menu. Takeoka (2007) uses a different approach to study subjective temporal resolution of uncertainty. He analyzes a different domain, where the DM chooses between what one might term "compound menus" (menus over menus etc.). We compare the two different approaches in Section 5.2, while reevaluating our domain in light of the results from Section 4.

More generally, our work is part of the preferences over menus literature initiated by Kreps (1979). Most papers in this literature study uncertainty over future tastes (and not over beliefs) without assuming an objective state space. Kreps (1979) studies preferences over menus of deterministic alternatives. Dekel, Lipman, and Rustichini (2001) extend Kreps' domain of choice to menus of lotteries. Some of the axioms that lead to the most general representation of second-order beliefs are adapted from Dekel et al.'s paper. Our proof of that theorem relies on a sequence of geometric arguments that establish the close connection between our domain and theirs. In the setting of preferences over menus of lotteries, Ergin and Sarver (2010) provide an alternative to Hyogo's (2007) approach of modeling costly information acquisition.

1.1. A formal preview of the representation results

Let S be a finite state space. An act is a mapping $f : S \rightarrow [0,1]$, where $[0,1]$ is interpreted as a utility space.[2] Let $\mathcal{F}$ be the set of all acts. Let $\mathcal{K}(\mathcal{F})$ be the set of all non-empty compact subsets of $\mathcal{F}$. Preferences are defined over $\mathcal{K}(\mathcal{F})$. Theorem 1 derives a (second-order beliefs) representation, in which the value of a set F is given by

$$V(F) = \int_{\Delta(S)} \max_{f \in F} \left(\textstyle\sum_{s \in S} f(s)\, \pi(s) \right) dp(\pi),$$

where $p(\cdot)$ is a unique probability measure on $\Delta(S)$, the space of all probability measures on S. The axioms that are equivalent to the existence of such representation are familiar from the literature on preferences over menus of lotteries– *Ranking, vNM Continuity, Nontriviality, and Independence*– adapted to our domain, in addition to *Dominance*, which implies monotonicity in payoffs, and *Set Monotonicity*, which captures preferences for flexibility.

We then study a specialized model in which signals are subsets of the state space, that is, elements of 2^S. We impose two additional axioms, Finiteness and Context Independence. *Finiteness* implies that the probability measure p in Theorem 1 has finite support. (Finiteness is obviously necessary since 2^S is finite.) To formulate *Context Independence*, we first identify through preferences a special subset of menus that we term *saturated* (Definition 6). We show that if F is saturated then $f \in F$ implies that there exists a posterior π_f in the support of p in Theorem 1, with the following properties: (i) f is optimal (maximizes the expected utility among all the acts in F) given π_f and only given π_f; and (ii) f yields positive payoffs only on those states that gets positive probability under π_f. Given these properties, Context Independence has the flavor of Savage's sure-thing principle. Suppose

[2] This allows us to abstract from deriving the DM's utility function over monetary prizes, which is a standard exercise.

that given his prior beliefs, the DM prefers committing to the act g to committing to the act f, where both g and f yield positive payoffs only on a subset $E \subset S$. The axiom then requires that the DM would prefer to replace f with g in any saturated menu that contains f. With these additional axioms, Theorem 3 derives an (information set) representation in which the value of a set F is given by

$$V(F) = \sum_{I \in 2^S} \max_{f \in F} \left[\sum_{s \in I} f(s) \frac{\mu(s)}{\mu(I)} \right] \rho(I),$$

where μ is a probability measure on S and ρ is a probability measure on 2^S, such that $\sum_{I \in 2^S | s \in I} \frac{\rho(I)}{\mu(I)} = 1$ for all $s \in S$. The pair (μ, ρ) is unique. The condition that $\sum_{I \in 2^S | s \in I} \frac{\rho(I)}{\mu(I)} = 1$ for all $s \in S$ implies that the probability of being in information set I when the state of the world is s is the same for all states $s \in I$. To say this differently, the DM behaves as if he can infer no information about relative probabilities from the information set.

A natural question is which information structures $\Psi \subseteq 2^S$ are admissible, in the sense that there exists an information set representation in which Ψ is the support of ρ. Theorem 4 shows that Ψ is admissible if and only if it is a *generalized partition*; $\Psi \subseteq 2^S$ is a generalized partition of a set $S' \subseteq S$ if there exists $k \geq 1$ and a function $\beta : \Psi \to \mathbb{N}_+$ such that for all $s \in S'$, $\sum_{I \in \Psi | s \in I} \beta(I) = k$. In this case we say that S' is covered k times by Ψ. Note that the usual notion of a set partition is implied if $k = 1$.

Lastly, we show that the same domain can capture the effect of subjective gradual resolution of uncertainty. To this end, we reinterpret menus as choice situations in which the opportunity to choose form the menu arrives randomly. We use the notion of saturated menus to impose an additional axiom, *Hierarchy*, which implies that the support of ρ in Theorem 3 has a hierarchical structure. This allows us to interpret information as becoming more precise over time: Theorem 7 provides an (exclusive tree) representation in which the value of a set F is given by

$$V(F) = \int_{[0,1]} \left\{ \sum_{P \in \mathcal{P}_t} \max_{f \in F} \left[\sum_{s \in S} f(s) \mu(s | P) \right] \mu(P) \right\} dt,$$

where μ is a probability measure on S and $\{\mathcal{P}_t\}$ is a filtration indexed by $t \in [0,1]$. The pair $(\mu, \{\mathcal{P}_t\})$ is unique. In this context, DM1 values binary bets more than DM2 if and only if $\{\mathcal{P}_t^1\}$ is finer than $\{\mathcal{P}_t^2\}$ (i.e., for any t, all events in $\{\mathcal{P}_t^2\}$ are measurable in $\{\mathcal{P}_t^1\}$). DM1 has more preference for flexibility than DM2 if and only if both also share the same prior beliefs (i.e., $\mu^1 = \mu^2$).

The remainder of the paper is organized as follows: Section 2 studies the most general model of uncertainty about future beliefs. Section 3 studies a special case in which signals correspond to information sets. Section 4 further specializes the model to situations in which uncertainty is expected to be resolved gradually over time, and the pattern of its resolution matters. Section 5 suggests a reinterpretation and an application of the model outlined in Section 4 to cases in which at any point in time the DM chooses an act from the menu and derives a utility flow from it. The section concludes by comparing our methodology to other approaches to the study of subjective temporal resolution of uncertainty. Most proofs are relegated to the appendix.

2. A general model of subjective learning

Let $S = \{s_1, ..., s_k\}$ be a finite state space. An act is a mapping $f : S \to [0, 1]$. Let $\mathcal{F}$ be the set of all acts. Let $\mathcal{K}(\mathcal{F})$ be the set of all non-empty compact subsets of $\mathcal{F}$. Capital letters denote sets, or menus, and small letters denote acts. For example, a typical menu is $F = \{f, g, h, ...\} \in \mathcal{K}(\mathcal{F})$. We interpret payoffs in $[0, 1]$ to be in "utils"; that is, we assume that the utility function over outcomes is known and payoffs are stated in its units. An alternative interpretation is that there are two monetary prizes $x > y$, and $f(s) = p_s(x) \in [0, 1]$ is the probability of getting the greater prize in state s.

Let $\succeq$ be a preference relation over $\mathcal{K}(\mathcal{F})$. The symmetric and asymmetric components of $\succeq$ are denoted by $\sim$ and $\succ$, respectively. We impose the following axioms on $\succeq$:

Axiom 1 (Ranking). *The relation $\succeq$ is a weak order.*

Definition 1. *Let $\alpha F + (1-\alpha) G := \{\alpha f + (1-\alpha) g : f \in F,\ g \in G\}$, where $\alpha f + (1-\alpha) g$ is the act that yields $\alpha f(s) + (1-\alpha) g(s)$ in state s.*

Axiom 2 (vNM Continuity). *If $F \succ G \succ H$ then there are $\alpha, \beta \in (0, 1)$, such that $\alpha F + (1-\alpha) H \succ G \succ \beta F + (1-\beta) H$.*

Axiom 3 (Nontriviality). *There are F and G such that $F \succ G$.*

Axiom 4 (Independence). *For all F, G, H, and $\alpha \in [0, 1]$,*

$$F \succeq G \Leftrightarrow \alpha F + (1-\alpha) H \succeq \alpha G + (1-\alpha) H.$$

In the domain of menus of acts, Axiom 4 implies that the DM's preferences must be linear in payoffs. This is plausible since we interpret payoffs in $[0, 1]$ directly as "utils", as discussed above.

Axiom 5 (Set monotonicity). *If $F \subset G$ then $G \succeq F$.*

Axiom 5 was first proposed in Kreps (1979). It captures preference for flexibility, that is, bigger sets are weakly preferred. The interpretation of $f(\cdot)$ as a vector of utils requires the following payoff-monotonicity axiom.

Axiom 6 (Domination). *If $f \geq g$ and $f \in F$ then $F \sim F \cup \{g\}$.*

Axioms 1-6 are necessary and sufficient for the most general representation of subjective learning, which is derived in Takeoka (2004).

Theorem 1 (Takeoka (2004)). *The relation $\succeq$ satisfies Axioms 1–6 if and only if it can be represented by:*

$$V(F) = \int_{\Delta(S)} \max_{f \in F} \left(\textstyle\sum_{s \in S} f(s) \pi(s) \right) dp(\pi),$$

where $p(\cdot)$ is a unique probability measure on $\Delta(S)$, the space of all probability measures on S.

Proof. See Appendix 6.1 ■

The representation in Theorem 1 suggests that the DM is uncertain about which first-order beliefs π he will have at the time he makes a choice from the menu.

Since our axioms are slightly different from Takeoka's and since his working paper is unpublished, for readers' convenience we present our proof in the appendix. Another related work, Dekel, Lipman, and Rustichini (2001), analyzes choice over menus of lotteries and finds a representation that suggests uncertainty about the DM's tastes (a relevant corrigendum is Dekel, Lipman, Rustichini, and Sarver (2007)). Our proof relies on a sequence of geometric arguments that establish the close connection between our domain and theirs. The parameter p is uniquely identified in the representation above, because p and π are required to be probability measures. Such natural normalization does not exist in Dekel et al. (2001, 2007), and therefore they can only jointly identify the parameters in their representation.

2.1. More preference for flexibility and the theorem of Blackwell

We now connect a notion of preference for flexibility with the DM's subjective learning.

Definition 2. *DM1 has more preference for flexibility than DM2 if for all $f \in \mathcal{F}$ and for all $G \in \mathcal{K}(\mathcal{F})$,*

$$\{f\} \succeq_1 G \text{ implies } \{f\} \succeq_2 G$$

Expressed in words, DM1 has more preference for flexibility than DM2 if whenever DM1 prefers to commit to a particular action rather than to retain an option to choose, so does DM2.[3] In what follows, when we discuss a particular individual i, we denote by V^i the representation of his preferences and by $\sigma(p^i)$ the corresponding support of his second-order beliefs.

Remark 1. *Definition 2 is equivalent to the notion that if DM1 and DM2 are endowed with the same act, then DM1 has a greater willingness to pay to acquire additional options. That is, for all $f, h \in \mathcal{F}$ with $f \geq h$ and for all $G \in \mathcal{K}(\mathcal{F})$,*

$$\{f\} \succeq_1 \{f - h\} \cup G \text{ implies } \{f\} \succeq_2 \{f - h\} \cup G,$$

where $(f - h)(s) = f(s) - h(s)$. The act h is interpreted as the state-contingent cost of acquiring the options in G. Definition 2 clearly implies this condition. The converse follows from taking $h = f$.

Definition 2, however, does not imply a greater willingness to pay to add options to any *given menu. In particular, even if DM1 has more preferences for flexibility than DM2, it may be possible to find $G \subset F \in \mathcal{K}(\mathcal{F})$ and an act h_c with $h_c(s) = c$ for all s and $f \geq h_c$, such that both $V^2(F - c) > V^2(G)$ and $V^1(F - c) \leq V^1(G)$ hold, where*

$$F - c := \{f - h_c \,|\, f \in F\}.$$

For example, suppose $S = \{s_1, s_2\}$ and that both DM1 and DM2 think the two states are equally likely. DM1 expects to learn the true state for sure, that is, $\sigma^1(p) = \{(1, 0), (0, 1)\}$, whereas DM2 expects to learn nothing, that is, $\sigma^2(p) = \{(0.5, 0.5)\}$. Let $G = \{(k, 0), (0, k)\}$ and let $F = \{(k, 0), (0, k), (k, k)\}$. Then $V^1(F - c) < V^1(G)$ for all $c > 0$ (since DM1 can guarantee himself a payoff of k from menu G), whereas for $c \in \left(0, \frac{k}{2}\right)$, $V^2(F - c) > V^2(G)$.[4]

Claim 1. *Suppose DM1 has more preference for flexibility than DM2. Then*

$$\{f\} \succeq_1 \{g\} \text{ if and only if } \{f\} \succeq_2 \{g\}.$$

[3] Definition 2 is analogous to the notion of "more aversion to commitment" as appears in Higashi, Hyogo, and Takeoka (2009, Definition 4.4, p. 1031) in the context of preferences over menus of lotteries.

[4] In fact, defining more preference for flexibility as a greater willingness to pay to add options to any given menu would result in an empty relation. To see this, suppose that $\succeq_1 \neq \succeq_2$ and, for simplicity, that $\sigma(p^1)$ and $\sigma(p^2)$ are finite. Using Theorem 1, there is a first-order belief π, such that $p^1(\pi) > p^2(\pi)$. It is easy to construct a menu that generates payoff $k - \delta$ under belief π and payoff k under any other belief. DM1 then would be willing to pay more than DM2 to add an act that yields k on π, hence DM2 would not have more preference for flexibility than DM1. But by a symmetric argument, DM1 would also not have more preference for flexibility than DM2.

Proof. Let $G = \{g\}$ for some $g \in \mathcal{F}$. Applying Definition 2 implies that if $\{f\} \sim_1 \{g\}$ then $\{f\} \sim_2 \{g\}$. That is, any indifference set of the restriction of $\succeq_1$ to singletons is a subset of some indifference set of the restriction of $\succeq_2$ to singletons. The linearity (in probabilities) of the restriction of $V^i(\cdot)$ to singletons implies that these indifference sets are planes that separate any n–dimensional unit simplex, for $n \leq (|S| - 1)$. Therefore, the indifference sets of the restriction of $\succeq_1$ and $\succeq_2$ to singletons must coincide. Since the restrictions of $\succeq_1$ and of $\succeq_2$ to singletons share the same indifference sets and since both relations are monotone, they must agree on all upper and lower contour sets. In particular, $\{f\} \succeq_1 \{g\}$ if and only if $\{f\} \succeq_2 \{g\}$. ■

We now compare *subjective* information systems in analogy to the notion of better information proposed by Blackwell (1951, 1953) in the context of objective information. Definition 3 below says that an information system is better than another one if and only if both systems induce the same distribution of prior probabilities, and all posteriors probabilities of the latter are a convex combination of the posterior probabilities of the former.

Definition 3. *DM1 expects to be better informed than DM2 if and only if DM2's distribution of first-order beliefs is a mean-preserving spread of DM1's (in the space of probability distributions), that is*

(i) Mean preserving:

$$\int_{\Delta(S)} \pi(s)\, dp^1(\pi) = \int_{\Delta(S)} \pi(s)\, dp^2(\pi)$$

for all $s \in S$; and

(ii) Spread (garbling): there exists a nonnegative function $k : \sigma(p^1) \times \sigma(p^2) \to \mathbb{R}_+$, such that

$$\int_{\sigma(p^2)} k(\pi, \pi')\, d\pi' = 1$$

for all $\pi \in \sigma(p^1)$, and

$$\pi'(s) = \int_{\sigma(p^1)} \pi(s)\, k(\pi, \pi')\, d\pi$$

for all $\pi' \in \sigma(p^2)$ and $s \in S$.

Theorem 2. *If DM1 and DM2 have preferences that can be represented as in Theorem 1, then DM1 has more preference for flexibility than DM2 if and only if DM1 expects to be better informed than DM2.*[5]

[5] The characterization of preference for flexibility via Blackwell's comparison of information systems is specific to our context, where this preference arises due to uncertainty about learning. Krishna and Sadowski (2011) provide an analogous result in a context where preference for flexibility arises due to uncertain tastes.

Proof. Blackwell (1953, Theorem 8) establishes that DM2's distribution of first-order beliefs is a mean-preserving spread of DM1's if and only if $V^1(G) \geq V^2(G)$ for any $G \in \mathcal{K}(\mathcal{F})$. At the same time, $V^1(\{f\}) = V^2(\{f\})$ for any $f \in \mathcal{F}$. Hence, $V^1(\{f\}) \geq V^1(G)$ implies $V^2(\{f\}) \geq V^2(G)$. Conversely, suppose $V^2(G) > V^1(G)$ for some $G \in K(\mathcal{F})$. Then continuity implies that there exists $f \in \mathcal{F}$ with $V^2(G) > V^2(\{f\}) = V^1(\{f\}) > V^1(G)$. ■

3. Subjective learning with objectively describable signals

The model in Section 2 is the most general model that captures subjective learning. In Theorem 2 we compare the behavior of two individuals who share the same prior beliefs but expect to learn differently. We would like to be able to perform such a comparison even if the two individuals disagree on their prior beliefs; for example, one individual might consider himself a better experimenter than the other, even though he holds more pessimistic beliefs about the state of the world. Disagreement on prior beliefs may not matter if we try to compare the amount of information the two individuals expect to learn contingent on the true state of the world. Distinct priors, however, generically imply that the contingent priors are also different. To see this, let, for $i = 1, 2$, μ^i be a vector of DMi's prior beliefs and let $a^i(s\,|s')$ be the probability he assigns to state s contingent on the true state being s'.[6] Then $A^i := (a^i(s\,|s'))_{s,s'}$ is a stochastic matrix and Bayes' law implies $\mu^i A^i = \mu^i$, that is, μ^i is the stationary distribution of A. If each entry of A is strictly positive, then A is an indecomposable matrix and the stationary distribution is unique. In that case, different priors, μ^1 and μ^2, must correspond to different stochastic matrices, A^1 and A^2. But since the rows of A^i are the state-contingent priors of DMi, there must be at least one state s, contingent on which a comparison as in Theorem 2 is impossible.

In order to compare the amount of information each DM expects to learn contingent on the state, we need to be able to describe information independently of the induced changes in beliefs. To this end, we now consider a more parsimonious model of learning in which signals correspond to information sets, that is, to learning a subset of the objective state space. The DM's beliefs can then be understood as uncertainty about the information set he will be in at the time of choosing from the menu. We maintain the assumptions of Theorem 1 and develop a language that allows us to formulate, in terms of behavior, the assumption that the DM cannot draw any inferences from learning an information set besides knowing that states outside that set were not realized. To say this differently, we axiomatize the most general representation in which the relative probability of any two states is the same across

[6] The probability individual i assigns to first-order belief $\pi \in \Delta(S)$ contingent on state $s' \in \sigma(\mu^i)$ is $l^i(\pi\,|s') := \frac{\pi(s')p^i(\pi)}{\mu^i(s')}$. Then $a^i(s\,|s') = \sum_{\pi \in \Delta(S)} \pi(s)\, l^i(\pi\,|s')$.

all information sets that contain them.[7] In Section 3.1, we further identify the largest class of (subjective) information systems that can accommodate this type of learning. This class generalizes the notion of modeling information as a partition of the state space. Finally, in Section 3.2, we compare two individuals according to the amount of information each expects to acquire without restricting them to have the same prior beliefs.

Since there are only finitely many distinct subsets of S, the support of the function p, $\sigma(p)$, in Theorem 1 must be finite. This restriction is captured by the following axiom:

Axiom 7 (Finiteness). *For all $F \in \mathcal{K}(\mathcal{F})$, there is a finite set $G \subseteq F$ with $G \sim F$.*[8]

Dekel, Lipman, and Rustichini (2009) show that Axiom 7 indeed implies that $\sigma(p)$ is finite. The intuition is clear: if for any F there is a finite subset G of F that is as good as F itself, then only a finite set of first-order beliefs can be relevant.

Definition 4. *Given $f \in \mathcal{F}$, let f_s^x be the act*

$$f_s^x(s') = \begin{cases} f(s') & \text{if } s' \neq s \\ x & \text{if } s' = s \end{cases}$$

Note that $\sigma(f) := \{s \in S \,|\, f(s) > 0\} = \{s \in S \,|\, f_s^0 \neq f\}$.

Definition 5. *A menu $F \in \mathcal{K}(\mathcal{F})$ is fat free if for all $f \in F$ and for all $s \in \sigma(f)$, $F \succ (F \backslash \{f\}) \cup \{f_s^0\}$.*

If a menu F is fat free, then for any act $f \in F$ and any state $s \in \sigma(f)$, eliminating s from $\sigma(f)$ reduces the value of the menu.[9] In particular, removing an act f from the fat-free menu F must make the menu strictly worse.

Definition 6. *A menu $F \in \mathcal{K}(\mathcal{F})$ is saturated if it is fat free and satisfies*

(i) for all $f \in F$ and $s \notin \sigma(f)$, there exists $\bar{\varepsilon} > 0$ such that $F \sim F \cup f_s^{\varepsilon}$ for all $\varepsilon < \bar{\varepsilon}$; and

(ii) If $G \nsubseteq F$ then $F \cup G \sim (F \cup G) \setminus \{g\}$ for some $g \in F \cup G$.

[7] I.e., S must be large enough to contain all random variables that the DM considers informative. Subjectivity, therefore, does not refer to the information content of a signal, but only to whether or not the DM learns a particular signal.

[8] We impose Axiom 7 mainly for clarity of exposition. Alternatively, it is possible to strengthen Definition 5, Definition 6, and Axiom 8 below to apply to situations where Finiteness may not hold. In that case, Axiom 7 is implied.

[9] Our notion resembles the notion of "fat-free acts" suggested by Lehrer (2008). An act f is fat-free if when an outcome assigned by f to a state is replaced by a worse one, the resulting act is strictly inferior to f. In our setting, a finite fat-free set contains acts, for all of which reducing an outcome in any state in the support results in an inferior set.

Definition 6 says that if F is a saturated menu, then (i) if an act $f \in F$ does not yield any payoff in some state, then the DM's preferences are insensitive to slightly improving f in that state; and, (ii) adding an act to a saturated menu implies that there is at least one act in the new menu that is not valued by the DM. In particular, the extended menu is no longer fat-free.

To better understand the notions of fat-free and saturated menus, consider the following example.

Example 1. *Suppose that there are two states $S = \{s_1, s_2\}$. If the act f yields positive payoffs in both states but only one of them is non-null, then $\{f\}$ is not fat-free. If both states are non-null and f does not yield positive payoffs on one of them, then the set $\{f\}$ is not saturated according to Definition 6 (i). If the two states are non-null and f yields positive payoffs in both, then $\{f\}$ is fat-free, but it is not necessarily saturated. For example, if the DM expects to learn the true state for sure, that is, $\sigma^1(p) = \{(1,0), (0,1)\}$, then for $g = (f(s_1) + \varepsilon, 0)$, both $\{f, g\} \succ \{f\}$ and $\{f, g\} \succ \{g\}$, which means that $\{f\}$ is not saturated according to Definition 6 (ii).*

Claim 2. *A saturated menu F, with $f(s) < 1$ for all $f \in F$ and all $s \in S$, always exists. Furthermore, if F is saturated, then F is finite.*

Proof. See Appendix 6.2 ■

In all that follows, we only consider saturated menus that contain acts f with $f(s) < 1$ for all $s \in S$. For ease of exposition, we refrain from always explicitly stating this assumption.

Claim 3. *If F is saturated, then F is isomorphic to the set of first-order beliefs.*

Proof. See Appendix 6.3 ■

Claim 3 connects the definition of a saturated menu with the idea that the DM might be required to make a decision when his state of knowledge is any one of his first-order beliefs from the representation of Theorem 1. Claim 3 then says that any act in a saturated menu is expected to be chosen under exactly one such belief.

The next claim demonstrates that the support of any act in a saturated menu coincides with that of the belief under which the act is chosen. For any act f in a given saturated menu F, let $\pi_f \in \sigma(p)$ be the belief such that $f = \underset{f' \in F}{\arg\max} \sum_{s \in S} f'(s) \pi_f(s)$. By Claim 3, π_f exists and is unique.

Claim 4. *If F is saturated and $f \in F$ then $\sigma(f) = \sigma(\pi_f)$.*

Proof. Suppose $f(s) > 0$ and $\pi_f(s) = 0$. Then $F \sim (F \backslash \{f\}) \cup \{f_s^0\}$, which is a contradiction to F being fat-free (and, therefore, to F being saturated.) Suppose $f(s) = 0$ and $\pi_f(s) > 0$. Then for any $\varepsilon > 0$, $F \prec F \cup \{f_s^\varepsilon\}$, which is a contradiction to F being saturated. ■

We are now ready to state the central axiom of this section.

Axiom 8 (Context independence). *Suppose F is saturated and $f \in F$. Then for all g with $\sigma(g) = \sigma(f)$,*

$$\{g\} \succeq \{f\} \text{ implies } (F \backslash \{f\}) \cup \{g\} \succeq F$$

Suppose the DM prefers committing to g over committing to f, where both g and f pay strictly positive amounts only on the event $\sigma(f)$. The axiom then requires that the DM would prefer to replace f with g on any saturated menu that contains f. In light of Claim 4, the axiom is similar to Savage's sure-thing principle. To see this, recall that Claim 4 suggests f is chosen from the saturated menu F only in the event $\sigma(f)$. We can rephrase the axiom as follows: whenever two acts, f and g, differ at most on the event $\sigma(f)$, then their unconditional ranking agrees with their ranking conditional on $\sigma(f)$.[10]

Definition 7. *The pair (μ, ρ) is an information set representation if μ is a probability measure on S and ρ is a probability measure on 2^S, such that $\sum_{I \in 2^S | s \in I} \frac{\rho(I)}{\mu(I)} = 1$ for all $s \in \sigma(\mu)$, and*

$$V(F) = \sum_{I \in 2^{\sigma(\mu)}} \max_{f \in F} \left[\sum_{s \in I} f(s) \frac{\mu(s)}{\mu(I)} \right] \rho(I)$$

represents $\succeq$.

Consider the probability of learning the information set I given that the true state s is contained in I, $\Pr(I | s \in I)$. Define $p(I) := \frac{\rho(I)}{\mu(I)}$. Observe that for any $s \in I$,

$$\Pr(I | s \in I) = \frac{\Pr(s | I) \rho(I)}{\mu(s)} = \frac{\mu(s)}{\mu(I)} \frac{\rho(I)}{\mu(s)} = \frac{\rho(I)}{\mu(I)} = p(I)$$

[10] The requirement that F is saturated guarantees that if F includes an act h that dominates g but not f, then preferences for which the unconditional ranking of f and g does not agree with their conditional ranking on $\sigma(f)$ are not precluded. For example, let $S = \{s_1, s_2\}$ and assume that the DM ex-ante thinks the two states are equally likely. Suppose that the DM expects to learn the true state for sure, that is $\sigma(p) = \{(1,0), (0,1)\}$. For $\varepsilon > 0$ small enough, consider the following three acts: $g = (1, \varepsilon)$, $h = (1, 2\varepsilon)$, and $f = \left(\frac{1}{3}, \frac{1}{3}\right)$. Let $F = \{h, f\}$. Then $\sigma(g) = \sigma(f)$ and $\{g\} \succ \{f\}$, but

$$(F \backslash \{f\}) \cup \{g\} = \{h, g\} \sim \{h\} \prec \{h, f\} = F.$$

The menu F, however, is not saturated (and is not even fat-free) since, for example, $F \sim (F \backslash \{f\}) \cup \{f_{s_1}^0\}$.

is independent of s. Since p is a probability measure on 2^S, consistency requires that

$$\sum\nolimits_{I\in 2^{\sigma(\mu)}|s\in I} \frac{\rho(I)}{\mu(I)} = \sum\nolimits_{I\in 2^{\sigma(\mu)}|s\in I} p(I) = 1,$$

as in Definition 7.

The fact that $\Pr(I\,|s\in I)$ is independent of s (conditional on $s\in I$) reflects the idea that the DM cannot draw any inferences from learning an information set other than that states outside that information set were not realized. Indeed, for any $s, s'\in I$,

$$\frac{\Pr(s\,|I)}{\Pr(s'\,|I)} = \frac{\mu(s)}{\mu(s')}$$

independent of I.

Theorem 3. *The relation $\succeq$ satisfies Axioms 1–8 if and only if it has an information set representation, (μ,ρ). Furthermore, the pair (μ,ρ) is unique.*

Proof. See Appendix 6.4 ■

In contrast to the representation in Theorem 1, the representation in Theorem 3 suggests that S is large enough to capture the subjective resolution of uncertainty. To say this differently, consider a subjective state space that includes all (possibly only privately observable) random variables the DM might consider informative about the objective state $s\in S$. This subjective state space might be larger than S. The representation suggests that any event in the larger subjective state space that the DM considers informative is measurable in S.

3.1. Admissible information structures

In Theorem 3, signals are identified with information sets and the relative probability of any two states is the same across all information sets that contain them. We now identify the class of information systems, Ψ, such that there is an information set representation (μ,ρ) with $\sigma(\rho)=\Psi$.

Definition 8. *A set $S'\subseteq S$ is covered k times by a collection of sets $\Psi\subseteq 2^S$ if there is a function $\beta:\Psi\to\mathbb{N}_+$, such that for all $s\in S'$, $\sum_{I\in\Psi|s\in I}\beta(I)=k$*

Definition 9. *A collection of sets $\Psi\subseteq 2^S$ is a generalized partition of a set $S'\subseteq S$, if there exists $k\geq 1$, such that S' is covered k times by Ψ.*

In the context of cooperative games, Shapley (1967) introduces the notion of a balanced collection of weights. Denote by $\mathcal{C}$ the set of all coalitions (subsets of the set N of players).

The collection $(\gamma_L)_{L\in\mathcal{C}}$ of numbers in $[0,1]$ is a balanced collection of weights if for every player $i\in N$, the sum of γ_L over all the coalitions that contain i is 1. Suppose $\Psi\subseteq 2^S$ is a generalized partition of a set $S'\subseteq S$. Then there exists $k\geq 1$ such that for all $s\in S'$, $\sum_{I\in\Psi|s\in I}\frac{\beta(I)}{k}=1$. In the terminology of Shapley, the collection $\left(\frac{\beta(I)}{k}\right)_{I\in\Psi}$ of numbers in $[0,1]$ is, thus, a balanced collection of weights.

To better understand the notion of generalized partition, consider the following example.

Example 2. *Suppose $S=\{s_1,s_2,s_3\}$. Any partition of S, for example $\{\{s_1\},\{s_2,s_3\}\}$, is a generalized partition of S (with $k=1$). A set that consists of multiple partitions, for example $\{\{s_1\},\{s_2,s_3\},\{s_1,s_2,s_3\}\}$, is a generalized partition of S (in this example with $k=2$). The set $\{\{s_2,s_3\},\{s_1,s_2,s_3\}\}$ is not a generalized partition of S, because $\sum_{I|s_1\in I}\beta(I)<\sum_{I|s_2\in I}\beta(I)$ for any $\beta:\{\{s_2,s_3\},\{s_1,s_2,s_3\}\}\to\mathbb{N}_+$. The set $\{\{s_2,s_3\},\{s_1\},\{s_2\},\{s_3\}\}$, however, is a generalized partition of S with*

$$\beta(I)=\begin{cases}2 & \text{if } I=\{s_1\}\\ 1 & \text{otherwise}\end{cases}$$

Lastly, the set $\{\{s_1,s_2\},\{s_2,s_3\},\{s_1,s_3\}\}$ is a generalized partition of S (with $k=2$), even though it does not contain a partition.

An empirical situation that gives rise to a generalized partition consisting of two partitions is an experiment that reveals the state of the world if it succeeds, and is completely uninformative otherwise. For a concrete example that gives rise to a generalized partition that does not contain a partition, consider the sequential elimination of n candidates, say during a recruiting process. If k candidates are to be eliminated in the first stage, then the resulting generalized partition is the set of all $(n-k)$-tuples.

Definition 10. *Given $\Psi\subseteq 2^S$, let $S_\Psi:=\left\{s\in S\,\middle|\,s\in\bigcup_{I\in\Psi}I\right\}$.*

Definition 11. *A collection $\Psi\subseteq 2^S$ is admissible if there exists an information set representation (μ,ρ) with $\sigma(\rho)=\Psi$.*

Theorem 4. *A collection Ψ is admissible if and only if Ψ is a generalized partition of S_Ψ.*

Proof. See Appendix 6.5 ■

To illustrate Theorem 4, let us consider a specific example. An oil company is trying to learn whether there is oil in a particular location. Suppose the company can drill a hole to determine accurately whether there is oil, $s=1$, or not, $s=0$. In that case, the company learns the partition $\{\{0\},\{1\}\}$ and $\rho(I)=\mu(I)$ provides an information set representation given the firm's prior beliefs μ on $S=\{0,1\}$.

Now suppose that with some positive probability the test may not be completed (for some exogenous reason, which is not indicative of whether there is oil or not). The company will either face the trivial partition $\{\{0,1\}\}$, or the partition $\{\{0\},\{1\}\}$, and hence $\Psi=\{\{0,1\},\{0\},\{1\}\}$. Suppose the company believes that the experiment will succeed with probability q. Then $\rho(\{0,1\})=1-q$, $\rho(\{0\})=q\mu(0)$ and $\rho(\{1\})=q\mu(1)$ provides an information set representation given the company's prior beliefs μ on $S=\{0,1\}$.

Finally, suppose the company is trying to assess the size of an oil field by drilling in l proximate locations and hence the state space is $\{0,1\}^l$. As before, any test may not be completed, independently of the other tests. This is an example of a situation where the state consists of l different attributes (i.e., the state space is a product space), and the DM may learn independently about any of them. Such learning about attributes also gives rise to a generalized partition that consists of multiple partitions and can be accommodated. To find an information set representation based on any generalized partition, Ψ, for which there is a collection Π of partitions whose union is Ψ, based on any probability distribution q on Π, and based on any measure μ on $S=\{0,1\}$, one can set $\rho(I)=\sum_{\mathcal{P}\in\Pi|I\in\mathcal{P}} q(\mathcal{P})\,\mu(I)$. We refer to the pair (q,Π) as a *random partition*.

Remark 2. *If the state space is defined via the value of all random variables the DM observes, then Bayesian learning gives rise to an information system that is a partition. Conversely, learning can always be described via a partition, if the state space is made sufficiently large. To attain a state space that is surely large enough, one could follow Savage and postulate the existence of a grand state space that describes all conceivable sources of uncertainty. However, a larger state space requires a much larger collection of acts, which poses a serious conceptual problem, as in many applications the domain of choice (the available acts) is given. In that sense, acts should be part of the primitives of the model.*[11] *Our approach instead identifies a behavioral criterion for checking whether a given state space (e.g. the one acts are naturally defined on in a particular application) is large enough: behavior satisfies Axiom 8 if and only if the resolution of any subjective uncertainty corresponds to an event (an information set) in the state space. Theorem 4 demonstrates*

[11] Gilboa, Postlewaite, and Schmeidler (2009a, 2009b) point out the problems involved in using an analytical construction, according to which states are defined as functions from acts to outcomes, to generate a state space that captures all conceivable sources of uncertainty. First, since all possible acts on this new state space should be considered, the new state space must be extended yet again, and this iterative procedure does not converge. Second, the constructed state space may include events that are never revealed to the DM, and hence some of the comparisons between acts may not even be potentially observable. (A related discussion appears in Gilboa (2009, Section 11.1.)

that this does not require a state space in which learning generates a partition. To emphasize our point, reconsider the drilling example, with $S = \{0, 1\}$ and a probability q for the test to be completed. This is a random partition with $p(\{0\}, \{1\}) = q$ and $p(\{0, 1\}) = 1 - q$. Suppose we enlarge the state space to be $S \times X$, where $X = \{success, failure\}$. While this state space naturally describes the DM's learning via a partition, acts that condition on X may not be available: it is plausible that the payoff of drilling rights does not depend on the success or failure of the test drill, but only on the availability of oil. Under our assumptions, the domain of acts that are defined on S is sufficient to allow the description of expected information as events.

3.2. Comparing valuations of binary bets

Fix $0 < k < c < 1 - k$ and $s, s' \in \sigma(\mu)$, such that $\{c\} \succ \{f\}$, where

$$f(\widehat{s}) = \begin{cases} c + k & \text{if } \widehat{s} = s \\ 0 & \text{if } \widehat{s} = s' \\ c & \text{otherwise} \end{cases}$$

Let

$$f'(\widehat{s}) = \begin{cases} c + k & \text{if } \widehat{s} = s \\ c & \text{otherwise} \end{cases}$$

Implicitly define $v(s)$ by $\{c + v(s)\} \sim \{f'\}$. The amount $v(s)$ is unique and nonnegative. It captures the DM's willingness to pay for additional payoffs in state s. Implicitly define $w(s, s')$ by $\{c + w(s, s')\} \sim \{c, f\}$. The amount $w(s, s')$ captures the DM's willingness to pay for the ability to bet on state s versus s'. Let

$$\zeta(s, s') = \begin{cases} \frac{w(s,s')}{v(s)} & \text{if } v(s) > 0 \\ 0 & \text{otherwise} \end{cases}$$

be the value the DM derives from being able to bet on state s versus s' in terms of the value of additional payoffs in state s. Corollary 1 in Appendix 6.6 establishes that $\zeta(s, s')$ is independent of any k and c that satisfy the premise above and that $\zeta(s, s') = \zeta(s', s)$.

Definition 12. *DM1 values binary bets more than DM2 if for all $s, s' \in S$*
(i) $v^1(s) = 0 \Leftrightarrow v^2(s) = 0$; and
(ii) $\zeta^1(s, s') \geq \zeta^2(s, s')$.

Given Theorem 3, a natural measure of the amount of information that a DM expects to receive is how likely he expects to be able to distinguish any state s from any other state s' whenever s is indeed the true state. Using Bayes' rule, $\Pr(\{I \,|s \in I,\, s' \notin I\} \,|s) = \sum_{I|s\in I,\ s'\notin I} \frac{\rho(I)}{\mu(I)}$.[12]

Theorem 5. *DM1 values binary bets more than DM2 if and only if $\sigma(\mu^1) = \sigma(\mu^2)$ and*

$$\sum_{I|s\in I,\ s'\notin I} \frac{\rho^1(I)}{\mu^1(I)} \geq \sum_{I|s\in I,\ s'\notin I} \frac{\rho^2(I)}{\mu^2(I)}$$

for all $s, s' \in \sigma(\mu^1)$

Proof. See Appendix 6.6 ■

Theorem 5 enables us to compare the behavior of two individuals who expect to learn different amounts of information, without requiring that they share the same prior beliefs. In contrast, Theorem 2 requires agreement on the prior beliefs.

We conclude this section by establishing that having more preferences for flexibility (Definition 2) is stronger than valuing binary bets more (Definition 12).

Theorem 6. *If DM1 has more preference for flexibility than DM2, then DM1 values binary bets more than DM2.*

Proof. See Appendix 6.7 ■

The Blackwell criterion for comparing information systems is often considered too strong because it only allows the comparison of information systems that generate identical underlying beliefs. We demonstrate in Theorem 2 that the behavioral counterpart of this criterion is the notion of "more preference for flexibility." The behavioral notion of "valuing binary bets more" is weaker, that is, it allows more comparisons, as established in Theorem 6. Suppose, for example, that both $\sigma(\rho^1)$ and $\sigma(\rho^2)$ form a partition of S. Then it is easy to verify that DM1 has more preference for flexibility than DM2 if and only if DM1's partition is finer and both share the same prior beliefs. In this example, the weaker comparison of "valuing binary bets more" corresponds exactly to dropping the requirement that the prior beliefs are the same.[13]

4. Subjective temporal resolution of uncertainty

Suppose that the DM anticipates uncertainty to resolve gradually over time. The pattern of resolution might be relevant if, for example, the time at which the DM has to choose an

[12] $\Pr(\{I \,|s \in I,\, s' \notin I\} \,|s) = \frac{\Pr\left(s\middle|\left\{I\middle|s\in I,\ s'\notin I\right\}\right)\Pr\left(\left\{I\middle|s\in I,\ s'\notin I\right\}\right)}{\mu(s)} = \frac{\left[\sum_{I|s\in I,\ s'\notin I} \frac{\mu(s)}{\mu(I)} \frac{\rho(I)}{\rho(\{I|s\in I,\ s'\notin I\})}\right]\rho\left(\left\{I\middle|s\in I,\ s'\notin I\right\}\right)}{\mu(s)}$ $= \sum_{I|s\in I,\ s'\notin I} \frac{\rho(I)}{\mu(I)}$.

[13] We do not provide a formal proof of this assertion at this point, as it is a corollary of Theorem 8 below.

alternative from the menu is random and continuously distributed over some interval, say $[0,1]$. An alternative interpretation is that at any given point in time $t \in [0,1]$ the DM chooses one act from the menu. At time 1, the true state of the world becomes objectively known. The DM is then paid the convex combination of the payoffs specified by all acts on the menu, where the weight assigned to each act is simply the amount of time the DM held it. That is, the DM derives a utility flow from holding a particular act, where the state-dependent flow is determined ex-post, at the point when payments are made. In both cases, the information available to the DM at any point in time t might be relevant for his choice. This section is phrased in terms of random timing of the second-stage choice. Section 5.1 discusses the utility flow interpretation in more detail.

In a context where the flow of information over time is objectively given, it is common to describe it as a filtered probability space, that is, a probability space with a filtration on its sigma algebra. We would like to replicate this description in the context of subjective learning. To that end we now refine the information set representation (μ, ρ) in Theorem 3, such that it can be interpreted as follows: the DM holds beliefs over the states of the world and has in mind a filtration indexed by continuous time. Using Bayes' law, the filtration and prior beliefs jointly generate a subjective temporal lottery. We achieve this refinement by imposing an additional axiom on $\succeq$, which uses the notion of saturated menus to imply that the support of ρ has a hierarchical structure. Our domain is rich enough to allow both the filtration, that is the timing of information arrival and the sequence of partitions induced by it, and the beliefs to be uniquely elicited from choice behavior.

Definition 13. *An act f contains act g if $\sigma(g) \subsetneq \sigma(f)$.*

Definition 14. *Acts f and g do not intersect if $\sigma(g) \cap \sigma(f) = \emptyset$.*

Axiom 9 (Hierarchy). *If F is saturated and $f, g \in F$ then either f and g do not intersect or one contains the other.*

In order to interpret two distinct information sets in $\sigma(\rho)$ that both contain state s as being relevant at different points in time, they must be ordered by set inclusion. Using Claim 3, this is the content of Axiom 9.

We now introduce exclusive trees. Such trees can be described as a pair of a probability measure μ on S and a filtration $\{\mathcal{P}_t\}$ indexed by $t \in [0,1]$.[14]

Definition 15. *The pair $(\mu, \{\mathcal{P}_t\})$ is an exclusive tree if μ is a probability measure on S and $\{\mathcal{P}_t\}$ is a filtration indexed by $t \in [0,1]$.*

[14] Slightly abusing notation, we will identify a filtration with a sequence of partitions of the state space.

Note that there can only be a finite number of times at which the filtration $\{\mathcal{P}_t\}$ becomes strictly finer. That is, there exists a finite set $\{t_1, ..., t_T\} \subset (0,1)$, such that $\mathcal{P}_{t'} \subset \mathcal{P}_t$ is equivalent to $t < t'$ if and only if $t' \in \{t_1, ..., t_T\}$. The definition does not require $\mathcal{P}_0 = \{S\}$.

Theorem 7. *The relation $\succeq$ satisfies Axioms 1–9 if and only if there is an exclusive tree, $(\mu, \{\mathcal{P}_t\})$, such that $\succeq$ is represented by*

$$V(F) = \int_{[0,1]} \left\{ \sum_{P \in \mathcal{P}_t} \max_{f \in F} \left[\sum_{s \in P} f(s)\, \mu(s) \right] \right\} dt.$$

The pair $(\mu, \{\mathcal{P}_t\})$ is unique.

Proof. See Appendix 6.8 ■

If the DM faces an (exogenously given) random stopping time that is uniformly distributed on $[0,1]$,[15] then Theorem 7 can be interpreted as if the DM holds prior beliefs μ and expects to learn over time as described by the filtration $\{\mathcal{P}_t\}$.

We now briefly sketch the proof of Theorem 7. Given an information set representation (μ, ρ) as in Theorem 3, Axiom 9 allows us to construct a random partition (q, Π) as defined at the end of Section 3.1 (i.e. $\rho(I) = \sum_{\mathcal{P} \in \Pi | I \in \mathcal{P}} q(\mathcal{P})\, \mu(I)$), where the partitions in Π can be ordered by increasing fineness. If the DM faces a random stopping time that is uniformly distributed on $[0,1]$, then it is natural to interpret $q(\mathcal{P})$ as the time for which the DM expects partition $\mathcal{P}$ to be relevant. This interpretation is captured in the time dependency of $\{\mathcal{P}_t\}$. The construction of (q, Π) is recursive. First, for each state $s \in S$, we find the largest set in $\sigma(\rho)$ that includes s. The collection of those sets constitutes $\mathcal{P}_1$. The probability $q(\mathcal{P}_1)$ corresponds to the smallest weight any of those sets is assigned by ρ relative to the prior μ, that is, $q(\mathcal{P}_1) = \min_{I \in \mathcal{P}_1} \left(\frac{\rho(I)}{\mu(I)} \right)$. To begin the next step, we calculate adjusted weights, ρ_1, as follows: for any set $I \in \mathcal{P}_1$, let $\rho_1(I) = \rho(I) - q(\mathcal{P}_1)\, \mu(I)$. For any set $I \in \sigma(\rho) \backslash \mathcal{P}_1$, let $\rho_1(I) = \rho(I)$. $\sigma(\rho_1)$ then consists of all sets $I \in \mathcal{P}_1$ with a relative weight $\frac{\rho(I)}{\mu(I)} > q(\mathcal{P}_1)$ and all sets in $\sigma(\rho) \backslash \mathcal{P}_1$. Recursively, construct $\mathcal{P}_n$ according to ρ_{n-1}. By Theorem 3, $\sum_{I \in 2^S | s \in I} \frac{\rho(I)}{\mu(I)} = 1$ for all $s \in \sigma(\mu)$, which guarantees that the inductive procedure is well defined. It must terminate in finite time due to the finiteness of 2^S.

Remark 3. *At the time of menu choice, the DM holds beliefs over all possible states of the world. If he expects additional information to arrive before time-zero (at which point his beliefs commence to be relevant for choice from the menu), then time-zero information is*

[15] It is straightforward to accommodate any other exogenous distribution of stopping times. An alternative interpretation is that the distribution of stopping times is not uniform because of an external constraint, but because the DM subscribes to the principle of insufficient reason, by which he assumes that all points in time are equally likely to be relevant for choice.

described by a non-trivial partition of $\widehat{S}$, the set of all non-null states, that is, $\mathcal{P}_0 \neq \left\{\widehat{S}\right\}$. If one wants to assume that the (subjective) flow of information cannot start before time-zero, then the following additional axiom is required.

Axiom 10 (Initial node). *If F is saturated, then there exists $f \in F$ such that f contains g for all $g \in F$ with $g \neq f$.*

Under the assumptions of Theorem 7, if $\succeq$ also satisfies Axiom 10, then $\mathcal{P}_0 = \left\{\widehat{S}\right\}$. That is, the tree $(\mu, \{\mathcal{P}_t\})$ has a unique initial node.

4.1. Revisiting the behavioral comparisons

We can characterize the notion of preference for flexibility and the value of binary bets via the DM's subjective filtration.

Definition 16. *DM1 learns earlier than DM2 if $\{\mathcal{P}_t^1\}$ is weakly finer than $\{\mathcal{P}_t^2\}$.*

Theorem 8. *(i) DM1 values binary bets more than DM2 if and only if DM1 learns earlier than DM2;*

(ii) DM1 has more preference for flexibility than DM2 if and only if DM1 learns earlier than DM2 and they have the same prior beliefs, $\mu^1 = \mu^2$.

Proof. See Appendix 6.9 ■

Theorem 8 shows that under the assumptions of Theorem 7, the characterization of "more preference for flexibility" differs from that of the weaker notion of "valuing binary bets more" solely by requiring that the prior beliefs are the same.

5. Discussion

5.1. A different interpretation: utility flow

In Section 4 we suggest that cases in which the DM derives a utility flow from choosing an act at any point in time can be accommodated in our setting. We now elaborate on this interpretation. Consider a company that produces laptop computers and is preparing the scheduled release of a new model. At any point in time prior to the launch, the company can choose one of many development strategies, each of which specifies how to allocate development effort among different features of the product. For example, one development strategy might divide the time equally between improving the screen and expanding the memory. Another might to focus exclusively on enlarging the keyboard. The value of the different collections of features at the time of launch depends on consumers' tastes and competing products, as summarized by the state of the world, and on the effort spent developing them. As the launch approaches, the company may become more informed about the underlying state of the world and may adjust its development strategy accordingly. Suppose that, given the state of the world, the value generated by the development process is the *sum* of the values added by the different strategies the company pursued prior to launch. The added value from any particular strategy is simply the value it would have generated had it been pursued consistently, weighted by the amount of time it was pursued. Formally, given a collection of possible development strategies F, let $a : [0,1] \rightarrow F$ be a development process, or a particular path of strategy choices, that is, $a(t)$ is the strategy $f \in F$ that DM chooses at time t. Given the state of the world $s \in S$, the payoff from the process a is

$$\int_{[0,1]} a(t)(s)\, dt$$

In light of this separability of payoffs over time, Theorem 7 provides an intuitive representation of choice between sets of development strategies. The representation suggests that given a set of strategies F, at every point in time the company chooses the strategy that performs best under its current beliefs: if its information at time t is P_t, then its strategy choice, $a(t)$, will satisfy

$$a(t) \in \arg\max_{f \in F} \left[\sum_{s \in S} f(s)\, \mu(s \,|P_t\,) \right]$$

Take Apple as an example of a company that many perceive as standing out from their competitors; it is generally accepted that Apple has "vision," the ability to identify the next big thing before its competitors. According to our behavioral comparison, Apple should derive more value from flexibility than the competition. At the same time, "vision" has no immediate implications for the amount of flexibility a firm chooses. One can think of research expenditures as a proxy for flexibility: the more a company spends on research, the more development options it has. Our predictions are then in line with the observation that Microsoft vastly outspends Apple on research to less effect, Apple gets more "bang for their research buck."[16]

5.2. Reevaluation of our domain

In this paper we study preferences over sets of feasible intermediate actions, or menus of acts. For the first two representation theorems (Theorems 1 and 3), we adopt the usual

[16]See http://gizmodo.com/#!5486798/research-and-development-apple-vs-microsoft-vs-sony

interpretation that the DM has to choose an alternative from a menu at some prespecified future point in time. While this interpretation of the domain allows preferences to be affected by the DM's expectations regarding the resolution of uncertainty, preferences are insensitive to the timing of resolution as long as all resolution happens before the choice of an alternative. An illustrative example is provided in Takeoka (2007), who proceeds to derive a subjective two-stage compound lottery by specifying the sets of feasible intermediate actions at *different* points in time, that is, by analyzing choice between what one might term "compound menus" (menus over menus etc.). The domain of compound menus provides a way to talk about compound uncertainty (without objective probabilities). It has the advantage that it can capture situations where the DM faces intertemporal trade-offs, for example if today's action may limit tomorrow's choices. However, while only the initial choice is modeled explicitly, the interpretation of choice on this domain now involves multiple stages, say 0, 1/2, and 1, at which the DM must make a decision. That is, the pattern of information arrival (or, at least, the collection of times at which an outside observer can detect changes in the DM's beliefs) is objectively given. In that sense, the domain only partially captures subjective temporal resolution of uncertainty. Furthermore, the domain of compound menus becomes increasingly complicated, as the resolution of uncertainty becomes finer.[17]

In Section 4 we take a different approach to study subjective temporal resolution of uncertainty: we specify a single set of feasible intermediate actions, which is the relevant constraint on choice at *all* points in time. At the first stage, the DM chooses a menu of acts and only this choice is modeled explicitly. The innovation lies in our interpretation of choice from the menu. Whether we think of an exogenous distribution for the stopping time or of a model where the DM derives a utility flow (as suggested in Section 5.1), the information that the DM has at any point in time might be relevant for the DM's ultimate choice from a menu. Our domain has the obvious disadvantage that it does not accommodate choice situations where the set of feasible actions may change over time. That said, our approach allows us (as we argue in the text) to uniquely pin down the timing of information arrival in continuous time, the sequence of induced partitions, and the DM's prior beliefs from the familiar and analytically tractable domain of menus of acts.

[17] Note that the set of menus over acts is infinitely dimensional. Hence, even the three-stage model considers menus that are subsets of an infinite dimensional space.

6. Appendix

6.1. Proof of Theorem 1

It is easily verified that any preferences with a second-order beliefs representation as in Theorem 1 satisfy the axioms. We proceed to show the sufficiency of the axioms.

We can identify $\mathcal{F}$ with the set of all $k-$dimensional vectors, where each entry is in $[0,1]$. For reasons that will become clear below, we now introduce an artificial state, s_{k+1}. Let

$$\mathcal{F}' := \left\{ f' \in [0,1]^k \times [0,k] \,\middle|\, \textstyle\sum_{i=1}^{k+1} f'(s_i) = k \right\}.$$

Note that the $k+1$ component in f' equals $k - \sum_{i=1}^{k} f'(s_i)$. For $f' \in \mathcal{F}'$, denote by $f'^k \in \mathcal{F}$ the vector that agrees with the first k components of f'. Since $\mathcal{F}$ and $\mathcal{F}'$ are isomorphic, we can look at preferences on $\mathcal{K}(\mathcal{F}')$, $\succeq_*$, defined by: $F' \succeq_* G' \Leftrightarrow F \succeq G$, where $F := \left\{ f \in \mathcal{F} \,\middle|\, f = f'^k \text{ for some } f' \in F' \right\}$ and analogously for G.

Claim 5. *The relation $\succeq_*$ satisfies the independence axiom.*

Proof. Using the definition of $\succeq_*$ and Axiom 4, we have, for all F', G', and H' in $\mathcal{K}(\mathcal{F}')$ and for all $\alpha \in [0,1]$,

$$F' \succeq_* G' \Leftrightarrow F \succeq G \Leftrightarrow \alpha F + (1-\alpha) H \succeq \alpha G + (1-\alpha) H \Leftrightarrow$$
$$(\alpha F + (1-\alpha) H)' \succeq_* (\alpha G + (1-\alpha) H)' \Leftrightarrow \alpha F' + (1-\alpha) H' \succeq_* \alpha G' + (1-\alpha) H'.$$

■

Let

$$\mathcal{F}'' := \left\{ f' \in [0,k]^{k+1} \,\middle|\, \textstyle\sum_{i=1}^{k+1} f'(s_i) = k \right\}.$$

Let $F^{k+1} := \left\{ \left(\frac{k}{k+1}, ..., \frac{k}{k+1} \right) \right\} \in \mathcal{K}(\mathcal{F}')$. Observe that for $F'' \in \mathcal{F}''$ and $\varepsilon < \frac{1}{k^2}$, $\varepsilon F'' + (1-\varepsilon) F^{k+1} \in \mathcal{K}(\mathcal{F}')$. Define $\succeq_{**}$ on $\mathcal{K}(\mathcal{F}'')$ by $F'' \succeq_{**} G'' \Leftrightarrow \varepsilon F'' + (1-\varepsilon) F^{k+1} \succeq_* \varepsilon G'' + (1-\varepsilon) F^{k+1}$ for all $\varepsilon < \frac{1}{k^2}$.

Claim 6. *The relation $\succeq_{**}$ is the unique extension of $\succeq_*$ to $\mathcal{K}(\mathcal{F}'')$ that satisfies the independence axiom.*

Proof. Note that the $(k+1)$-dimensional vector $\left(\frac{k}{k+1}, ..., \frac{k}{k+1} \right) \in int\mathcal{F}' \subset \mathcal{F}''$, hence $F^{k+1} \subset int\mathcal{F}' \subset \mathcal{F}''$. We now show that $\succeq_{**}$ satisfies independence. For any $F'', G'', H'' \in \mathcal{K}(\mathcal{F}'')$

and $\alpha \in [0,1]$,

$$\begin{aligned}
&F'' \succeq_{**} G'' \Leftrightarrow \varepsilon F'' + (1-\varepsilon) F^{k+1} \succeq_* \varepsilon G'' + (1-\varepsilon) F^{k+1} \Leftrightarrow \\
&\quad \alpha \left(\varepsilon F'' + (1-\varepsilon) F^{k+1}\right) + (1-\alpha) \left(\varepsilon H'' + (1-\varepsilon) F^{k+1}\right) \\
&\quad = \varepsilon \left(\alpha F'' + (1-\alpha) H''\right) + (1-\varepsilon) F^{k+1} \succeq_* \\
&\quad \alpha \left(\varepsilon G'' + (1-\varepsilon) F^{k+1}\right) + (1-\alpha) \left(\varepsilon H'' + (1-\varepsilon) F^{k+1}\right) \\
&\quad = \varepsilon \left(\alpha G'' + (1-\alpha) H''\right) + (1-\varepsilon) F^{k+1} \Leftrightarrow \alpha F'' + (1-\alpha) H'' \succeq_{**} \alpha G'' + (1-\alpha) H''
\end{aligned}$$

The first and third $\Leftrightarrow$ is by the definition of $\succeq_{**}$. The second $\Leftrightarrow$ is by Claim 5.[18] This argument shows that a linear extension exists. To show uniqueness, let $\widehat{\succeq}$ be any preference relation over $\mathcal{K}(\mathcal{F}'')$, which satisfies the independence axiom. By independence, $F'' \mathrel{\widehat{\succeq}} G'' \Leftrightarrow \varepsilon F'' + (1-\varepsilon) F^{k+1} \mathrel{\widehat{\succeq}} \varepsilon G'' + (1-\varepsilon) F^{k+1}$. Since $\widehat{\succeq}$ extends $\succeq_*$, they must agree on $\mathcal{K}(\mathcal{F}')$. Therefore, $\varepsilon F'' + (1-\varepsilon) F^{k+1} \mathrel{\widehat{\succeq}} \varepsilon G'' + (1-\varepsilon) F^{k+1} \Leftrightarrow \varepsilon F'' + (1-\varepsilon) F^{k+1} \succeq_* \varepsilon G'' + (1-\varepsilon) F^{k+1}$. By combining the two equivalences above, we conclude that defining $\widehat{\succeq}$ by $F'' \mathrel{\widehat{\succeq}} G'' \Leftrightarrow \varepsilon F'' + (1-\varepsilon) F^{k+1} \succeq_* \varepsilon G'' + (1-\varepsilon) F^{k+1}$ is the only admissible extension of $\succeq_*$. ■

The domain $\mathcal{K}(\mathcal{F}'')$ is formally equivalent to that of Dekel, Lipman, Rustichini, and Sarver (2007, henceforth DLRS) with $k+1$ prizes. (The unit simplex is obtained by rescaling all elements of $\mathcal{F}''$ by $1/k$, that is, by redefining $\mathcal{F}''$ as $\left\{f' \in [0,1]^{k+1} : \sum_{i=1}^{k+1} f'(s_i) = 1\right\}$.) Applying Theorem 2 in DLRS,[19] one obtains the following representation of $\succeq_{**}$:

$$\widehat{V}(F'') = \int_{\mathcal{M}(S)} \max_{f'' \in F''} \left(\sum_{s \in S \cup \{s_{k+1}\}} f''(s)\, \widehat{\pi}(s)\right) d\widehat{p}(\widehat{\pi})$$

where $\mathcal{M}(S) := \left\{\widehat{\pi} \,\middle|\, \sum_{s \in S \cup \{s_{k+1}\}} \widehat{\pi}(s) = 0 \text{ and } \sum_{s \in S \cup \{s_{k+1}\}} (\widehat{\pi}(s))^2 = 1\right\}$. Given the normalization of $\widehat{\pi} \in \mathcal{M}(S)$, $\widehat{p}(\cdot)$ is a unique probability measure. Note that $\widehat{V}$ also represents $\succeq_*$ when restricted to its domain, $\mathcal{K}(\mathcal{F}')$.

[18] The $(=)$ sign in the third and in fifth lines are due to the fact that F^{k+1} is a singleton menu. For a singleton menu $\{f\}$ and $\alpha \in (0,1)$,

$$\alpha \{f\} + (1-\alpha)\{f\} = \{f\}$$

while, for example,

$$\alpha \{f,g\} + (1-\alpha)\{f,g\} = \{f, g, \alpha f + (1-\alpha) g, \alpha g + (1-\alpha) f\},$$

is not generally equal to $\{f,g\}$.

[19] DLRS provide a supplemental appendix which shows that, for the purpose of the theorem, their stronger continuity assumption can be relaxed to the weaker notion of vNM continuity used in the present paper.

We aim for a representation of $\succeq$ of the form

$$V(F) = \int_{\Delta(S)} \max_{f \in F} \left(\sum_{s \in S} f(s) \pi(s) \right) dp(\pi),$$

where $f(\cdot)$ is a vector of utils and $p(\cdot)$ is a unique probability measure on $\Delta(S)$, the space of all probability measures on S.

We now explore the additional constraint imposed on $\widehat{V}$ by Axiom 6 and the definition of $\succeq_*$.

Claim 7. *$\widehat{\pi}(s_{k+1}) \leq \widehat{\pi}(s)$ for all $s \in S$, $\widehat{p}-$almost surely.*

Proof. Suppose there exists some event $E \subset \mathcal{M}(S)$ with $\widehat{p}(E) > 0$ and $\widehat{\pi}(s_{k+1}) > \widehat{\pi}(s)$ for some $s \in S$ and all $\widehat{\pi} \in E$. Let $f' = (0, 0, ..., 0, \varepsilon, 0, ..., k - \varepsilon)$, where ε is received in state s and $k - \varepsilon$ is received in state s_{k+1}. Let $g' = (0, 0, ..0, 0, 0, ..., k)$. Then $\{f', g'\} \succ_* \{f'\}$. Take $F' = \{f'\}$ (so that $F' \cup \{g'\} \succ_* F'$). But note that Axiom 6 and the definition of $\succeq_*$ imply that $F' \sim_* F' \cup \{g'\}$, which is a contradiction. ■

Given our construction of $\widehat{V}$, there are two natural normalizations that allow us to replace the measure $\widehat{p}$ on $\mathcal{M}(S)$ with a unique probability measure p on $\Delta(S)$.

First, since s_{k+1} is an artificial state, the representation should satisfy $\pi(s_{k+1}) = 0$, $p-$almost surely. For all $s \in S$ and for all $\widehat{\pi}$, define $\xi(\widehat{\pi}(s)) := \widehat{\pi}(s) - \widehat{\pi}(s_{k+1})$. Since $\sum_{i=1}^{k+1} f'(s_i) = k$ and ξ simply adds a constant to every $\widehat{\pi}$,

$$\arg\max_{f'' \in F''} \left(\sum_{s \in S \cup \{s_{k+1}\}} f''(s) \xi(\widehat{\pi}(s)) \right) = \arg\max_{f'' \in F''} \left(\sum_{s \in S \cup \{s_{k+1}\}} f''(s) \widehat{\pi}(s) \right)$$

for all $\widehat{\pi} \in \sigma(\widehat{p})$, the support of $\widehat{p}$. Furthermore, by Claim 7, $\xi(\widehat{\pi}(s)) \geq 0$ for all $s \in S$, $\widehat{p}-$almost surely.

Second, we would like to transform $\xi \circ \widehat{\pi}$ into a probability measure π. Let

$$\pi(s) := \xi(\widehat{\pi}(s)) / \left(\sum_{s' \in S} \xi(\widehat{\pi}(s')) \right).$$

(recall that $\xi(\widehat{\pi}(s_{k+1})) = 0$). Since this transformation affects the relative weight given to event $E \subset \mathcal{M}(S)$ in the representation, we need p to be a probability measure on I that offsets this effect, as implied by the uniqueness in DLRS. Hence, we have the Radon-Nikodym derivative

$$\frac{dp(\pi)}{d\widehat{p}(\widehat{\pi})} = \frac{\sum_{s \in S} \xi(\widehat{\pi}(s))}{\int_{\mathcal{M}(S)} \left(\sum_{s \in S} \xi(\widehat{\pi}(s)) \right) d\widehat{p}(\widehat{\pi})}.$$

Therefore, $\succeq$ is represented by

$$V(F) = \int_{\Delta(S)} \max_{f \in F} \left(\sum_{s \in S} f(s) \pi(s) \right) dp(\pi).$$

6.2. Proof of Claim 2

We will construct a menu that satisfies Definition 6 with $f(s) < 1$ for all $f \in F$ and all $s \in S$. Let $F_{\Delta(S)} := \{f \in \mathcal{F} : \|f\|_2 = 1\}$ be the positive segment of the $k-1$ dimensional unit sphere. There is an isomorphism between $\Delta(S)$ and $F_{\Delta(S)}$ with the mapping $\pi \to \arg\max_{f \in F_S} \left(\sum_{s \in S} f(s) \pi(s) \right)$. For $\mathcal{L} \subset \Delta(S)$ let $F_{\mathcal{L}} \subset F_{\Delta(S)}$ be the image of $\mathcal{L}$ under this mapping. Finiteness of $\sigma(p)$ implies that $F_{\sigma(p)}$ is finite. Let $f_{\sigma(p),\pi} := \arg\max_{f \in F_{\sigma(p)}} \left(\sum_{s \in S} f(s) \pi(s) \right)$ and (implicitly) define $\pi_{\sigma(p),f}$ by $f = \arg\max_{f \in F_{\sigma(p)}} \left(\sum_{s \in S} f(s) \pi_{\sigma(p),f}(s) \right)$. Because $F_{\Delta(S)}$ is the positive segment of a sphere, $\pi(s) > 0$ for $\pi \in \sigma(p)$ if and only if $f_{\sigma(p),\pi}(s) > 0$. This implies that $F_{\sigma(p)} \succ F_{\sigma(p)} \backslash \{f\} \cup \{f_s^0\}$ for all $f \in F_{\sigma(p)}$ and $s \in S$ with $f(s) > 0$. Hence, $F_{\sigma(p)}$ is fat-free (Definition 5). We claim that $F_{\sigma(p)}$ is a saturated menu. Consider condition (i) in Definition 6. If $f(s) = 0$, then $\pi_{\sigma(p),f}(s) = 0$. Hence, there exists $\bar{\varepsilon} > 0$ such that $F_{\sigma(p)} \sim F_{\sigma(p)} \cup \left\{ f_s^{f(s)+\varepsilon} \right\}$ for all $\varepsilon < \bar{\varepsilon}$. Finally, consider condition (ii) in Definition 6. Let $G \nsubseteq F_{\sigma(p)}$. If $F_{\sigma(p)} \cup G \sim F_{\sigma(p)}$ then the condition is trivially satisfied. Suppose $F_{\sigma(p)} \cup G \succ F_{\sigma(p)}$. Then, there exist $\pi \in \sigma(p)$ and $g \in G$ with $\sum_{s \in S} g(s) \pi(s) > \sum_{s \in S} f_{\sigma(p),\pi}(s) \pi(s)$. Then $F_{\sigma(p)} \cup G \sim \left(F_{\sigma(p)} \cup G \right) \backslash \left\{ f_{\sigma(p),\pi} \right\}$.

6.3. Proof of Claim 3

If F is saturated and $f \in F$, then there exists π such that $f = \arg\max \left(\sum_{s \in S} f(s) \pi(s) \right)$ (if not, then $F \sim F \backslash \{f\}$). We should show that if $f = \arg\max \left(\sum_{s \in S} f(s) \pi(s) \right)$, then for all $\pi' \neq \pi$, $f \notin \arg\max \left(\sum_{s \in S} f(s) \pi'(s) \right)$. Suppose to the contrary that there exist $\pi \neq \pi'$ such that $f = \arg\max \left(\sum_{s \in S} f(s) \pi(s) \right)$ and $f \in \arg\max \left(\sum_{s \in S} f(s) \pi'(s) \right)$. Then $f(s) > 0$ for all $s \in \sigma(\pi) \cup \sigma(\pi')$ by Definition 6 (i). We construct an act f', which does better than f with respect to belief π' and does not change the arg max with respect to any other belief in which f was not initially the best. Since $\pi \neq \pi'$, there exist two states, s and s', such that $\pi'(s) > \pi(s)$ and $\pi'(s') < \pi(s')$. Let

$$f'(\widehat{s}) = \begin{cases} f(\widehat{s}) \text{ if } \widehat{s} \notin \{s, s'\} \\ f(\widehat{s}) + \varepsilon \text{ if } \widehat{s} = s \\ f(\widehat{s}) - \delta \text{ if } \widehat{s} = s' \end{cases},$$

where $\varepsilon, \delta > 0$ are such that:
(1) $\varepsilon \pi' (s) - \delta \pi' (s') > 0$, and
(2) $\varepsilon \pi (s) - \delta \pi (s') < 0$.
The two conditions can be summarized as $\frac{\varepsilon}{\delta} \in \left(\frac{\pi'(s')}{\pi'(s)}, \frac{\pi(s')}{\pi(s)} \right) \subset (0, \infty)$. Note that one can make ε and δ sufficiently small (while maintaining their ratio fixed) so that, by continuity, f' does not change the arg max with respect to any other belief in which f was not initially the best. Hence $f' \notin F$ and $F \cup f' \succ F \cup f' \backslash \{g\}$ for all $g \in F \cup f'$, which is a contradiction to F being saturated.

6.4. Proof of Theorem 3

To show that the axioms are necessary for the representation, we only verify that the representation implies Axiom 8, as the other axioms follow exactly as in the case of Theorem 1. Suppose then that F is saturated with $f \in F$, and let g satisfy $\sigma (g) = \sigma (f)$ and $\{g\} \succeq \{f\}$, which implies that

$$\begin{aligned} V(\{g\}) - V(\{f\}) &= \textstyle\sum_{I \in 2^{\sigma(\mu)}} \sum_{s \in I} [g(s) - f(s)] \, \mu(s) \dfrac{\rho(I)}{\mu(I)} \\ &= \textstyle\sum_{s \in S} \sum_{I \in 2^{\sigma(\mu)} | s \in I} [g(s) - f(s)] \, \mu(s) \dfrac{\rho(I)}{\mu(I)} \\ &= \textstyle\sum_{s \in S} [g(s) - f(s)] \, \mu(s) \sum_{I \in 2^{\sigma(\mu)} | s \in I} \dfrac{\rho(I)}{\mu(I)} \\ &= \textstyle\sum_{s \in S} [g(s) - f(s)] \, \mu(s) \geq 0. \end{aligned} \tag{1}$$

Since F is saturated, Claim 3 and Claim 4 imply that there exists $I_f \in \sigma(\rho)$ such that

$$\begin{aligned} V(F) &= \left[\textstyle\sum_{s \in I_f} f(s) \, \mu(s) \right] \dfrac{\rho(I_f)}{\mu(I_f)} + \textstyle\sum_{I \in 2^{\sigma(\mu)} / I_f} \max\limits_{f' \in F / \{f\}} \left[\textstyle\sum_{s \in I} f(s) \dfrac{\mu(s)}{\mu(I)} \right] \rho(I) \\ &\leq \left[\textstyle\sum_{s \in I_f} g(s) \, \mu(s) \right] \dfrac{\rho(I_f)}{\mu(I_f)} + \textstyle\sum_{I \in 2^{\sigma(\mu)} / I_f} \max\limits_{f' \in F / \{f\}} \left[\textstyle\sum_{s \in I} f(s) \dfrac{\mu(s)}{\mu(I)} \right] \rho(I) \\ &\leq V((F \backslash \{f\}) \cup \{g\}), \end{aligned}$$

where the first inequality uses Equation (1) and the second inequality is because the addition of the act g might increase the value of the second component. Therefore, $(F \backslash \{f\}) \cup \{g\} \succeq F$.

The sufficiency part of Theorem 3 is proved using the following claims:

Claim 8. *Suppose F is saturated and $f \in F$. Then for all g with $\sigma(g) = \sigma(f)$,*

$$\{g\} \succ \{f\} \text{ implies } (F \backslash \{f\}) \cup \{g\} \succ F.$$

Proof. For $\varepsilon > 0$ small enough, let

$$h(s) = \begin{cases} f(s) + \varepsilon & \text{if } s \in \sigma(f) \\ 0 & \text{if } s \notin \sigma(f) \end{cases}$$

Then $\{g\} \succ \{h\}$ and $\sigma(h) = \sigma(g)$. Theorem 1 implies that $F \cup \{h\} \succ F$. Let

$$F' := \left\{ \arg\max_{f' \in F \cup \{h\}} \left(\textstyle\sum_{s \in S} f'(s) \pi(s) \right) \middle| \pi \in \sigma(p) \right\}.$$

Then $F' \sim F \cup \{h\}$ and F' is saturated. By Axiom 8,

$$F' \backslash \{h\} \cup \{g\} \succeq F'$$

Furthermore, $F' \backslash \{h\} \subseteq F \backslash \{f\}$ and, by Axiom 5 (Set Monotonicity), $F \backslash \{f\} \cup \{g\} \succeq F' \backslash \{h\} \cup \{g\}$. Collecting all the preference rankings established above completes the proof:

$$F \backslash \{f\} \cup \{g\} \succeq F' \backslash \{h\} \cup \{g\} \succeq F' \sim F \cup \{h\} \succ F$$

■

Claim 9. *If $\pi, \pi' \in \sigma(p)$ and $\pi \neq \pi'$ then $\sigma(\pi) \neq \sigma(\pi')$*

Proof. Suppose there are $\pi, \pi' \in \sigma(p)$, $\pi \neq \pi'$, but $\sigma(\pi) = \sigma(\pi')$. Let F_M be the saturated menu constructed in Claim 2. Then there are $f, g \in F_M$ with $f \neq g$ but $\sigma(f) = \sigma(g)$. Without loss of generality, suppose that $\{g\} \succeq \{f\}$. For $\varepsilon > 0$ small enough, let

$$h(s) = \begin{cases} g(s) + \varepsilon & \text{if } s \in \sigma(f) \\ 0 & \text{if } s \notin \sigma(f) \end{cases}$$

and let

$$F := \left\{ \arg \max_{f \in F_M \cup \{h\}} \left(\textstyle\sum_{s \in S} f(s) \pi(s) \right) | \pi \in \sigma(p) \right\}$$

F is a saturated menu with $F \sim F_M \cup \{h\}$. For $\varepsilon > 0$ small enough, $f, h \in F$. Furthermore, $\{h\} \succ \{g\} \succeq \{f\}$. Then, by Claim 8 $F \backslash \{f\} = (F \backslash \{f\}) \cup \{h\} \succ F$, which contradicts Axiom 5. ■

So far we have established that in Theorem 1 we can replace the integral over $\Delta(S)$ according to the measure p with a summation over 2^S according to the measure ρ. The uniqueness of ρ is implied by the uniqueness of p in Theorem 1.

$$V(F) = \sum_{I \in 2^S} \max_{f \in F} \left[\sum_{s \in S} f(s)\, \pi(s|I) \right] \rho(I)$$

Let $\mu(s) = \sum_{I|s \in I} \pi(s|I)\, \rho(I)$. The uniqueness of (π, p) in Theorem 1 implies that $\mu(s)$ is unique as well.

Claim 10. *For all $s, s' \in I \in \sigma(\rho)$,*

$$\frac{\pi(s|I)}{\pi(s'|I)} = \frac{\mu(s)}{\mu(s')}$$

Proof. Suppose to the contrary that there are $s, s' \in I \in \sigma(\rho)$ such that

$$\frac{\pi(s|I)}{\pi(s'|I)} < \frac{\mu(s)}{\mu(s')}.$$

Given a saturated menu F, let $f_I := \arg\max_{f \in F} \sum_{\widehat{s} \in S} f(\widehat{s})\, \pi(\widehat{s}|I)$. By continuity, and since $f_I(s') > 0$, there exists an act h with

$$h(\widehat{s}) = \begin{cases} f_I(\widehat{s}) \text{ if } \widehat{s} \notin \{s, s'\} \\ f_I(\widehat{s}) + \varepsilon \text{ if } \widehat{s} = s \\ f_I(\widehat{s}) - \delta \text{ if } \widehat{s} = s' \end{cases},$$

where $\varepsilon, \delta > 0$ are such that:

(1) $\varepsilon\mu(s) - \delta\mu(s') > 0$, and

(2) $\varepsilon\pi(s|I) - \delta\pi(s'|I) < 0$

Note that using Claim 3 and Claim 4 one can make ε and δ sufficiently small (while maintaining their ratio fixed), so that, by continuity and finiteness of $\sigma(\rho)$, h does not change the arg max with respect to any other belief in $\sigma(\rho)$. Then $\{h\} \succeq \{f_I\}$, but $F \succ F \backslash \{f_I\} \cup \{h\}$, which contradicts Axiom 8. ■

Claim 11. *For all $s \in S$ and all $I \in \sigma(\rho)$, $\pi(s|I) = \frac{\mu(s)}{\mu(I)}$.*

Proof. Using Claim 10,

$$\mu(I) := \sum_{s' \in I} \mu(s') = \frac{\mu(s)}{\pi(s|I)} \sum_{s' \in I} \pi(s'|I) = \frac{\mu(s)}{\pi(s|I)}$$
$$\Rightarrow \pi(s|I) = \frac{\mu(s)}{\mu(I)}$$

■

Claim 12. *For all* $s \in \sigma(\mu)$, $\sum_{I \in 2^S | s \in I} \frac{\rho(I)}{\mu(I)} = 1$.

Proof. Using Claim 11,

$$\mu(s) := \sum_{I \in 2^S | s \in I} \pi(s \,| I) \rho(I) = \sum_{I \in 2^S | s \in I} \frac{\rho(I)}{\mu(I)} \mu(s)$$
$$\Rightarrow \sum_{I \in 2^S | s \in I} \frac{\rho(I)}{\mu(I)} = 1$$

■

6.5. Proof of Theorem 4

We have already observed that an information set representation exists if and only if there exists $p : \Psi \to (0,1]$, such that for all I, $p(I) = \Pr(I \,| s \in I)$ for any $s \in I$. We now show that such a p exists if and only if Ψ is a generalized partition of S_Ψ.

(if) Let Ψ be a generalized partition of S_Ψ. Let $k \geq 1$ be the number of times that S_Ψ is covered by Ψ. Set $p(I) = \frac{\beta(I)}{k}$ for all $I \in \Psi$.
(only if) Suppose that $p(I) \in \mathbb{Q} \cap (0,1]$ for all $I \in \Psi$. Rewrite the vector p by expressing all entries using the smallest common denominator, $\xi \in \mathbb{N}_+$. Then Ψ is a generalized partition of size ξ. To see this, let $\beta(I) := \xi p(I)$ for all $I \in \Psi$. Then

$$\sum_{I \in \Psi | s \in I} \beta(I) = \sum_{I \in \Psi | s \in I} \xi p(I) = \xi \sum_{I \in \Psi | s \in I} \Pr(I \,| s \in I) = \xi$$

for all $s \in S_\Psi$.

It is thus left to show that if there exists $p \in (0,1]^{|\Psi|}$ such that for all $I \in \Psi$, $p(I) = \Pr(I \,| s \in I)$ for any $s \in I$, then there is also $p' \in [\mathbb{Q} \cap (0,1]]^{|\Psi|}$ with this property. Note that p is a solution for the system of linear equations $\mathbf{A}p = \mathbf{1}$, where $\mathbf{A}$ is a $|S_\Psi| \times |\Psi|$ matrix with entries $a_{i,j} \in \{0,1\}$, p is a $|\Psi| \times 1$ vector, and $\mathbf{1}$ is a $|S_\Psi| \times 1$ vector of ones.

Let $\widehat{P}$ be the set of solutions for the system $\mathbf{A}p = \mathbf{1}$. Then, there exists $X \in \mathbb{R}^k$ (with $k \leq |\Psi|$) and an affine function $f : X \to \mathbb{R}^{|\Psi|}$ such that $\widehat{p} \in \widehat{P}$ implies $\widehat{p} = f(x)$ for some $x \in X$. We first make the following two observations:

(i) there exists f as above, such that $x \in \mathbb{Q}^k$ implies $f(x) \in \mathbb{Q}^{|\Psi|}$;

(ii) there exists an open set $\widetilde{X} \subseteq \mathbb{R}^k$ such that $f(x) \in \widehat{P}$ for all $x \in \widetilde{X}$

To show (i), apply the Gauss elimination procedure to get f and X as above. Using the assumption that $\mathbf{A}$ has only rational entries, the Gauss elimination procedure (which involves a sequence of elementary operations on $\mathbf{A}$) guarantees that $x \in \mathbb{Q}^k$ implies $f(x) \in \mathbb{Q}^{|\Psi|}$.

To show (ii), suppose first that $p^* \in \widehat{P} \cap (0,1)^{|\Psi|}$ and $p^* \notin \mathbb{Q}^{|\Psi|}$. By construction, $p^* = f(x^*)$, for some $x^* \in X$. Since $p^* \in (0,1)^{|\Psi|}$ and f is affine, there exists an open ball $B_{\varepsilon}(x^*) \subset \mathbb{R}^k$ such that $f(x) \in \widehat{P} \cap (0,1)^{|\Psi|}$ for all $x \in B_{\varepsilon}(x^*)$, and in particular for $x' \in B_{\varepsilon}(x^*) \cap \mathbb{Q}^k$ $(\neq \phi)$. Then $p' = f(x') \in [\mathbb{Q} \cap (0,1]]^{|\Psi|}$. Lastly, suppose that $p^* \in \widehat{P} \cap (0,1]^{|\Psi|}$ and that there are $0 \leq l \leq |\Psi|$ sets $I \in \Psi$, for which $p(I)$ is uniquely determined to be 1. Then set those l values to 1 and repeat the above procedure for the remaining system of $|\Psi| - l$ linear equations.

6.6. Proof of Theorem 5

By Theorem 3, $v^i(s) = k\mu^i(s)$ for $i = 1, 2$. Consider the set $\{c, f\}$. Since conditional on any $I \ni s, s'$

$$\frac{\Pr(s\,|I)}{\Pr(s'\,|I)} = \frac{\mu(s)}{\mu(s')}$$

and since $\{c\} \succ \{f\}$, $\sum_{\widehat{s} \in I} f(\widehat{s}) \frac{\mu(\widehat{s})}{\mu(I)} > c \sum_{\widehat{s} \in I} \frac{\mu(\widehat{s})}{\mu(I)}$ if and only if $s \in I$ but $s' \notin I$. These are the only events in which DM expects to choose f from $\{c, f\}$. Therefore, $w^i(s, s') = k\mu^i(s) \Pr^i(\{I\,| s \in I,\, s' \notin I\}\,|s)$ and

$$\zeta_i(s, s') = \Pr{}^i(\{I\,| s \in I,\, s' \notin I\}\,|s) = \sum\nolimits_{I|s \in I,\, s' \notin I} \frac{\rho^i(I)}{\mu^i(I)}.$$

Corollary 1. *$\zeta(s, s')$ is independent of any k and c (that satisfy the premise in the beginning of the section) and $\zeta(s, s') = \zeta(s', s)$.*

Proof. The proof of Theorem 5 establishes that $\zeta(s, s') = \sum_{I|s \in I,\, s' \notin I} \frac{\rho(I)}{\mu(I)}$ independently of k and c. Theorem 3 implies that $\sum_{I|s \in I,\, s' \notin I} \frac{\rho(I)}{\mu(I)} = 1 - \sum_{I|s \in I,\, s' \notin I} \frac{\rho(I)}{\mu(I)} = \sum_{I|s \in I,\, s' \notin I} \frac{\rho(I)}{\mu(I)}$, and hence $\zeta(s, s') = \zeta(s', s)$. ■

6.7. Proof of Theorem 6

We first establish that in the context of Theorem 3, Definition 3 can be rewritten as follows:

Definition 17. *DM2's distribution of first-order beliefs is a mean-preserving spread of DM1's if and only if*

(i) $\mu^1 = \mu^2$, and

(ii) for all $I \in 2^S$

$$\sum\nolimits_{I' \subseteq I} \rho^1(I') \geq \sum\nolimits_{I' \subseteq I} \rho^2(I')$$

In light of Theorem 2 and the finiteness of $\sigma(\rho)$, it is sufficient to establish the following:

Claim 13. *DM1 has more preference for flexibility than DM2 if and only if items (i) and (ii) in Definition 17 hold.*

Proof. **(if)** DM1 expects to be better informed than DM2; He, therefore, expects to be able to imitate DM2's choice from any menu by simply ignoring the additional information (with an appropriate probability he pretends to be in a larger information set). Hence, he expects to derive weakly more value from any menu. Since both derive the same value from singletons, where there is no choice to be made from the menu (and therefore information is irrelevant), DM1 must weakly prefer a menu over a singleton whenever DM2 does.

(only if)

(i) Taking $G = \{g\}$ implies that they have the same preferences on singletons, and hence the same beliefs.

(ii) Suppose that there is $I \in 2^S$ with $\sum_{I' \subseteq I} \rho_2(I') > \sum_{I' \subseteq I} \rho_1(I')$. Obviously I is a strict subset of the support of μ. Define the act

$$f := \begin{cases} \delta > 0 \text{ if } s \in I \\ 0 \text{ if } s \notin I \end{cases}$$

Let c denote the constant act that gives $c > 0$ in every state, such that $\delta > c > \frac{\mu(I)}{\mu(I'')}\delta$ for all I'' that are a strict super set of I. Then $V_i(\{f, c\}) = c + (\delta - c)\sum_{I' \subseteq I} \rho_i(I')$. Finally, pick c' such that

$$(\delta - c)\sum_{I' \subseteq I} \rho_2(I') > c' - c > (\delta - c)\sum_{I' \subseteq I} \rho_1(I')$$

to find $\{f, c\} \succ_2 \{c'\}$ but $\{c'\} \succ_1 \{f, c\}$, and hence DM1 cannot have more preference for flexibility than DM2. ∎

We are now ready to prove Theorem 6, which states that if DM1 has more preferences for flexibility than DM2, then DM1 values binary bets more than DM2.

Proof of Theorem 6. Suppose $\mu^1 = \mu^2$. Then DM1 values binary bets more than DM2 if and only if $\sum_{I|s \in I,\, s' \notin I} \frac{\rho^1(I) - \rho^2(I)}{\mu(I)} \geq 0$ for all $s, s' \in \sigma(\mu^1)$. In particular, for any $I \in \sigma(\rho)$ the condition holds for all $s, s' \in I$, and thus $\sum_{I' \subseteq I} \frac{\rho^1(I') - \rho^2(I')}{\mu(I')} \geq 0$ must hold. We now show that this condition is implied by item *(ii)* in Definition 17. That is, we show that if there exists an I for which $\sum_{I' \subseteq I} \frac{\rho^2(I') - \rho^1(I')}{\mu(I')} > 0$, then there exists I'' such that $\sum_{I' \subseteq I''} \rho^2(I') > \sum_{I' \subseteq I''} \rho^1(I')$. Suppose this is not the case. Then $\rho^1(I) - \rho^2(I) \geq 0$ for any singleton I. For any I with $|I| = 2$ we must have $\rho^2(I) - \rho^1(I) \leq \sum_{I' \subseteq I} \rho^1(I') - \rho^2(I')$. But then, since μ is increasing with respect to the order of set inclusion, $\frac{\rho^2(I) - \rho^1(I)}{\mu(I)} \leq \sum_{I' \subseteq I} \frac{\rho^1(I') - \rho^2(I')}{\mu(I')}$ also holds. Continue inductively in this manner to establish that for any I we must have $\rho^2(I) - \rho^1(I) \leq \sum_{I' \subseteq I} \rho^1(I') - \rho^2(I')$, which implies that there is no I for which $\sum_{I' \subseteq I} \frac{\rho^2(I') - \rho^1(I')}{\mu(I')} > 0$, a contradiction. ∎

6.8. Proof of Theorem 7

It is easy to check that any preferences with an exclusive tree representation as in Theorem 7 satisfy Axiom 9. The rest of the axioms are satisfied since Theorem 7 is a special case of Theorem 3.

To show sufficiency, first observe that by Axiom 9 and Claim 3, I,$I' \in \sigma(\rho)$ implies that either $I \subset I'$, or $I' \subset I$, or $I \cap I' = \emptyset$. This guarantees that for any $M \subset \sigma(\rho)$ and $s \in \sigma(\mu)$, $\arg\max_{I \in M} \{|I| \, | s \in I\}$ is unique if it exists.

For any state $s \in \sigma(\mu)$, let $I_1^s = \arg\max_{I \in \sigma(\rho)} \{|I| \, | s \in I\}$. Define $T_1 := \{I_1^s \, | s \in \sigma(\mu)\}$. Let $\eta_1 = \min_{I \in T^1} \left(\frac{\rho(I)}{\mu(I)}\right)$. Set

$$\rho_1(I) = \begin{cases} \rho(I) - \eta_1 \mu(I) & \text{if } I \in T_1 \\ \rho(I) & \text{if } I \notin T_1 \end{cases}$$

Let $\rho_n : \sigma(\rho) \to [0,1]$ for $n \in \mathbb{N}$. Inductively, if for all $s \in \sigma(\mu)$ there exists $I \in \sigma(\rho_n)$ such that $s \in I$, then for any $s \in \sigma(\mu)$ let $I_{n+1}^s = \arg\max_{I \in \sigma(\rho_n)} \{|I| \, | s \in I\}$. Define $T_{n+1} := \{I_{n+1}^s \, | s \in \sigma(\mu)\}$. Let $\eta_{n+1} = \min_{I \in T_{n+1}} \left(\frac{\rho_n(I)}{\mu(I)}\right)$. Set

$$\rho_{n+1}(I) = \begin{cases} \rho_n(I) - \eta_{n+1} \mu(I) & \text{if } I \in T_{n+1} \\ \rho_n(I) & \text{if } I \notin T_{n+1} \end{cases}$$

Let $N+1$ be the first iteration in which there exists $s \in \sigma(\mu)$ which is not included in any $I \in \sigma(\rho_N)$. Axiom 7 implies that N is finite and that $(T^n)_{n=1,..,N}$ is a sequence of increasingly finer partitions, that is, for $m > n$, $I_m^s \subseteq I_n^s$ for all s, with strict inclusion for some s.

Claim 14. *$\rho(I) = \mu(I) \sum_{n \leq N | I \in T_n} \eta_n$ for all $I \in \sigma(\rho)$.*

Proof. First note that by the definition of N, $\rho(I) \geq \mu(I) \sum_{n \leq N | I \in T_n} \eta_n$ for all $I \in \sigma(\rho)$. If the claim were not true, then there would exist $I' \in \sigma(\rho)$ such that $\rho(I') > \mu(I') \sum_{n \leq N | I' \in T_n} \eta_n$. Pick $s' \in I'$. At the same time, by the definition of N, there exists $s'' \in \sigma(\mu)$ such that if $s'' \in I \in \sigma(\rho)$ then $\rho(I) = \mu(I) \sum_{n \leq N | I \in T_n} \eta_n$. We have,

$$\begin{aligned} \mu(s'') &= \sum_{I \in \sigma(\rho)} \Pr(s'' | I) \rho(I) = \sum_{I \in \sigma(\rho)} \Pr(s'' | I) \mu(I) \sum_{n \leq N | I \in T_n} \eta_n \\ &= \sum_{n \leq N} \Pr\left(s'' \middle| I_n^{s''}\right) \mu\left(I_n^{s''}\right) \eta_n = \mu(s'') \sum_{n \leq N} \eta_n \end{aligned}$$

where the last equality follows from Claim 11. Therefore, $\sum_{n\leq N}\eta_n = 1$. At the same time

$$\begin{aligned}\mu(s') &= \textstyle\sum_{I\in\sigma(\rho)} \Pr(s'\,|I)\,\rho(I) > \sum_{I\in\sigma(\rho)} \Pr(s'\,|I)\,\mu(I)\sum_{n\leq N|I\in T_n}\eta_n \\ &= \textstyle\sum_{n\leq N} \Pr\left(s'\left|I_n^{s'}\right.\right)\mu\left(I_n^{s'}\right)\eta_n = \mu(s')\sum_{n\leq N}\eta_n = \mu(s'),\end{aligned}$$

which is a contradiction. ■

Claim 14 implies that $\sigma(\rho_{N+1}) = \emptyset$. Let $\eta_m := 0$ and for $t \in [0,1)$ define the filtration $\{\mathcal{P}_t\}$ by

$$\mathcal{P}_t := T_n, \text{ for } n \text{ such that } \textstyle\sum_{m=0}^{n-1}\eta_m \leq t < \sum_{m=0}^{n}\eta_m.$$

The pair $(\mu, \{\mathcal{P}_t\})$ is thus an exclusive tree.

Claim 15. *If $\succeq$ also satisfies Axiom 10, then $\mathcal{P}_0 = \{\sigma(\mu)\}$*

Proof. Suppose to the contrary, that there are $\{S', S''\} \subset \mathcal{P}_0$ such that $S' \cap S'' = \emptyset$ and $S' \cup S'' \subseteq \sigma(\mu)$. Then, any saturated F includes some act h with $\sigma(h) \subset S'$ and another act g with $\sigma(g) \subset S''$, but it does not include an act that contains both h and g, which contradicts Axiom 10. ■

Claim 16. *If $\left(\widehat{\mu}, \left\{\widehat{\mathcal{P}}_t\right\}\right)$ induces a representation as in Theorem 7, then $\left(\widehat{\mu}, \left\{\widehat{\mathcal{P}}_t\right\}\right) = (\mu, \{\mathcal{P}_t\})$.*

Proof. μ is unique according to Theorem 3. Suppose that $\{\mathcal{P}_t\} \neq \left\{\widehat{\mathcal{P}}_t\right\}$. Then, without loss of generality, there exists t and $I \in \sigma(\rho)$, such that $I \in \mathcal{P}_t$ and $\widehat{I} \subset I \in \widehat{\mathcal{P}}_t$. Let $M = \{I' \in \sigma(\rho) : I \subseteq I'\}$. Then, according to $(\mu, \{\mathcal{P}_t\})$, $\rho(M) \geq t$, while according to $\left(\widehat{\mu}, \left\{\widehat{\mathcal{P}}_t\right\}\right)$, $\rho(M) < t$, which is a contradiction. ■

6.9. Proof of Theorem 8

(i) DM1 does not learn earlier than DM2 $\Leftrightarrow$ there exists t such that $\mathcal{P}_t^1$ is not finer than $\mathcal{P}_t^2$ $\Leftrightarrow$ there exists two states s, s', such that $s, s' \in I$ for some $I \in \mathcal{P}_t^1$, but $s, s' \notin I'$ for any $I' \in \mathcal{P}_t^2$ $\Leftrightarrow$ $\Pr^2(\{I\,|s \in I,\, s' \notin I\}\,|s) = \zeta^2(s, s') \geq 1 - t$, but $\Pr^1(\{I\,|s \in I,\, s' \notin I\}\,|s) = \zeta^1(s, s') < 1 - t \Leftrightarrow$ DM1 does not value binary bets more than DM2.

(ii) For $i = 1, 2$, let

$$t^i(I) = \min\left\{t\left|I \text{ is measurable in } \mathcal{P}_t^i\right.\right\}$$

if defined, otherwise let $t^i(I) = 1$. Let

$$\Delta^i(I) = \max\left\{t\left|I \in \mathcal{P}_t^i\right.\right\} - \min\left\{t\left|I \in \mathcal{P}_t^i\right.\right\}$$

if defined, otherwise let $\Delta^i(I) = 0$. Under the assumptions of Theorem 7,

$$\sum_{I' \subseteq I} \rho^i(I') = \sum_{I' \subseteq I} \mu^i(I') \Delta^i(I) = \mu^i(I) \left(1 - t^i(I)\right)$$

Hence, DM1 learns more than DM2 if and only if $\mu^1 = \mu^2$ and $t^1(I) \leq t^2(I)$ for all I, which is equivalent to $\{\mathcal{P}_t^1\}$ being weakly finer than $\{\mathcal{P}_t^2\}$.

References

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

[2] Blackwell, D. (1951) "The comparison of experiments," In: J. Neyman, Editor, Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability, Univ. of California Press, Berkeley, 93–102.

[3] Blackwell, D. (1953), "Equivalent comparison of experiments," *Ann. Math. Statist.*, 24, 265–272.

[4] Dekel, E., B. Lipman, and A. Rustichini (2001), "Representing Preferences with a Unique Subjective State Space," *Econometrica*, 69, 891–934.

[5] Dekel, E., B. Lipman, and A. Rustichini (2009), "Temptation-driven preferences," *Review of Economic Studies*, 76 (3), 937–971.

[6] Dekel, E., B. Lipman, A. Rustichini, and T. Sarver (2007), "Representing Preferences with a Unique Subjective State Space: Corrigendum," *Econometrica*, 75, 591–600. See also supplemental appendix.

[7] Dillenberger, D., J. Lleras, and P. Sadowski (2011), "Partitional Information," mimeo.

[8] Ergin, H., and T. Sarver (2010), "A Unique Costly Contemplation Representation," *Econometrica*, 78, 1285-1339.

[9] Gilboa, I. (2009), " Theory of Decision under Uncertainty," Cambridge University Press.

[10] Gilboa, I., A. Postlewaite, and D. Schmeidler (2009), "Is It Always Rational to Satisfy Savage's Axioms?," *Economics and Philosophy*, 25, 285–296.

[11] Gilboa, I., A. Postlewaite, and D. Schmeidler (2009), "Rationality of Belief Or: Why Savage's axioms are neither necessary nor sufficient for rationality," *Synthese*, forthcoming.

[12] Gollier, C., (2001), "The Economics of Risk and Time," The MIT Press, 0-262-07215-7.

[13] Higashi, Y., K. Hyogo, and N. Takeoka (2009), "Subjective Random Discounting and Intertemporal Choice," *Journal of Economic Theory*, 144 (3), 1015–1053.

[14] Hyogo, K. (2007), "A Subjective Model of Experimentation," *Journal of Economic Theory*, 133, 316–330.

[15] Kreps, D. (1979), "A Representation Theorem for "Preference for Flexibility," *Econometrica*, 47, 565–577.

[16] Krishna, R. V., and P. Sadowski (2011), "Dynamic Preference for Flexibility," mimeo.

[17] Lehrer, E. (2008), "Partially Specified Probabilities: Decisions and Games," *American Economic Journal: Microeconomics*, forthcoming.

[18] Savage, J. (1954), "The Foundations of Statistics." New York: John Wiley & Sons.

[19] Shapley, L. S. (1967), "On Balanced Sets and Cores." *Naval Research Logistics Quarterly,* 14, 453–460.

[20] Takeoka, N. (2007), "Subjective Probability over a Subjective Decision Tree." *Journal of Economic Theory*, 136, 536–571.

[21] Takeoka, N. (2004), "Subjective Prior over Subjective States, Stochastic Choice, and Updating," unpublished manuscript.