Evaluating Ambiguous Random Variables and Updating by Proxy[†]

Faruk Gul

and

Wolfgang Pesendorfer

Princeton University

June 2018

Abstract

We introduce a new theory of belief revision under ambiguity. It is recursive (random variables are evaluated by backward induction) and consequentialist (the conditional expectation of any random variable depends only on the values the random variable attains on the conditioning event). Agents experience no change in preferences but may not be indifferent to the timing of resolution of uncertainty. We provide three main theorems: the first relates our rule to standard Bayesian updating; the others characterize the dynamic behavior of an agent who adopts our rule.

[†] This research was supported by grants from the National Science Foundation. We thank Tomasz Strzalecki for insightful comments on an earlier version of the paper.

1. Introduction

Consider the following Ellsberg-style experiment: a ball is drawn from an urn consisting of blue and green balls of unknown proportions and a fair coin is flipped. The decision maker is ambiguity averse and receives \$1 if the ball is blue and the coin comes up heads or if the ball is green and the coin comes up tails. Prospect 1 (Figure 1 below) describes a version of this experiment in which the agent first learns the color of the ball drawn and then the outcome of the coin flip whereas in Prospect 2, she learns the outcome of the coin flip before the ball is drawn from the urn.

Prospect 1 describes a situation in which the coin flip hedges the ambiguity of the draw and, thus, we expect its value to be the same as the value of a \$1 bet on the outcome of a fair coin. This interpretation conforms with the definition of hedging in standard models of ambiguity (Gilboa and Schmeidler (1989), Schmeidler (1989), Klibanoff, Marinacci and Mukherji (2005)). In particular, all axiomatic models set in the Anscombe-Aumann framework assume that the coin flip hedges the ambiguity in Prospect 1. The situation is different in Prospect 2; in this case, it seems plausible that the coin flip does not hedge the ambiguous draw from the urn. If it did, agents could eliminate ambiguity by randomizing over acts, as Raiffa (1961) argues.

The implicit assumption underlying much of the ambiguity literature is that ex ante randomization does not hedge while ex post randomization does. We explicitly make this assumption.[1] Thus, an ambiguity averse decision maker may prefer Prospect 1 to Prospect 2 even though these two prospects differ only in the order in which uncertainty resolves.

[1] This view, however, is not unanimous. For example, Ellis (2016) assumes that the decision maker can hedge with a coin flip even if the coin flip is revealed before the unambiguous state. Saito (2015) develops a subjective hedging model with axioms on preferences over sets of Anscombe-Aumann acts. His model identifies a preference parameter that measures the extent to which the decision maker finds hedging through randomization feasible. In our model, the sequencing of uncertainty resolution serves a role similar to Saito's parameter.

Figure 1

The main contribution of this paper is a model of belief updating consistent with the description above. Specifically, we formulate an updating rule that ensures that rolling back payoffs in the ball-first tree (Prospect 1) yields values consistent with hedging while the corresponding calculation in a coin-first tree (Prospect 2) yields values consistent with no hedging.

Our theory is recursive and consequentialist. Recursivity means that random variables that resolve gradually are evaluated by backward induction. There is no "preference change" and no preference for commitment (i.e., there is no dynamic inconsistency). Consequentialism means that the conditional expectation of any random variable depends only on the values the random variable attains on the conditioning event. However, our belief revision rule cannot, in general, satisfy the law of iterated expectation since the decision makers need not be indifferent to the manner in which uncertainty resolves. As the opening example illustrates, this sensitivity to the order of resolution of uncertainty is necessary to match the standard notion of hedging in the ambiguity literature.

As in Kreps and Porteus (1978), our model posits an intrinsic preference for how uncertainty resolves. Whether the coin toss occurs first or the ball is drawn first matters despite the fact that no decisions are made in between these two events. In the Kreps-Porteus model, the agent cares about the timing of resolution of uncertainty because she values knowing earlier or because she prefers to stay uninformed longer. Hence, the carriers of utility are the beliefs that the agent has at a given moment in time. Therefore, the amount of time that elapses between the coin toss and the draw from the urn matters in the Kreps-Porteus model. In our model, the agent cares about the *sequencing* of the resolution of uncertainty. Hence, she cares about whether the coin is tossed first or the ball is drawn first but not about how much time elapses between these two events.

The primitives of our model are random variables. These random variables could be interpreted as the composition of a general act and a cardinal utility index. A *simple evaluation* E associates a real number, $E(X)$, with each random variable X and represents the overall value of that random variable (or the utility of the underlying act). Our first result (Proposition 1) ensures that simple evaluations are Choquet integrals with respect to a totally monotone capacity.

To study updating, we define two-stage random variables. In the first stage, she learns the element of a partition $\mathcal{P} = \{B_1, \ldots, B_k\}$ of the state space. In the second stage, the agent learns the state. A *compound evaluation*, $E_{\mathcal{P}}(X)$, is the value of a random variable X when the partition is $\mathcal{P}$. Even though the agent takes no action at the intermediate stage, the compound evaluation of X depends on the way uncertainty resolves. Assume, for the moment, that we have an updating rule and let $E(X|B)$ denote the conditional evaluation of X given $B \in \mathcal{P}$. As usual, we define $E(X \mid \mathcal{P})$ to be the random variable Y such that

$$Y(s) = E(X \mid B) \text{ for } B \text{ such that } s \in B \in \mathcal{P}$$

Then, we can state our recursivity assumption as follow:

$$E_{\mathcal{P}}(X) = E(E(X \mid \mathcal{P})) \tag{1}$$

that is, the compound evaluation is computed by backward induction; first for every state s in $B \in \mathcal{P}$, we set $Y(s)$ equal to the conditional expectation $E(X \mid B)$ of X and then, we evaluate Y. The missing ingredient in this formulation is the new updating rule $E(\cdot \mid \cdot)$.

Our main result, Theorem 1, characterizes a new updating rule assuming the following weakening of the law of iterated expectation: $E(X \mid B) \leq c$ for all $B \in \mathcal{P}$ implies $E(X) \leq c$. We can paraphrase this condition as saying that *not all news can be bad news*. The law of iterated expectation implies this condition but the converse is not true. In particular, it allows for the possibility that all news is good news. To see why ambiguity can lead to this violation of the law of iterated expectation, consider again the introductory example in Figure 1, above. Let Prospect 3 be the version of Prospect 1 in which the coin toss and the draw from the urn are *simultaneous* and assume that simultaneous randomization does not hedge the ambiguous draw from the urn. Thus, the ambiguity averse agent assigns the same value to Prospects 2 and 3. If c is the value of a \$1 bet on the coin, then we expect the value of Prospect 3 to be less than c. By contrast, in Prospect 1 the conditional value of the prospect is c under either outcome of the drawn from the urn and, therefore, all news is good news. In Prospect 1, uncertainty resolves in a manner that eliminates ambiguity which leads to the seemingly paradoxical situation that all news is good news.

For most of the analysis, we assume that the information that agents receive is a partition of the "payoff relevant" state space; that is, on the domain of the random variables in question. In sections 5 and 6, we allow for a more general information structures that may not be of a partition form. Thus, agents receive signals that may affect the likelihood of states without necessarily ruling them out. We call compound evaluations with this more general information structure, general compound evaluations. The more general informational structure is better suited for studying situations in which the analyst does not know the information structure of the agent. Our second main result shows that, given our updating rule, every generalized compound evaluation corresponds to the maxmin rule with some set of priors. Our third main result is the converse of our our second: every maxmin evaluation can be approximated arbitrary closely by some generalized compound evaluation. Hence, our model provides an interpretation of maxmin expected utility as Choquet expected utility with gradual resolution of uncertainty.

1.1 Related Literature

As is typical for ambiguity models, our theory does not satisfy the law of iterated expectation; that is, $E(X)$ need not equal $E(E(X|\mathcal{P}))$. Epstein and Schneider (2003) ensure that the law of iterated expectation is satisfied by restricting the conditioning events.[2] In contrast to Epstein and Schneider (2003), we consider all conditioning events. Restricting the possible conditioning events renders ambiguity theory incapable of analyzing problems

[2] Specifically, Epstein and Schneider (2003) restrict attention to event partiti?ons that ensure that the priors of for the maxmin expectation satisfy a rectangularity condition.

in which decision makers choose or affect their information as is the case in recent models of costly information processing, persuasion, and in many other traditional models in information economics.

Siniscalchi (2011) analyzes choice in decision trees (i.e., trees with both chance and decision nodes). He axiomatizes the following behavior: initially, the agent has a set of priors and at each decision node derives a new set by updating prior-by-prior. The agent makes decisions at each node according to the maxmin criterion given her current set of priors and is indifferent to how the remaining uncertainty will resolve. As is well-know, maxmin expected utility with prior-by-prior updating leads to dynamic inconsistency. Siniscalchi's agent deals with this inconsistency by treating the behavior of her future selves as constraints. Thus, as in Strotz (1955), the agent is sophisticated and engages in consistent planning. In contrast, our agent cares about how uncertainty resolves. Since gradually resolving random variables are evaluated recursively, dynamic consistency is guaranteed.

Hanany and Klibanoff (2009) posit a non-consequentialist updating rule. In their model, the value of $E(X|B)$ changes according to the ex ante optimal plan. Thus, Hanany and Klibanoff (2009) give up consequentialism and enforce equality of $E(X)$ and $E(E(X|\mathcal{P}))$. By contrast, we attribute the difference between $E(X)$ and $E(E(X|\mathcal{P}))$ to differences in the sequence of resolution of uncertainty.[3] As a result, ambiguity does not generate a preference for commitment in our model.

The two most studied updating rules under ambiguity are *prior-by-prior* updating and the Dempster-Shafer rule. The latter was introduced by Dempster (1967) and Shafer (1976) and axiomatized by Gilboa and Schmeidler (1993). The former was analyzed by Wasserman and Kadane (1990), Jaffray (1992) and axiomatized by Pires (2002). Below, we provide examples that show that prior-by-prior updating and the Dempster-Shafer rule violate the "not all news can be bad news" condition. Seidenfeld and Wassermann (1993) observe this feature of prior-by-prior updating and call it *dilation*.

[3] Note, however, that in our model the resolution of uncertainty only affects the agent's evaluation if he is not an expected value maximizer while Kreps and Porteus (1978) consider expected utility maximizers who care about the manner in which uncertainty resolves.

2. Evaluations

Let $S = \{1, 2, \ldots\}$ be the set of all states. Any (finite) subset of S is a (finite) event; the letters A, B, C, D denote generic events; $\mathcal{P}, \mathcal{Q}$ denote generic partitions of S. A *random variable* is a function from S to the set of non-negative real numbers. We let X, Y and Z denote generic random variables. For any A and random variable X, we write X_A to denote the random variable Y such that $Y(s) = X(s)$ for $s \in A$ and $Y(s) = 0$ otherwise. We identify a constant random variable X with the corresponding constant x; thus 1_A describes the random variable that has value 1 at $s \in A$ and 0 otherwise. The restriction to positive random variables is for convenience.

Our primitive is a function E that assigns a non-negative real number to each random variable. A random variable Y is *E-null* if $E(X+Y) = E(X)$. An event A is E-null if X_A is null for all X. We say that A is a support for E if $E(X) = E(X_A)$ for all X.

Random variables can be interpreted as the values of a cardinal utility index; the evaluation of these random variables can be interpreted as the overall utility of the corresponding act. By now, the axioms needed to derive the utility index for maxmin/Choquet expected utility theory are well-understood.[4] Once the utility index is at hand, translating our assumptions on the evaluation operator to preference statements is straightforward. We will impose three properties on E and call the functions that satisfy these properties *simple evaluations*. Our first assumption is *finiteness and normalization*:

P1: *E has a finite support and $E(1) = 1$.*

The existence of a finite support facilitates a simpler treatment. The requirement $E(1) = 1$ is a normalization. We call the next property the *Lebesgue property* since it is related to the definition of the Lebesgue integral. This property plays the same role here as does comonotonic independence in Choquet expected utility theory and is closely related to a property called put-call parity in Cerreia-Vioglio, Maccheroni, Marinacci (2015).

P2: $E(X) = E(\min\{X, \gamma\}) + E(\max\{X - \gamma, 0\})$ *for all* $\gamma \geq 0$.

Property 3 is *superhedging*. It is stronger than the standard assumption of uncertainty aversion and ensures that the capacity associated with the Choquet-integral representation

[4] See Schmeidler (1989), Gilboa and Schmeidler (1989), and Ghirardato and Marinacci (2001) for a unifying treatment.

of the evaluation is totally monotone. Given any collection of random variables $\mathcal{X} = (X_1, \dots, X_n)$, $\alpha \geq 0$ and $A \subset S$, let

$$K_\alpha(\mathcal{X}, A) = |\{i \,|\, X_i(s) \geq \alpha \text{ for all } s \in A\}|$$

where $|A|$ denotes the cardinality of the set A. The collection $\mathcal{X} = (X_1, \dots, X_n)$ set-wise dominates $\mathcal{Y} = (Y_1, \dots, Y_n)$ if $K_\alpha(\mathcal{X}, A) \geq K_\alpha(\mathcal{Y}, A)$ for all α, A.

P3: *$(X_1, \dots, X_n)$ set-wise dominates $(Y_1, \dots, Y_n)$ implies $\sum_i E(X_i) \geq \sum_i E(Y_i)$.*

To interpret P3, we extend the agents utility to lotteries over random variables. P3 then says that the agent prefers a random draw from the collection $\mathcal{X}$ over a random draw from $\mathcal{Y}$ if $\mathcal{X}$ set-wise dominates $\mathcal{Y}$.

A (simple) capacity is a function $\pi : 2^S \to [0, 1]$ such that (i) $\pi(\emptyset) = 0$, (ii) there is a finite D such that $\pi(A) = \pi(A \cap D)$ for all A, (iii) $\pi(A) \leq \pi(B)$ whenever $A \subset B$. We say that $\mu : 2^S \to I\!R$ is the Möbius transform of π if

$$\pi(A) = \sum_{B \subset A} \mu(B)$$

A simple, inductive argument establishes that every capacity has a unique Möbius transform μ_π. The capacity π is *totally monotone* if $\mu_\pi(A) \geq 0$ for all A. Total monotonicity implies supermodularity; that is,

$$\pi(A \cup B) + \pi(A \cap B) \geq \pi(A) + \pi(B)$$

for all $A \in 2^S$.

For any random variable X and capacity π, the Choquet integral of X with respect to the capacity π is defined as follows: let $\alpha_1 > \cdots > \alpha_n$ be all of the nonzero values that X takes and set $\alpha_{n+1} = 0$. Then

$$\int f d\pi := \sum_{i=1}^{n} (\alpha_i - \alpha_{i+1}) \pi(\{s \,|\, X(s) \geq \alpha_i\})$$

Proposition 1 below establishes that every simple evaluation has a Choquet integral representation with a totally monotone capacity. It is clear that the capacity π in the integral representation of E is unique. Hence, we call this π the capacity of E and call E the evaluation with capacity π. All proofs are in the appendix.

Proposition 1: *E is a simple evaluation if and only if there is a totally monotone capacity π such that $E(X) = \int X d\pi$ for all X.*

Every simple evaluation can also be described as a maxmin evaluation. Let $\Delta(\pi) := \{p \,|\, p(A) \geq \pi(A) \text{ for all } A\}$ be the core of E's capacity π and define

$$E^m(X) = \min_{p \in \Delta(\pi)} \int X dp$$

Since a totally monotone capacity is supermodular, it follows (Schmeidler (1989)) that $E(X) = E^m(X)$ for all X.

The set D is a support for π if $\pi(D \cap A) = \pi(A)$ for all A. Note that a set is a support for E if and only if it is a support for E's capacity. Every capacity and hence every simple evaluation has a (unique) minimal support; that is, a support that is contained in every other support. We let $S_E = S_\pi$ denote this support. We call the set $S_X = \{s \,|\, X(s) > 0\}$, the support of X. Let $\mathcal{E}$ denote the set of all simple evaluations and let $\mathcal{E}^o \subset \mathcal{E}$ denote the subset of simple evaluations that have a probability as their capacity. Hence, the elements of $\mathcal{E}^o$ are expectations.

We conclude this section with a definition of unambiguous random variables and events. Let E be a simple evaluation with capacity π. A random variable Y is E-*unambiguous* if $E(X + Y) = E(X) + E(Y)$ for all X. An event A is E-unambiguous if $E(X_A) + E(X_{A^c}) = E(X)$ for all X. Define $\mathcal{A}_\pi(s) = \{A \subset S \,|\, s \in A, \mu_\pi(A) \neq 0\}$.

Proposition 2: *Let E be a simple evaluation with capacity π. The following conditions are equivalent: (i) B is E-unambiguous. (ii) $A_\pi(s) \subset B$ for all $s \in B$. (iii) $\pi(B)+\pi(B^c) = 1$.*

Parts (i) and (ii) of Proposition 2 show that our definition of unambiguous events coincides with the standard definition of unambiguous events based on capacities. Part (iii) provides an alternative characterization based on the Möbius transform.

3. Compound Evaluations and Proxy Updating

A two-stage random variable is a pair $(X, \mathcal{P})$ such that X is a random variable and $\mathcal{P}$ is a finite partition. Let $E_{\mathcal{P}}(X)$ denote the (compound) evaluation of the two stage random variable $(X, \mathcal{P})$. For any E and E-nonnull event B, let $E(\cdot \,|\, B)$ denote the conditional of E given B. This conditional is itself a simple evaluation. Hence, an updating rule maps every simple evaluation E and E-nonnull B to a simple evaluation with support contained in B. If B is E-null, we let $E(\cdot \,|\, B)$ be unspecified. For every X and finite partition $\mathcal{P}$ of S, let $E(X \,|\, \mathcal{P})$ denote the random variable Y such that $Y(s) = E(X \,|\, B)$ whenever $s \in B$ and B is E-nonnull. The agent evaluates the two-stage random variable $(X, \mathcal{P})$ recursively; that is,

$$E_{\mathcal{P}}(X) = E(E(X \,|\, \mathcal{P}))$$

Note that we are able to avoid specifying $E(\cdot \,|\, B)$ for E-null B because $E(\cdot \,|\, B)$ is multiplied by 0 when computing $E(E(X \,|\, \mathcal{P}))$.

We interpret our updating rule as a two-stage procedure. In the first stage, the agent forms a proxy of the original capacity that renders information cells unambiguous. Then, she updates the proxy according to Bayes' rule. Let π be the capacity of a simple evaluation E and let μ_π be the corresponding Möbius transform. The proxy depends on the information partition $\mathcal{P}$; accordingly, we write $\pi^{\mathcal{P}}$ for π's proxy and $\mu_\pi^{\mathcal{P}}$ the proxy's Möbius transform. Equation (6), below, defines the Möbius transform of the proxy: for all events A,

$$\mu_\pi^{\mathcal{P}}(A) = \sum_{B \in \mathcal{P}} \sum_{\{D : D \cap B = A\}} \frac{|A|}{|D|} \cdot \mu_\pi(D) \tag{2}$$

Then, π's proxy capacity is the capacity $\pi^{\mathcal{P}}$ such that $\pi^{\mathcal{P}}(A) = \sum_{C \subset A} \mu_\pi^{\mathcal{P}}(C)$. We refer to the evaluation corresponding to $\pi^{\mathcal{P}}$ as the proxy evaluation $E^{\mathcal{P}}$, that is, $E^{\mathcal{P}}(X) = \int X d\pi^{\mathcal{P}}$. Given the proxy capacity $\pi^{\mathcal{P}}$, the agent updates according to Bayes' rule:

$$\pi(A \,|\, B) := \frac{\pi^{\mathcal{P}}(A \cap B)}{\pi^{\mathcal{P}}(B)} \tag{3}$$

Definition: *Let π be the capacity of the simple evaluation E, let B be an E-nonnull event and let $\pi(\cdot \,|\, B)$ be as defined in equations (2) and (3). The proxy update, $E(\cdot \,|\, B)$, of E is the evaluation with capacity $\pi(\cdot \,|\, B)$; that is, $E(X \,|\, B) = \int X d\pi(\cdot \,|\, B)$ for all X.*

The proxy of a capacity modifies it in a manner that renders every element of the information partition unambiguous. To do so, the Möbius transform weight of any event A is apportioned to events $A \cap B$ for every B in the information partition, in proportion to the number of elements in $A \cap B$. As a result of this apportionment, no event that intersects multiple elements of the information partition is given positive weight by the Möbius transform of the proxy capacity. Then, Proposition 2 implies that every element of the information partition is $E_{\mathcal{P}}$-unambiguous. This feature of the proxy reflects the view that once an event occurs, it is no longer ambiguous and ensures that the proxy of π entails less ambiguity than π; that is,

$$E_{\mathcal{P}}(X) \geq E(X)$$

for all X. If π is a probability, its proxy coincides with the original capacity since $\mu_\pi(A) = 0$ for all non-singleton events. More generally, a capacity and its proxy coincide whenever every $B \in \mathcal{P}$ is $E-$unambiguous. This follows since, by Proposition 2(ii), if B is unambiguous, there is no C with $\mu_\pi(C) > 0$ that intersects both B and its complement. Then, equation (2) implies $\mu_\pi^{\mathcal{P}} = \mu_\pi$. Finally, note that $\pi(\cdot \mid B)$ does not depend on the information partition $\mathcal{P}$ as long as $B \in \mathcal{P}$. This follows since, by equation (2), $\pi^{\mathcal{P}}(A) = \pi^{\mathcal{P}'}(A)$ for all $A \subset B \in \mathcal{P} \cap \mathcal{P}'$.

An alternative way to describe the proxy updating formula is the following: call E an *elementary evaluation* if there is some finite set D such that $E(X) = \min_{s \in D} X(s)$ for all random variables X. The capacity of an elementary evaluation has the form

$$\pi(A) = \begin{cases} 1 & \text{if } D \subset A \\ 0 & \text{otherwise.} \end{cases}$$

We let E_D denote the elementary evaluation with support D and let π_D denote its capacity. It is easy to see that every capacity can be expressed as a convex combination of elementary evaluations; that is,

$$\pi = \sum_{D \subset S} \mu_\pi(D) \cdot \pi_D$$

where μ_π is the Möbius transform of π. Then, the proxy update of π conditional any nonnull B is:

$$\pi(C \mid B) = \frac{\sum_{D \subset S} \mu_\pi(D) \frac{|B \cap D|}{|D|} \pi_{D \cap B}(C)}{\sum_{D \subset S} \mu_\pi(D) \frac{|B \cap D|}{|D|}} \tag{4}$$

In the proof of Theorem 1, we establish the equivalence of the proxy updating described in equation (4) and the rule described by equations (2) and (3).

As we show in the next section, proxy updating satisfies the "not all news can be bad news" property:

$$E(X \mid B) \leq \alpha \text{ for all } B \in \mathcal{P} \text{ implies } E(X) \leq \alpha \tag{C4}$$

Property (C4) distinguishes proxy updating from the best known existing rules, prior-by-prior updating and the Dempster-Shafer rule. To facilitate comparisons between proxy updating and these two alternatives, first we will define the compound evaluations implied by prior-by-prior updating and by the Dempster-Shafer rule.

Recall that $\Delta(\pi)$ are the set of priors in the core of π. Let $\Delta^B(\pi) = \{p(\cdot \mid B) \mid p \in \Delta(\pi) \text{ and } p(B) > 0\}$ be the set of (Bayesian) updated priors in $\Delta(\pi)$ conditional on B. Given a simple evaluation E and E-nonnull set B, the prior-by-prior updating rule yields the conditional evaluation $E^m(\cdot \mid B)$ such that

$$E^m(X \mid B) = \min_{p \in \Delta^B(\pi)} \int X dp$$

A simple evaluation, E, together with the prior-by-prior updating rule yields the following compound evaluation

$$E^m_{\mathcal{P}}(X) = E(E^m(X \mid \mathcal{P}))$$

where $E^m(X \mid \mathcal{P})$ is the random variable that yields $E^m(X \mid B)$ at any state $s \in B$.

Dempster-Shafer theory provides an alternative updating rule for totally monotone capacities. The DS-updating formula is as follows: for any E with capacity π and E-nonnull B, define the conditional capacity $\pi^{ds}(\cdot \mid B)$ as follows:

$$\pi^{ds}(A \mid B) = \frac{\pi(A \cup B^c) - \pi(B^c)}{1 - \pi(B^c)}$$

Then, $E^{ds}(X \mid B) = \int X d\pi^{ds}(\cdot \mid B)$ is the conditional evaluation for the Dempster-Shafer rule. The corresponding compound evaluation is

$$E^{ds}_{\mathcal{P}}(X) = E(E^{ds}(X \,|\mathcal{P}))$$

The following examples contrast the proxy rule with prior-by-prior updating and the Dempster-Shafer rule.

3.1 Example 1: Proxy rule vs Prior by Prior updates

In this example, two balls are drawn consecutively from separate urns. The first draw is unambiguous while the second draw is not. The agent bets on the first draw after observing the second. The first ball is drawn from an urn consisting of one ball labeled R and a second ball labeled G. If R is drawn, the second ball is drawn from urn I; if G is drawn, the second draw is from urn II. Both urn I and urn II contain 12 balls, each either red or green. Urn I contains at least 4 red and at least 2 green balls while urn II contains least 4 green and at least 2 red balls.

The decision maker observes the second draw (b or w) and, conditional on that draw, evaluates the random variable X that yields 1 in case the first draw was R, and 0 otherwise. Let $S = \{Rr, Wb, Gr, Gg\}$ denote the state space, let $R := \{Rr, Wb\}, B = \{Gr, Gg\}$ denote the events corresponding to the first draw and let $w = \{Rr, Gr\}, b = \{Wb, Gg\}$ denote the events corresponding to the second draw. Assume that the agent translates the description above into the totally monotone capacity π such that

$$\pi(Rr) = \pi(Gg) = 1/6$$

$$\pi(Rg) = \pi(Gr) = 1/12$$

$$\pi(R) = \pi(B) = 1/2$$

The capacity of all other events is found by extending the above values additively. The Möbius transform of π is μ_π where

$$\mu_\pi(Rr) = \mu_\pi(Gg) = 1/6$$

$$\mu_\pi(Rg) = \mu_\pi(Gr) = 1/12$$

$$\mu_\pi(R) = \mu_\pi(B) = 1/4$$

For all other events $A \subset S$, $\mu_\pi(A) = 0$. For $\mathcal{P} = \{w, b\}$ the Möbius transform of the proxy satisfies

$$\mu_\pi^{\mathcal{P}}(Rr) = \frac{1}{6} + \frac{1}{2} \cdot \frac{1}{4} = \frac{7}{24} = \mu_\pi^{\mathcal{P}}(Gg)$$
$$\mu_\pi^{\mathcal{P}}(Rg) = \frac{1}{6} + \frac{1}{2} \cdot \frac{1}{4} = \frac{5}{24} = \mu_\pi^{\mathcal{P}}(Gr)$$

and $\mu_\pi^{\mathcal{P}}(A) = 0$ for all other $A \subset S$. Therefore, the proxy updates $\pi^{\mathcal{P}}(R|r), \pi^{\mathcal{P}}(G|r)$ are

$$\pi^{\mathcal{P}}(R|r) = \frac{7/24}{12/24} = 7/12$$
$$\pi^{\mathcal{P}}(R|g) = \frac{5/24}{12/24} = 5/12$$

Thus, under the proxy rule $E(X|r) = 7/12$ and $E(X|g) = 5/12$.

Next, we compute the conditional evaluations with the prior-by-prior rule. The core of the capacity π is the following set of priors:

$$\Delta(\pi) = \{p \,|\, p_{Rr} = x, p_{Rg} = 1/2 - x, p_{Gr} = 1/2 - y, p_{Gg} = y \text{ for } x, y \in [1/6, 5/12]\}$$

A straightforward calculation then yields the following set of posteriors:

$$\Pr(R|r) \in [1/3, 5/6]$$
$$\Pr(R|g) \in [1/6, 2/3]$$

With the maxmin criterion applied to the set of updated beliefs, the bet on R is worth 1/3 after observing a red ball and 1/6 after observing a green ball, both less the ex ante value of 1/2. Thus, if a maxmin agent updates prior-by-prior, then *all news can be bad news.* Following a green draw, the agent evaluates X *as if* urn I has 4 red balls and 8 green balls and urn II has of 8 red balls and 4 green balls. Following a green draw, the agent evaluates X as if urn I had of 10 red balls and 2 green balls while urn II had of 10 green balls and 2 red balls.[5] By contrast, under the proxy rule, for the purpose of updating the agent creates "proxy urns" I* and II* that do not vary with the conditioning event. In this example, the proxy urn I* has 7 green balls and 5 red balls while II* has 7 green balls and 5 red balls.

[5] In this example, the set of updated beliefs conditional on any cell in the information partition is a superset of the set of priors. Seidenfeld and Wassermann (1993) observe this feature of prior-by-prior updating and call it *dilation*. Their paper characterizes when it occurs.

3.2 Example 2: Proxy Updating vs the Dempster-Shafer rule

Consider the following modified version of the previous example. As above, the first ball is drawn from an urn that has two balls, one labeled k, the other labeled u. If the ball labeled k is chosen, a second ball is drawn from urn K; if the balllabeled u is chosen, the second ball is drawn from urn U. Urn U contains 4 red or green balls; at least 1 red, at least 1 green. Urn K contains 2 red balls and 2 green balls. The decision maker observes the draw from the second urn (red or green) and, conditional on that draw, evaluates the random variable X that yields 1 in case the first draw was the ball labeled k, and 0 otherwise.

Let $S = \{kg, kr, ug, ur\}$ and let $k = \{kg, kr\}, u = \{ug, ur\}, g = \{kg, ug\}, r = \{kr, ur\}$ be the events corresponding to the first and second draws respectively. Assume the agent translates the description above into the totally monotone capacity π where

$$\pi(ug) = \pi(ur) = 1/8$$

$$\pi(kg) = \pi(kr) = 1/4$$

$$\pi(k) = \pi(u) = 1/2$$

The capacity of all other events is determined by extending the values above additively. According to the Dempster-Shafer rule, the conditional capacity satisfies, $\pi^{ds}(k|r) = 2/5 = \pi^{ds}(k|g)$. Thus, all news about the random variable X is bad news. The Dempster-Shafer rule determines the posterior likelihood of k given r as if the agent picks a "maximum likelihood" urn consistent with the description of urn U. Thus, upon observing a red ball, the agent updates as if urn U consisted of 3 red balls and 1 green ball, while upon observing a green ball, the agent updates as if urn U consisted of 3 green balls and 1 red ball.[6] By contrast, proxy updating implies that observing a red or green draw provides no information about X in this case. Under the proxy rule, for the purpose of updating the agent creates a "proxy urn" U^* that does not vary with the conditioning event and contains 2 red and 2 green balls.

[6] In this example, prior-by-prior updating coincides with Dempster-Shafer updating. Both rules imply that all news is bad news.

4. Representation Theorem

To derive the proxy rule, we impose 4 conditions on the conditional evaluation. The first condition relates to elementary evaluations: recall that E_D is the (elementary) evaluation with capacity π_D such that $\pi_D(A) = 1$ if $D \subset A$ and $\pi_D(A) = 0$ otherwise. Our first property specifies the conditional elementary evaluations:

C1: *If B is E-nonnull, then $E(\cdot \mid B) = E(\cdot \mid B \cap S_E)$; in particular, $E_D(\cdot \mid B) = E_{D \cap B}$.*

The first part asserts that only the part of the conditioning event contained in the support of E matters. The second part states that the conditioning an elementary evaluation with support D yields an elementary evaluation with support D intersected with the conditioning event. Property C1 is uncontroversial: it is satisfied by all updating rules including Bayesian updating of probabilities, prior-by-prior updating and Dempster-Shafer updating.

Our second property, C2, asserts that if E_0 is a mixture of E_1 and E_2, then the conditional of E_0 must also be a mixture of the conditionals of E_1 and E_2. In addition, C2 ensures that the support of the conditional of a mixture of two evaluations is the same as the union of the supports of the conditionals of the two individual evaluations.

It is difficult to interpret C2 as a behavioral condition because it relates the conditionals of two evaluations to the conditionals of a third one. Nevertheless, it is a simple and easily interpretable condition that is satisfied by Bayes' law (i.e., when calculating conditional expectations), by prior-by-prior updating, and by Dempster-Shafer updating.

C2: *Let $E_0 = \lambda E_1 + (1-\lambda)E_2$ for $\lambda \in (0,1)$ and let B be E_1-nonnull. Then, $E_0(\cdot|B) = \alpha E_1(\cdot|B) + (1-\alpha)E_2(\cdot|B)$ for some $\alpha \in (0,1]$; $\alpha < 1$ if and only if B is E_2-nonnull.*

The third property, symmetry, says that the updating rule is neutral with respect to labeling of the states. Let $h : S \to S$ be a bijection. We write $X \circ h$ for the random variable Y such that $Y(s) = X(h(s))$ and $h(A)$ for the event $A' = \{h(s) \mid s \in A\}$. C3 below, like C1 and C2 is satisfied by all known updating rules. Nevertheless, providing a version of our updating rule that does not impose C3 is not difficult. We discuss such updating rules following Corollary 1 below.

C3: *If $E(X \circ h) = E(X)$ for all X, then $E(X \circ h \,|\, h(B)) = E(X \,|\, B)$.*

Our final property is weak dynamic consistency. This property is what enables us to distinguish our model from prior-by-prior updating and Dempster-Shafer updating since the latter two models do not satisfy it. Weak dynamic consistency reflects the view that gradual resolution of uncertainty reduces ambiguity and, therefore, cannot render the compound evaluation uniformly worse. To put it differently, not all news can be bad news.

C4: *If $E(X|B) \leq c$ for all E-nonnull $B \in \mathcal{P}$, then $E(X) \leq c$.*

Theorem 1: *A conditional evaluation satisfies C1-C4 if and only if it is the proxy update.*

We defined proxy updating as an application of Bayes' rule to a modified capacity (the proxy capacity). Next, we provide two other equivalent definitions, the first in terms of Shapley values of appropriately defined games, the second in terms of the core of the capacity.

Given the capacity π, interpret $\pi(A)$ as the value of "coalition" $A \subset S$. Then, let $\rho_\pi(s)$ denote the Shapley value of "player" s in the "game" π.[7] Without risk of confusion, we identify ρ_π with its additive extension to the power set of S; that is $\rho_\pi(A) = 0$ whenever $A \cap S_\pi = \emptyset$, $\rho_\pi(A) = \sum_{s \in A \cap S_\pi} \rho(s)$ otherwise. Let $\rho_\pi^D(s)$ denote the Shapley value of s in the game π^D where $\pi^D(A) = \pi(A \cap D)$ for all A. Again, identify ρ_π^D with its additive extension to the set of all subsets of S. Corollary 1 shows that the updated capacity is the ratio of Shapley values:

Corollary 1: *For any nonnull $B \in \mathcal{P}$ and $D := A \cup B^c$, $\pi(A \,|\, B) = \frac{\rho_\pi^D(A \cap B)}{\rho_\pi^D(B)}$.*

Without property C3 the updating rule is not unique. Specifically, replacing the Shapley value with a *weighted* Shapley value (see Kalai and Samet (1987)) in Corollary 1

[7] Let Θ denote the set of all permutations of S_π. Hence, Θ is the set of all bijections from S_π to the set $\{1, \ldots, |S_\pi|\}$. Then, for any $\theta \in \Theta$, let $\theta^s = \{\hat{s} \in S_\pi \,|\, \theta(\hat{s}) \leq \theta(s)\}$. The Shapley value of s is defined as follows:

$$\rho_\pi(s) = \frac{1}{|\Theta|} \sum_{\theta \in \Theta} [\pi(\theta^s) - \pi(\theta^s \backslash \{s\})]$$

yields an updating rule that satisfies C1, C2 and C4. Thus, the role of C3 is to yield a uniform weight on all states.

Next, we relate proxy updates to the prior-by-prior rule. Recall that $\Delta(\pi)$ denotes the core of the capacity π. Under the prior-by-prior rule, the agent updates every prior in $\Delta(\pi)$ and evaluates conditional acts with the least favorable posterior. The proxy rule can be interpreted as a modified version of the prior-by-prior rule in which, for updating purposes, the agent considers only a subset of the priors in $\Delta(\pi)$. This subset depends on the information partition. Recall that ρ_π is the Shapley value of π and let

$$\Delta^{\mathcal{P}}(\pi) = \{p \in \Delta(\pi) \,|\, p(B) = \rho_\pi(B) \text{ for all } B \in \mathcal{P}\}$$

The set $\Delta^{\mathcal{P}}(\pi)$ is the subset of $\Delta(\pi)$ consisting of priors that assign $\rho_\pi(B)$ to each cell B of the information partition; $\Delta^{\mathcal{P}}(\pi)$ is also the core of the proxy $\pi^{\mathcal{P}}$.

As we noted above, our model draws a distinction between the agent's ex ante and ex post perceptions of ambiguity. Once an information cell, B, is revealed, it is no longer deemed ambiguous and the agent assigns probability $\rho_\pi(B)$ to it (and assigns probability $1 - \rho^\pi(B)$ to B^c). When updating, she only considers priors that agree with the Shapley value on elements of the information partition. Corollary 2 shows that proxy updating is equivalent to this modified prior-by-prior rule:

Corollary 2: *If $E(\cdot|B)$ is the proxy update of the simple evaluation E then*

$$E(X \,|\, B) = \min_{p \in \Delta^{\mathcal{P}}(\pi)} \frac{1}{p(B)} \int X_B dp$$

for all random variables X and nonnull $B \in \mathcal{P}$.

As we noted earlier, proxy updating violates the law of iterated expectation. However, there are information structures for which the law of iterated expectation is satisfied. Epstein and Schneider (2003) identify a necessary and sufficient condition (rectangularity) on $\pi, \mathcal{P}$ that ensures that the law of iterated expectation is satisfied with prior-by-prior updating. The proposition below shows that the same condition, applied to our setting, is also necessary and sufficient for satisfying the law of iterated expectation with updating by proxy and with the Dempster-Shafer rule. We replace Epstein and Schneider's rectangularity condition with a new condition that is easier to interpret in our setting. We say that the partition $\mathcal{P} = \{B_1, \ldots, B_n\}$ is π-extreme if, for all i, either (i) B_i is unambiguous or (ii) $\mu_\pi(A) \cdot \mu_\pi(C) > 0$ and $A \cap B_i \neq \emptyset \neq C \cap B_i$ imply $A \cap B_i = C \cap B_i$. Thus, a partition is π-extreme if each of its elements, B_i, is either unambiguous or totally ambiguous in the sense that every positive probability element of the Möbius transform of π that intersects B_i contains all nonnull states of B_i. Proposition 3, below, shows that $\pi-$extremeness is necessary and sufficient for the law of iterated expectation under any of the three updating rules.

Proposition 3: *Let π be the capacity of the simple evaluation E and let $\mathcal{P}$ be a partition. Then, the following statements are equivalent: (i) $E_{\mathcal{P}}(X) = E(X)$ for all X, (ii) $E^m_{\mathcal{P}}(X) = E(X)$ for all X, (iii) $E^{ds}_{\mathcal{P}}(X) = E(X)$ for all X, (iv) $\mathcal{P}$ is π-extreme.*

As Proposition 3 shows, when conditioning on ambiguous information, only the most trivial form of ambiguity is consistent with the law of iterated expectation. In particular, there must either be no ambiguity or complete ambiguity within any cell of the information partition. Moreover, all three rules are equivalent when it comes to their adherence to the law of iterated expectation.

Our setting is less general than the setting of Epstein and Schneider (2003) because we restrict the sets of probabilities to be the core of a totally monotone capacity. Thus, the recursive prior-by-prior model and Epstein and Schneider's rectangularity condition apply to a broader class of models than the other two rules and π-extremeness.

5. Proxy Updating and Inference: An Example

In this section, we apply proxy updating to a standard inference problem. Let $\{0, 1\}^n$ be sequences of n possible signal realizations and let $\Theta = \{\theta_1, \theta_2\}$, $0 < \theta_1 < 1/2 < \theta_2 < 1$ be the possible values of the parameter. The state space is $S = \Theta \times \{0, 1\}^\infty$.

The decision maker observes a sequence of n signals $y = (y_1, \ldots, y_n) \in T^n$ and draws inferences about the parameter $\theta \in \Theta$. The signal induces a partition $\mathcal{P} = \{\Theta \times \{y\} \,|\, y \in T^n\}$. Let $\#y = \sum_{i=1}^n y_i$.

The decision maker's simple evaluation with n signals has capacity π_n. For this capacity, the parameters are unambiguous: $\pi_n(\{\theta_i\} \times T) = p_i$ and $p_1 + p_2 = 1$; if $A \subset \{\theta_1\} \times T$, $B \subset \{\theta_2\} \times T$, then $\pi_n(A \cup B) = \pi_n(A) + \pi_n(B)$. However, there is ambiguity about the signal: if $A \neq \emptyset$ is a strict subset of $\{\theta_1\} \times T$, then

$$\pi_n(A) = (1-\epsilon)p_i \sum_{(\theta_i, y) \in A} \theta_i^{\#y}(1-\theta_i)^{n-\#y}$$

If $\epsilon = 0$, then this is the setting of a standard Wald (1945) problem in which the y_i's follow a Bernoulli distribution with parameter θ_i. If $\epsilon > 0$, then the signal is ambiguous. This example describes an agent who believes that with probability $1-\epsilon$ a typical Bernoulli source generates the signals and with probability ϵ the signal is maximally ambiguous.

The decision maker's goal is to evaluate the random variable that yields 1 if the value of the parameter is θ_1 and 0 otherwise. With some abuse of notation, we write θ_i to denote the event $\{\theta_i\} \times T$. Hence, 1_{θ_1} is the uncertain prospect that the decision maker needs to evaluate. Let $E(1_{\theta_1} \,|\, y)$ be the conditional evaluation of this random variable given $\Theta \times \{y\}$ for $y \in \{0,1\}^n$. It is straightforward to verify that $E(1_{\theta_1} \,|\, y) = E(1_{\theta_1} \,|\, y')$ when $\#y = \#y'$. Consider a sequence of the above inference problems indexed by the number of signals n. Let $r_i := \frac{\ln(1/2) - \ln\theta_i}{\ln(1-\theta_i) - \ln\theta_i}$. The following proposition characterizes the limit of this evaluation as n goes to infinity.

Proposition 4: *Assume $0 < \epsilon < 1$ and $\alpha := \lim_{n\to\infty} \#y/n$. Then,*

$$\lim E(1_{\theta_1} \,|\, y) = \begin{cases} 0 & \textit{if } \alpha_1 < r_1 \\ p_1 & \textit{if } r_1 < \alpha < r_2 \\ 1 & \textit{otherwise.} \end{cases}$$

To illustrate the result in Proposition 4, assume that $\theta_1 = 1/3, \theta_2 = 2/3$. If $\epsilon = 0$, then, $\lim E(1_{\theta_1} \,|\, y) = 0$ for $\alpha < 1/2$ and $\lim E(X \,|\, y) = 1$ for $\alpha > 1/2$. By contrast, if $\epsilon > 0$, then

$$\lim E(1_{\theta_1} \,|\, y) = \begin{cases} 0 & \text{if } \alpha < .42 \\ p_1 & \text{if } .42 < \alpha < .58 \\ 1 & \text{otherwise.} \end{cases}$$

Thus, the effect of ambiguity is to create a region (between .42 and .58 in the example) where the signal is deemed uninformative.

We can compare Proposition 4 to the corresponding results for prior-by-prior updating and the Dempster-Shafer rule. Let $E^m(1_{\theta_1} \,|\, y)$ and $E^{ds}(1_{\theta_1} \,|\, y)$, respectively be the conditional evaluations of the decision maker's uncertain prospect with prior-by-prior and Dempster-Shafer updating. As is the case with the proxy rule, it is straightforward to verify that $E^m(1_{\theta_1} \,|\, y)$ and $E^{ds}(1_{\theta_1} \,|\, y)$ depend on y only through $\#y$.

Proposition 5: *If* $\lim \#y/n$ *exists, then* $\lim E^m(1_{\theta_1} \,|\, y) = 0$ *and* $\lim E^{ds}(1_{\theta_1} \,|\, y) = p_1$.

For large n, the Dempster-Shafer rule predicts that the agent deems the signal uninformative irrespective of its realization. To understand the reason for this result, first note that the with Dempster-Shafer rule, the conditional value of 1_{θ_1} depends on the "maximum likelihood" prior among those in the core of π_n. For large n, the maximum likelihood prior is the one that places an ϵ mass on the realized sequence conditional on either value of the parameter. It is easy to see that - for that prior - the signal is uninformative when n is large.

The prior-by-prior rule chooses the worst prior among all of the updated priors in the core of π_n. For one such prior, conditional on θ_2, the realized signal sequence has probability ϵ while, conditional on θ_1, the realized sequence has the corresponding Bernoulli probability times $1 - \epsilon$. It is straightforward to verify that - for that prior - the updated value of 1_{θ_1} converges to zero.

In the example above, there is no prior ambiguity regarding the parameter θ; that is, $\pi_n(A \cup B) = \pi_n(A) + \pi_n(B)$ for $A \subset \{\theta_1\} \times T, B \subset \{\theta_2\} \times T$. In such situations, the proxy rule yields no ambiguity regarding θ conditional on observing the signal sequence. By contrast, with prior-by-prior updating, the posterior over θ "inherits" ambiguity from the ambiguous signal. In the limit, the posterior ambiguity is extreme as the set of posterior probabilities of θ_1 converges to the whole interval $[0, 1]$ for all signal sequences. It is straightforward to generalize the example above so that the parameter is ambiguous. In that case, the conditional (proxy-)evaluation about the parameter will also exhibit ambiguity.

6. Revealed Preference Implications

Our agent's evaluation of a compound random variable depends on how uncertainty resolves. In this section, we analyze the dynamic behavior consistent with our model. The agent evaluates random variables according to the function $E_{\mathcal{P}}$ such that

$$E_{\mathcal{P}}(X) = E(E(X \,|\, \mathcal{P}))$$

where E is a simple evaluation and $E(\cdot \,|\, \mathcal{P})$ is the conditional valuation defined in section 3. The goal of this section is to relate $E_{\mathcal{P}}$ to standard models of choice under ambiguity.

Example: Let $S = \{1, 2, 3\}$ and let state 1 be unambiguous: $(\pi(\{1\}) = 1/3 = 1 - \pi(\{2, 3\}))$. States 2 and 3 are ambiguous and $\pi(\{2\}) = \pi(\{3\}) = 0$. The information partition is $\mathcal{P} = \{\{1, 2\}, \{3\}\}$. The compound evaluation $E_{\mathcal{P}} = E(E(\cdot \,|\, \mathcal{P}))$ for this example has the following maxmin representation: let $\Delta = \{p \,|\, p_1 = p_2, 0 \leq p_3 \leq 2/3\}$. Then, it is easy to verify that the compound evaluation is the maxmin evaluation with the set of priors Δ; that is,

$$E_{\mathcal{P}}(X) = \min_{p \in \Delta} \int X dp$$

Note that the right-hand side of the above equation does not have a Choquet integral representation. Thus, compound evaluations encompass behavior more general than simple evaluations. As we show below, the features of this example generalize: compound evaluations always have a maxmin representation and, conversely, for every maxmin evaluation there is a compound evaluation that approximates it.

So far, we have only considered signals that form a partition of the payoff relevant states. This information structure is rich enough for our results on updating but it is too sparse to capture the range of possible compound evaluations. Therefore, we extend information structures to include signals that do not correspond to a partition of the payoff relevant states.

Let the set T represent the possible signals. Let E be a simple evaluation on S, the payoff relevant states, and let π be its capacity. Then, the simple evaluation $(S \times T, E^e)$ is an *extension of* (S, E) if its capacity, π^e, satisfies $\pi^e(A \times T) = \pi(A)$ for all $A \subset S$. For any random variable X on S, define the extended random variable X^e on $S \times T$ as follows:

$X^e(s,t) = X(s)$ for all $(s,t) \in S \times T$. Let $\mathcal{T} = \{S \times \{t\} \,|\, t \in T\}$ be the partition of $S \times T$ induced by the signal T.

Then, the general compound evaluation $E_{\mathcal{T}}$ is defined as follows:

$$E_{\mathcal{T}}(X) = E^e(E^e(X^e \,|\, \mathcal{T}))$$

where $E^e(X^e \,|\, \mathcal{T})$ is the proxy update of the simple evaluation E^e defined above. Hence, a general compound evaluation assigns to every random variable, X, the compound evaluation of its extension, X^e, to some $S \times T$ where T is the set of possible realizations of some signal.

Theorem 2 characterizes all generalized compound evaluations:

Theorem 2: *Let $(S \times T, E^e)$ be an extension of (S, E). Then, there is a compact, convex set of simple probabilities Δ, each with support contained in S_π, such that*

$$E_{\mathcal{T}}(X) = \min_{p \in \Delta} \int X dp$$

Theorem 2 shows that any generalized compound evaluation can be represented as a maxmin evaluation on the payoff relevant states. Theorem 3 provides a converse: any maxmin evaluation can be approximated by a generalized compound evaluation on the payoff relevant states. For any set of states $\hat{S}$, let $\Delta_{\hat{S}}$ denote the set of probabilities with support contained in $\hat{S}$ and let $\mathcal{V} = \{X \,|\, X(s) \leq 1\}$.[8]

Theorem 3: *For any nonempty finite $\hat{S}$, compact, convex set of probabilities $\Delta \subset \Delta_{\hat{S}}$ and $\epsilon > 0$, there is a general compound evaluation $E_{\mathcal{T}}$ such that, for all $X \in \mathcal{V}$,*

$$\left| E_{\mathcal{T}}(X) - \min_{p \in \Delta} \int X dp \right| < \epsilon$$

[8] The approximation result holds for any uniformly bounded set of random variables. Setting the bound to 1, as we have done here, is without loss of generality.

Theorems 2 and 3 provide a tight connection between our theory of updating and maxmin evaluations. Note that *simple* evaluations constitute a subset of maxmin evaluations - those for which the set of probabilities forms the core of a totally monotone capacity. In contrast, compound evaluations are more general; they can approximate *all* maxmin evaluations.

7. Conclusion

We have developed and analyzed a consequentialist and recursive model of updating under ambiguity. In our model, the agent is dynamically consistent but is not indifferent to the order of resolution of uncertainty. We derived an updating rule that yields behavior consistent with some of the implicit assumptions in the ambiguity literature: objective uncertainty (coin toss) hedges ambiguity when it resolves after the ambiguous event (ball drawn from an urn with unknown proportions of red and green balls) but provides no hedge if it resolves before the ambiguous event.

We have considered random variables that resolve in at most two periods; we have not dealt with arbitrary filtrations. Since conditional evaluations, $E(\cdot \mid B)$, are themselves simple evaluations, this is essentially without loss of generality. The extension of our results to more general dynamic settings is immediate.

For the most part, we have interpreted the order of resolution of uncertainty as an observable. We can also interpret the information structure as *subjective.* In other words, the information structure may reflect the sequence in which the agent envisages the uncertainty resolving. With this interpretation, two agents facing the same random variable may perceive different information structures. An agent who believes randomization hedges ambiguity behaves *as if* the ball is drawn first while an agent who does not believe randomization hedges ambiguity behaves as if the coin is tossed first. Intermediate situations between these two extremes can be modeled with the general information structures and Theorems 2 and 3 characterize the observable implications of that model.

In future research, we plan to apply our updating rule to standard dynamic problems in statistical decision theory, suitably modified to incorporate ambiguity. These applications are likely to require a generalization of our updating rule to settings with a continuum of states and continuous random variables.

8. Appendix

8.1 Proof of Proposition 1

Verifying that E defined by $E(X) = \int X d\pi$ for some capacity π satisfies P1 and P2 is straightforward. Next, we will show that if this capacity is totally monotone, then P3 is also satisfied.

Since π, the capacity of E, is totally monotone, it has a non-negative Möbius transform μ_π. Note that

$$E(Z) = \sum_{B \neq \emptyset} \mu_\pi(B) \cdot \min_{s \in B} Z(s)$$

for all Z. The display equation above is easy to verify using the definition of the Choquet integral and the definition of the Möbius transform; it is also well-known (see for example Gilboa (1994)). Let $\mathcal{X} = (X_1, \ldots, X_n)$ and $\mathcal{Y} = (Y_1, \ldots, Y_n)$, let $\alpha_1 > \ldots > \alpha_k$ be all of the strictly positive values attained by any random variable in $\mathcal{X}$ or $\mathcal{Y}$ and let $\alpha_{k+1} = 0$, Then, the display equation above yields

$$\begin{aligned}
\sum E(X_i) &= \sum_i \sum_{A \neq \emptyset} \mu_\pi(A) \cdot \min_{s \in A} X_i(s) = \sum_{A \neq \emptyset} \mu_\pi(A) \sum_i \cdot \min_{s \in A} X_i(s) \\
&= \sum_{A \neq \emptyset} \mu_\pi(A) \sum_{j=1}^{k} K_{\alpha_j}(\mathcal{X}, A) \cdot (\alpha_j - \alpha_{j+1}) \\
&\leq \sum_{A \neq \emptyset} \mu_\pi(A) \sum_{j=1}^{k} K_{\alpha_j}(\mathcal{Y}, A) \cdot (\alpha_j - \alpha_{j+1}) = \sum E(Y_i)
\end{aligned}$$

as desired.

Next, we will prove that P1-P3 yield the Choquet integral representation and that the corresponding capacity is totally monotone. First, note that E must be monotone; that is, $X(s) \geq Y(s)$ for all $s \in S$ implies $E(X) \geq E(Y)$. This follows immediately from P3. Define π such that $\pi(A) = E(1_A)$. Next, we will show that $E(\gamma_A) = \gamma\pi(A)$. First, by P2, $E(0) = E(0) + E(0)$ and hence $E(0) = 0$ and therefore, the desired result holds if $\gamma = 0$. Next, assume that $\gamma > 0$ is a rational number. Then $\gamma = \frac{k}{n}$ for integers $k, n > 0$. Arguing as above, by invoking P2 repeatedly, we get $E(\gamma_A) = kE(\frac{1}{n} \cdot 1_A) = \frac{k}{n}E(1_A) = \gamma E(1_A)$. Suppose $E(\gamma_A) > \gamma\pi(A)$ for some irrational γ. If $\pi(A) = 0$, choose a rational $\delta > \gamma$

and invoke the monotonicity and the result established for the rational case to get $0 = E(\delta_A) \geq E(\gamma_A) > 0$, a contradiction. Otherwise, choose $\delta \in (\gamma, E(\gamma_A)/\pi(A))$ and again, invoke the previous argument and monotonicity to get $E(\gamma_A) > \delta\pi(A) = E(\delta_A) \geq E(\gamma_A)$, a contradiction. A symmetric argument for the $E(\gamma_A) < \gamma$ case yields another contradiction. Hence, $E(\gamma_A) = \gamma\mu(A)$.

Then, by applying the fact established in the previous paragraph and P2 repeatedly, we get

$$E(X) = \sum_{i=1}^{n} (\alpha_i - \alpha_{i+1})\pi(\{s \mid X(s) \geq \alpha_i\})$$

where $\alpha_i, \ldots, \alpha_n$ are all of the nonzero values that X can take on S_π and $\alpha_{n+1} = 0$.

To conclude the proof, we will show that π is totally monotone. Let n be the cardinality of a non-empty subset of the support of π. Without loss of generality, we identify $\{1, \ldots, n\}$ with this set and let N be the set of all subsets of $\{1, \ldots, n\}$. Let N^o (N^e) be the set of all subsets of $\{1, \ldots, n\}$ that have an odd (even) number of elements.

First, consider the case in which n is an even number. Let $\mathcal{X} = (1_B)_{B \in N^e}$ and $\mathcal{Y} = (1_B)_{B \in N^o}$. It is easy to verify that the cardinality of the sets $\mathcal{X}$ and $\mathcal{Y}$ are the same: 2^{n-1}. We will show that $K_\alpha(\mathcal{X}, A) \geq K_\alpha(\mathcal{Y}, A)$. Choose $A \subset N$ such that $k = |A| < n$. Then, it is easy to verify that for $\alpha \in (0, 1]$,

$$|K_\alpha(\mathcal{X}, A) - K_\alpha(\mathcal{Y}, A)| = \left| \sum_{m=k}^{n} (-1)^m \binom{n-k}{m-k} \right| = (1-1)^{n-k} = 0$$

For $|A| = n$ and $\alpha \in (0, 1]$, $K_\alpha(\mathcal{X}, A) - K_\alpha(\mathcal{Y}, A) = 1 - 0 = 1$. That $K_\alpha(\mathcal{X}, A) - K_\alpha(\mathcal{Y}, A) = 0$ in all other cases is obvious. By P3, $\sum_{B \in N^e} E(1_B) - \sum_{B \in N^o} E(1_B) \geq 0$. Recall that $\sum E(X) = \sum_{A \neq \emptyset} \mu_\pi(A) \cdot \min_{s \in A} X(s)$ for all X. Hence,

$$\begin{aligned} 0 \leq \sum_{B \in N^e} E(1_B) - \sum_{B \in N^o} E(1_B) &= \sum_{B \in N^e} \sum_{A \subset B} \mu_\pi(A) - \sum_{B \in N^o} \sum_{A \subset B} \mu_\pi(A) \\ &= \sum_{B \in N^e} 2^{n-|B|-1} - \sum_{B \in N^o} \sum_{A \subset B} \mu_\pi(A) \\ &= \mu_\pi(\{1, \ldots, n\}) \end{aligned}$$

Next, consider the case in which n is an odd number. Let $\mathcal{X} = (1_B)_{B \in N^o}$, $\mathcal{Y} = (1_B)_{B \in N^e}$ and repeat the arguments above to establish $\mu_\pi(\{1, \ldots, n\}) \geq 0$ for all odd n. $\square$

8.2 Proof of Proposition 2

Define $\mathcal{A}_\pi = \{A \subset S \,|\, \mu_\pi(A) \neq 0\}$. To prove (iii) implies (ii), let $s \in B$ be such that $A_\pi(s) \cap B^c \neq \emptyset$. Then, there exists $A \in \mathcal{A}_\pi$ such that $A \cap B \neq \emptyset \neq A \cap B^c$. It follows that $\pi(B) + \pi(B^c) \leq 1 - \mu_\pi(A) < 1$ as desired. Next, assume (ii) holds. Recall that $\int X d\pi = \sum_{C \in \mathcal{A}_\pi} \min_{s \in C} X(s)\mu_\pi(C)$. Hence, $E(X) = E(X_B) + E(X_{B^c})$ for all X, proving (i). Finally, if (i) holds, then $1 = E(1_B) + E(1_{B^c}) = \pi(B) + \pi(B^c)$ and hence (iii) follows. □

8.3 Proof of Theorem 1

In the following lemmas, we assume that $E(\cdot \,|\, \cdot)$ satisfies C1-C4. Let π^B be the capacity of $E(\cdot \,|\, B)$. To prove the only if part of Theorem 1, we will show that $\pi^B = \pi(\cdot|B)$ where $\pi(\cdot \,|\, B)$ is as defined by (2) and (3) above.

Lemma 1: *Let $E \in \mathcal{E}^o$ and π be the probability of E. Then $\pi^B(A) = \pi(A \cap B)/\pi(B)$ for all A.*

Proof: Let δ_s be the probability such that $\delta_s(\{s\}) = 1$. By the hypothesis of the lemma, $\pi = \sum_{s \in D} \alpha_s \delta_s$ for some finite set D and some $\alpha_s \in (0, 1]$ such that $\sum_D \alpha_s = 1$. If D is a singleton, then the result follows from property C1. Thus, assume the result holds for all D' with cardinality $k \geq 1$ and let D have cardinality $k + 1$. If $B \cap D = D$, then write E as a convex combination of some E' with capacity $\pi' = \sum_{i=1}^k \alpha'_s \delta_s$ and $\hat{E}$ with capacity $\delta_{\hat{s}}$. Then, property C2 and the inductive hypothesis imply that $\pi^B = \sum_{s \in D} \gamma_s \delta_s$ for some $\gamma_s \in (0, 1]$ such that $\sum_{s \in D} \gamma_s = 1$. By C4, $\gamma_s \geq \alpha_s$ for all s and, therefore, $\gamma_s = \alpha_s$ as desired. If $\emptyset \neq B \cap D \neq D$, then again write E as a convex combination of some E' with capacity $\pi' = \sum_{s \in B \cap D}^k \alpha'_s \delta_s$ and $\hat{E}$ with capacity $\hat{\pi} = \sum_{s \in D \setminus B} \alpha'_s \delta_s$. Then, the result follows from property C2 and the inductive hypothesis. □

For Lemma 2, we consider evaluations E, E_1 and E_D such that $E = \frac{1}{2}E_D + \frac{1}{2}E_1$. Let π, π_1 and π_D respectively, be the capacities of E, E_1 and E_D respectively. Clearly, $\pi = \frac{1}{2}\pi_D + \frac{1}{2}\pi_1$. Finally, for any E-nonnull B, let π^B denote the capacity of $E(\cdot \,|\, B)$.

Lemma 2: *Let $C = \{s_1, \ldots, s_k\}$, let $\pi_1 = \frac{1}{k}\sum_{s \in C} \delta_s$, $C \cap D = \emptyset$ and $|D| = k$. Let $\mathcal{P} = \{B_1, \ldots, B_k\}$ be such that $B_i \cap C = \{s_i\}$ for all $i \geq 1$. Then,*

$$\pi^{B_i} = \frac{|B_i \cap D|\pi_{B_i \cap D} + \delta_{s_i}}{|B_i \cap D| + 1}$$

Proof: Lemma 1, C1 and C2 imply that there is $a_i \geq 0$ such that

$$\pi^{B_i} = \frac{a_i}{1+a_i}\pi_{B\cap D} + \frac{1}{1+a_i}\delta_{s_i} \tag{$A1$}$$

and $a_i > 0$ if and only if $B_i \cap D \neq \emptyset$. Next, we show that $\sum_{i=1}^k a_i = k$. First, assume $\sum_{i=1}^k a_i > k$. Let X be the following random variable:

$$X(s) = \begin{cases} 1+a_i & \text{if } s = s_i \in C \\ 0 & \text{otherwise} \end{cases}$$

Then, equation (A1) above implies $E(X|B_i) = 1$ for all i. Also, $E(X) = \sum_{i=1}^k \frac{1+a_i}{2}\frac{1}{k} = \frac{1}{2} + \frac{1}{2k}\sum_{s=1}^k a_i > 1$. Since this violates condition C4, we conclude that $\sum_{i=1}^k a_i \leq k$. Next, assume $\sum_{i=1}^k a_i < k$. Choose $r > \max\{1+a_i \,|\, i \in C\}$. Let Y be the following random variable:

$$Y(s) = \begin{cases} r-1-a_i & \text{if } s = s_i \in C \\ r & \text{otherwise} \end{cases}$$

Then, equation (A1) above implies $E(Y|B_i) = r-1$ for all i. Furthermore, $E(Y) = r - \sum_{i=1}^k \frac{1+a_i}{2}\frac{1}{k} = r - \frac{1}{2} - \frac{1}{2k}\sum_{i=1}^k a_i > r-1$. Again, this violates C4, and therefore, the assertion follows.

Let $\{B_1, \ldots, B_k\}$ be such that $B_i \cap D = \{s_i'\}$ and each B_i has the same cardinality. Then, C3 implies $a_1 = a_i$ for all i. Since $\sum_{i=1}^k a_i = k$, we have $a_1 = 1$. By C1, $\pi^B = \pi^{B'}$ if $B \cap S_\pi = B' \cap S_\pi$, and, therefore, by (A1), $\pi^{B'} = \frac{1}{2}\pi_{B'\cap D} + \frac{1}{2}\delta_{s_i}$ for all B' such that $B' \cap S_\pi = \{s_i, s_i'\}$ for $s_i \in C$ and $s_i' \in D$.

Next, let B_1 be such that $|B_1 \cap D| = m \leq k$ and choose B_i for $i = 2, \ldots, k$ such that $B_i \cap D$ is either a singleton or empty. Then, by the argument above, $a_i = 1$ if $B_i \cap D$ is a singleton. Moreover, by C2, $a_i = 0$ if $B_i \cap D = \emptyset$. Since $\sum_{i=1}^k a_i = k$, it follows that $a_1 = m$, as desired. $\square$

Lemma 3: *Let $s \in B, s \notin D$. If $\pi = \alpha\pi_D + (1-\alpha)\delta_s$, then*

$$\pi^B = \frac{\alpha\frac{|B\cap D|}{|D|}\pi_{B\cap D} + (1-\alpha)\delta_s}{\alpha\frac{|B\cap D|}{|D|} + 1 - \alpha}$$

Proof: Let $C \subset S$ satisfy $C \cap D = \emptyset, |C| = |D| = k$, $s \notin C$ and $B \cap C = \emptyset$. Since D is a finite set and S is countable, a set C with these properties exists. Let $\pi_1 = \frac{1}{k}\sum_{\hat{s}\in C}\delta_{\hat{s}}$. Let

$$\begin{aligned}
\pi_{12} &= \frac{1}{2}\pi_1 + \frac{1}{2}\pi_D \\
\pi_{13} &= \alpha\pi_1 + (1-\alpha)\delta_s \\
\pi_{23} &= \alpha\pi_D + (1-\alpha)\delta_s \\
\pi_0 &= \frac{\alpha}{1+\alpha}\pi_1 + \frac{\alpha}{1+\alpha}\pi_D + \frac{1-\alpha}{1+\alpha}\delta_s
\end{aligned}$$

Let $B' = B \cup \{\bar{s}\}$ for $\bar{s} \in C$. Lemmas 1 and 2 imply that

$$\begin{aligned}
\pi_{12}^{B'} &= \frac{|B \cap D|\pi_{B\cap D} + \delta_{\bar{s}}}{|B \cap D| + 1} \\
\pi_{13}^{B'} &= \frac{(1-\alpha)k\delta_s + \alpha\delta_{\bar{s}}}{(1-\alpha)k + \alpha}
\end{aligned}$$

By C2, there are $\alpha_1, \alpha_2 \in (0,1)$ and $\gamma_1, \gamma_2, \gamma_3 \in (0,1), \sum \gamma_i = 1$ such that

$$\begin{aligned}
\pi_0^{B'} &= \alpha_1\pi_{12}^{B'} + (1-\alpha_1)\delta_s \\
&= \alpha_2\pi_{13}^{B'} + (1-\alpha_2)\pi_{D\cap B} \\
&= \gamma_1\delta_s + \gamma_2\delta_{\bar{s}} + \gamma_3\pi_{D\cap B}
\end{aligned}$$

Let $r = |D \cap B|$ and note that the preceding five equations yield

$$\begin{aligned}
\gamma_1 &= (1-\alpha)k/((1-\alpha)k + \alpha(r+1)) \\
\gamma_2 &= \alpha/((1-\alpha)k + \alpha(r+1)) \\
\gamma_3 &= \alpha r/((1-\alpha)k + \alpha(r+1))
\end{aligned}$$

By C2, there is $\alpha_3, \beta \in (0,1)$ such that

$$\begin{aligned}
\pi_0^{B'} &= \alpha_3\pi_{23}^{B'} + (1-\alpha_3)\delta_{\bar{s}} \\
&= \alpha_3\left(\beta\pi_{D\cap B} + (1-\beta)\delta_s\right) + (1-\alpha_3)\delta_{\bar{s}}
\end{aligned}$$

which, in turn, implies that $\beta = \alpha r/((1-\alpha)k + \alpha r)$. Thus,

$$\pi_{23}^{B'} = \pi^B = \frac{\alpha r\pi_{D\cap B} + (1-\alpha)k\delta_s}{\alpha r + (1-\alpha)k}$$

as desired. $\square$

Lemma 4: *Let $\pi_i = \pi_{D_i}$ for $i = 1, \dots, n$ and let $\pi = \sum_{i=1}^n \alpha_i \pi_i$, for $\alpha_i \in (0,1)$ such that $\sum_{i=1}^n \alpha_i = 1$. Let $A_i = D_i \cap B$ and let $k_i = |A_i|/|D_i|$. If B is $\pi-$nonnull, then $\pi^B = \sum_{i=1}^n \alpha_i k_i \pi_{A_i} / \sum_{i=1}^n \alpha_i k_i$.*

Proof: Let $B' = B \cup C$ for some $C = \{s_1, \dots, s_n\}$ such that $C \cap (\bigcup D_i) = \emptyset$. Let $\hat{\pi} = \frac{1}{2}\sum_{i=1}^n \alpha_i \pi_i + \frac{1}{2}\pi_0$ where $\pi_0 = \sum_{s\in C}^n \delta_s / n$. Let $\gamma = n/(1+n)$ and $\tilde{\pi} = \gamma \sum_{i=1}^n \alpha_i \pi_i + (1-\gamma)\delta_{s_i}$. By C2,

$$\tilde{\pi}^{B'} = \hat{\pi}^{B'}$$

Then, Lemma 3 and applying C2 repeated implies that there are $\beta_i > 0$ with $\sum \beta_i = 1$ such that

$$\hat{\pi}^{B'} = \sum_{i=1}^n \beta_i \frac{n\alpha_i k_i \pi_{A_i} + \delta_{s_i}}{n\alpha_i k_i + 1}$$

By (C3), permuting the elements of C does not alter the above equation; that is, replacing each s_i with $s_{\theta(i)}$ where θ is a permutation of C changes neither the β_i's nor the $\hat{\pi}^{B'}(\{s_i\})$'s. Therefore, there exists $w > 0$ such that, for all $i = 1, \dots, n$,

$$\frac{\beta_i}{n\alpha_i k_i + 1} = w/n$$

Define $w_i = \beta_i \frac{n\alpha_i k_i}{n\alpha_i k_i + 1}$. For j such that $k_j > 0$, it follows that $w_i/w_j = \alpha_i k_i / \alpha_j k_j$. Therefore,

$$\hat{\pi}^{B'} = (1-w)\sum_{i=1}^n \frac{\alpha_i k_i \pi_{A_i}}{\sum_{i=1}^n \alpha_i k_i} + w\pi_0$$

By C2, $\pi^{B'} = \pi^B = \sum_{i=1}^n \frac{\alpha_i k_i \pi_{A_i}}{\sum_{i=1}^n \alpha_i k_i}$, as desired. $\square$

Proof of Theorem 1: To prove the only if part of the theorem, let π be the capacity of E and μ_π be π's Möbius transform. Since π is totally monotone, $\mu_\pi(A) \geq 0$ for all A. Define $\mathcal{A}_\pi = \{D \subset S \,|\, \mu_\pi(D) \neq 0\}$ and note that $\mathcal{A}_\pi$ is a finite set; that is, $\mathcal{A}_\pi = \{D_1, \dots, D_n\}$. Let $\alpha_i = \mu(D_i)$. Then, $\pi = \sum_{i=1}^n \alpha_i \pi_{D_i}$. If B is π-nonnull, then, by Lemma 4, for all

$C \subset B$,

$$\begin{aligned}
\pi^B(C) &= \frac{\sum_{i=1}^n \alpha_i \frac{|B\cap D_i|}{|D_i|}\pi_{D_i\cap B}(C)(}{\sum_{i=1}^n \alpha_i \frac{|B\cap D_i|}{|D_i|}} \\
&= \frac{\sum_{i=1}^n \sum_{A\subset C\cap B} \sum_{\{i:D_i\cap B=A\}} \frac{|A|}{|D_i|}\mu(D_i)}{\sum_{i=1}^n \sum_{A\subset B} \sum_{\{i:D_i\cap B=A\}} \frac{|A|}{|D_i|}\mu(D_i)} \\
&= \frac{\pi^{\mathcal{P}}(C\cap B)}{\pi^{\mathcal{P}}(B)} \\
&= \pi(C\,|\,B)
\end{aligned}$$

where the last equality holds for any $\mathcal{P}$ such that $B \in \mathcal{P}$.

For the if part of the theorem, note that C1, C2 and C3 are immediate. It remains to prove that proxy updating satisfies property C4. Define $E^{\mathcal{P}}$ to be the evaluation with capacity $\pi^{\mathcal{P}}$, the proxy of π. Let E_D be an elementary evaluation and let π_D be its capacity. It is easy to verify that $E_D(X) = \min_{s\in D} X(s) \leq E_D^{\mathcal{P}}(X)$. Next, consider any simple evaluation E and let π be its capacity. As we noted above, $\pi = \sum_{i=1}^n \alpha_i \pi_{D_i}$ for some finite collection $\{D_i\}$ and coefficients $\alpha_i \geq 0$ such that $\sum_{i=1}^n \alpha_i = 1$. It is straightforward to verify that $E^{\mathcal{P}} = \sum \alpha_i E_{D_i}^{\mathcal{P}}$. It follows that $E(X) \leq E^{\mathcal{P}}(X)$ for all X. Since $E^{\mathcal{P}}(X) = \sum_{B\in\mathcal{P}} \pi^{\mathcal{P}}(B)E(X|B)$, we have $E(X) \leq E(X|B)$ for some $B \in \mathcal{P}$. $\square$

8.4 Proof of Corollaries 1 and 2

The Shapley value of i is defined as follows:

$$\rho_\pi(i) = \frac{1}{|\Theta|}\sum_{\theta\in\Theta}[\pi(\theta^i) - \pi(\theta^i\backslash\{i\})]$$

Without risk of confusion, we identify ρ_π with its additive extension to the power set of S, that is $\rho_\pi(\emptyset) = 0$, $\rho_\pi(A) = \sum_{i\in A\cap S_\pi} \rho(i)$ whenever $A \neq \emptyset$. Let $\pi^{\mathcal{P}}$ be the proxy of π.

Lemma 5: (i) $\rho_\pi(B) = 0$ if and only if $\pi^{\mathcal{P}}(B) = 0$. (ii) $\pi^{\mathcal{P}}(A\cap B)\cdot\rho_\pi(B) = \rho_\pi^D(A\cap B)\cdot\pi^{\mathcal{P}}(B)$ for all partitions $\mathcal{P}$, $B \in \mathcal{P}$ and $D = (A\cap B)\cup B^c$.

Proof: To see why (i) is true, note that the Shapley value of any $s \in S$ can be expressed in terms of the Möbius transform as follows:

$$\rho_\pi(s) = \sum_{A\ni s}\frac{\mu_\pi(A)}{|A|} \tag{A2}$$

Equation (A2) follows easily from the definition of the Shapley value and the definition of the Möbius transform. Part (i) follows from equation (A2).

Each $B \in \mathcal{P}$ is $E^{\mathcal{P}}$-unambiguous and hence, by Proposition 2 and equation (A2), $\rho_\pi(B) = \pi^{\mathcal{P}}(B)$ for all $B \in \mathcal{P}$. To conclude the proof of part (ii), we need to show that $\pi^{\mathcal{P}}(A) = \rho_\pi^D(A)$ for all $A \subset B \in \mathcal{P}$. That $\sum_{C \subset A} \mu^{\mathcal{P}}(C) = \rho^D(A)$ follows from the definition of the Möbius transform and equation (A2) applied to the game π^D for $A \subset B$ and $D = A \cup B^c$. Hence, $\rho_\pi^D(A \cap B) = \pi^{\mathcal{P}}(A \cap B)$ for all $B \in \mathcal{P}$ and all A, proving part (ii). □

Proof of Corollaries 1 and 2: Corollary 1 follows from Lemma 5. It remains to prove Corollary 2. We first prove it for an elementary evaluation E_D. Let π_D be the corresponding capacity. Then, ρ_D such that

$$\rho_D(s) = \begin{cases} 1/|D| & \text{if } s \in D \\ 0 & \text{otherwise} \end{cases}$$

is the Shapley value of π_D. Recall that Δ_D is the set of probabilities with support D. It is easy to verify that $\Delta(\pi_D) = \Delta_D$ and, using the definition of $\pi_D^{\mathcal{P}}$, for $\mathcal{P} = \{B_1, \dots, B_k\}$, we have

$$\begin{aligned} \Delta(\pi_D^{\mathcal{P}}) &= \sum_{i=1}^{k} \frac{|B_i \cap D|}{|D|} \Delta_{D \cap B_i} \\ &= \{p \in \Delta(\pi_D) \,|\, p(B_i) = \rho_D(B_i) \text{ for all } i = 1, \dots, k\} \end{aligned}$$

Thus, the corollary follows for all elementary evaluations. Let $E = \alpha_i \sum_{i=1}^{n} E_i$ where each $E_i = E_{D_i}$ is an elementary evaluation and let $\mathcal{P} = \{B_1, \dots, B_k\}$ and let π be the capacity for E. Then, by the linearity of the Shapley value,

$$\begin{aligned} \Delta(\pi) &= \sum_{j=1}^{n} \alpha_j \sum_{i=1}^{k} \frac{|B_i \cap D_j|}{|D_j|} \Delta_{D_j \cap B_i} \\ &= \{p \in \Delta(\pi) \,|\, p(B_i) = \sum_{j=1}^{n} \alpha_j \rho_{D_j}(B_i) = \rho_\pi(B_i) \text{ for all } i = 1, \dots, k\} \end{aligned}$$

and, therefore, the corollary follows. □

8.5 Proof of Proposition 3

Recall that the partition $\mathcal{P} = \{B_1, \dots, B_n\}$ is π-*extreme* if for all i, either (i) B_i is unambiguous or (ii) $\mu_\pi(A) \cdot \mu_\pi(C) > 0$ and $A \cap B_i \neq \emptyset \neq C \cap B_i$ imply $A \cap B_i = C \cap B_i$.

Thus, a partition is π-extreme if each of its elements, B_i, is either unambiguous or totally ambiguous in the sense that any positive probability element of the Möbius transform of π that intersects B_i contains all nonnull states of B_i. Showing that π-extremeness implies the law of iterated expectation for any of the three updating rules is straightforward and omitted. Below, we prove the "only if" part.

Assume $\mathcal{P}$ is not π-extreme. Then, there exists i and A, C such that $\mu_\pi(C)\mu_\pi(A) > 0$, $A \cap B_i, C \cap B_i$ and $(A \backslash C) \cap B_i$ non-empty. Without loss of generality, assume $i = 1$ and let X^z be the random variable such that

$$X^z(s) = \begin{cases} 1 & \text{if } s \in B_1^c \\ z & \text{if } s \in C \cap B_1 \\ 0 & \text{otherwise.} \end{cases}$$

Therefore, for $z \geq 1$,

$$E(X^z) = \pi(C \cup B_1^c) + \pi(C \cap B_1)(z - 1) \tag{A3}$$

Since $\mu_\pi(A) > 0$ and $A \cap B_1$ is not a subset of C, for z close to 1, $E(X^z \mid B_1) < 1 = E(X^z \mid B_j)$ for $j \neq 1$. Then,

$$E_{\mathcal{P}}(X^z) = (1 - \pi(B_1^c))E(X^z \mid B_1) + \pi(B_1^c) = (1 - \pi(B_1^c))\frac{z\pi^{\mathcal{P}}(C \cap B_1)}{\pi^{\mathcal{P}}(B_1)} + \pi(B_1^c) \tag{A4}$$

It is easy to check that $\pi^{\mathcal{P}}(D) \geq \pi(D)$ for any $\mathcal{P}$ and D. Since $\mu(A), \mu(C) > 0$ and $A \cap B_1^c \neq 0 \neq C \cap B_1^c$, $\pi^{\mathcal{P}}(B_1) < 1 - \pi(B_1^c)$ and hence $(1 - \pi(B_1^c))\frac{\pi^{\mathcal{P}}(C \cap B_1)}{\pi^{\mathcal{P}}(B_1)} > \pi(C \cap B_1)$.Then, (A3) and (A4) imply that a small increase in z near $z = 1$ will increase $E_{\mathcal{P}}(X^z)$ more than it increases $E(X^z)$ establishing that the two cannot be equal for all z. Hence, law of iterated expectation fails for proxy updating.

Next, it is easy to verify that near $z = 1$, $E^m(X^z \mid B_1) = \frac{z\pi(C)}{1 - \pi(B_1^c)} < 1$ and hence,

$$E^m_{\mathcal{P}}(X^z) = (1 - \pi(B_1^c))E^m(X^z \mid B_1) + \pi(B_1^c) = z\pi(C) + \pi(B_1^c) \tag{A5}$$

Since $C \cap B_1^c \neq \emptyset$ and $\pi(C \cap B_1) \leq \pi(C) - \mu_\pi(C)$, $\pi(C \cap B_1) < \pi(C)$. Then, (A3) and (A5) imply that a small increase in z at $z = 1$ will increase $E^m_{\mathcal{P}}$ more than it increases $E(X^z)$ proving that law of iterated expectation fails with prior-by-prior updating.

Finally, note that $E^{ds}(X^z \mid B_1) = \frac{z(\pi(C\cup B^c)-\pi(B_1^c))}{1-\pi(B_1^c)}$. Since, $\pi(C \cup B_1^c) \leq 1 - \mu_\pi(A)$, $\pi(C \cup B^c) < 1$ and therefore,

$$E^{ds}_{\mathcal{P}}(X^z) = (1 - \pi(B_1^c))E^{ds}(X^z \mid B_1) + \pi(B_1^c) = z(\pi(C \cup B^c) - \pi(B_1^c)) + \pi(B_1^c) \quad (A6)$$

Since π is totally monotone, it is supermodular. Therefore, $\pi(C \cup B_1^c) - \pi(B_1^c) \geq \pi(C) \geq \pi(C \cap B_1^c) - \mu_\pi(A)$. Again, the last inequality follows from the fact that $A \cap B_1^c \neq \emptyset$. Hence, $\pi(C \cup B_1^c) - \pi(B_1^c) > \pi(C \cap B_1)$ and therefore, (A3) and (A6) imply that a small increase in z at $z = 1$ will increase $E^{ds}_{\mathcal{P}}(X^z \mid \mathcal{P}) = E(E^{ds}(X^z \mid \mathcal{P})$ more than it increases $E(X^z)$ proving that Dempster-Shafer updating fails the law of iterated expectation.

□

8.6 Proof of Propositions 4 and 5

Applying the definition of $\pi^{\mathcal{P}}$ we obtain,

$$\pi^{\mathcal{P}}(\theta_i, y) = p_i \left(\epsilon \left(\frac{1}{2}\right)^n + (1-\epsilon)\theta_i^{\#y}(1-\theta_i)^{n-\#y} \right)$$

and, therefore,

$$\begin{aligned} E(X \mid y) &= \frac{\epsilon p_1 \left(\frac{1}{2}\right)^n + (1-\epsilon)p_1\theta_1^{\#y}(1-\theta_1)^{n-\#y}}{\epsilon\left(\frac{1}{2}\right)^n + (1-\epsilon)\left(p_1\theta_1^{\#y}(1-\theta_1)^{n-\#y} + p_2\theta_2^{\#y}(1-\theta_2)^{n-\#y}\right)} \\ &= \frac{\epsilon p_1 \left(\frac{1}{2}\right)^n + (1-\epsilon)p_1 R_\alpha(\theta_1)^n}{\epsilon\left(\frac{1}{2}\right)^n + (1-\epsilon)\left(p_1 R_\alpha(\theta_1)^n + p_2 R_\alpha(\theta_2)^n\right)} \end{aligned}$$

where $\alpha = \#y/n$ and $R_\alpha(\theta) = \theta^\alpha(1-\theta)^{(1-\alpha)}$. Recall that $\theta_1 < 1/2 < \theta_2$ and consider $\alpha \leq 1/2$. Then, $R_\alpha(\theta_2) < \frac{1}{2}$ and $R_\alpha(\theta_1) > (<)\frac{1}{2}$ if and only if $\alpha < (>)\frac{\ln(1/2)-\ln(\theta_1)}{\ln(1-\theta_1)-\ln(\theta_1)} = r_1$. Similarly, if $\alpha \geq 1/2$, then $R_\alpha(\theta_1) < \frac{1}{2}$ and $R_\alpha(\theta_2) > (<)\frac{1}{2}$ if and only if $\alpha > (<)r_2$. Applying these facts to the limit, as n goes to infinity, of right-hand side of the equation above proves Proposition 4. The proof of Proposition 5 is straightforward and, therefore, omitted. □

8.7 Proof of Theorems 2 and 3

Proof of Theorem 2: Assume that every $t \in T$ is nonnull; that is, $\pi^e([S \times \{t\}] \cup A) > \pi^e(A)$ for some $A \subset S \times T$. This assumption is without loss of generality since we can eliminate all null t to obtain $\hat{\mathcal{T}} \subset \mathcal{T}$, construct the set of priors on $S \times \hat{T}$ as described below, prove the desired result for this $\hat{T}$ and then extend those priors to $S \times T$ by setting every $\mathcal{P}(s,t)$ equal to zero for all $t \in T \backslash \hat{T}$ all priors p.

For each $t \in T$, let $E(X \,|\, t) = E^e(X \,|\, S \times \{t\})$. By definition, the capacity of E^e is totally monotone. Then, Theorem 1 ensures that the capacity of $E^e(\cdot \,|\, S \times \{t\})$ is also totally monotone and hence supermodular. Then, there is a convex, compact set of probabilities Δ_t on $S \times T$ such that

$$E(Z \,|\, t) = \min_{p_t \in \Delta_t} \int Z dp_t$$

for all random variables Z on $S \times T$. By Theorem 1, $p_t(S \times \{t\}) = 1$ for all $p_t \in \Delta_t$. Identify each $p_t \in \Delta_t$ with a probability $\hat{p}$ on S such that $\hat{p}_t(A) = p_t(A \times \{t\})$.

Next, let $\hat{\pi}$ be the capacity on T such that $\hat{\pi}(R) = \pi^e(S \times R)$ for all $R \subset T$. Again, since π^e is totally monotone, so is $\hat{\pi}$ and therefore, there is a convex, compact set of probabilities Δ^* on T such that

$$E(\hat{Z}) = \min_{q \in \Delta^*} \int \hat{Z} dq$$

for all random variables $\hat{Z}$ on T.

Define a convex, compact set of probabilities, Δ, on S as follows:

$$\Delta = \left\{ p = \sum_{t \in T} q(t) \cdot p_t \;\middle|\; q \in \Delta^*, p_t \in \Delta_t \text{ for all } t \in T \right\}$$

Clearly, for any $\mathcal{T}$-measurable Z on $S \times T$, we have

$$E^e(Z) = \min_{q \in \Delta} \int Z dq$$

The first and the last display equations above yield, for any random variable X on S,

$$\begin{aligned} E^e(E^e(X^e \,|\, \mathcal{T})) &= \min_{q \in \Delta^*} \sum_{t \in T} q(t) \min_{p_t \in \Delta_t} \sum_{s \in S} X^e(s,t) p_t(s,t) \\ &= \min_{q \in \Delta^*} \sum_{\omega \in T} q(t) \min_{p_t \in \Delta_t} \sum_{s \in S} X(s) \hat{p}_t(s) \\ &= \min_{p \in \Delta} \sum X(s) p(s) \end{aligned}$$

Hence, $E_{\mathcal{T}}(X) = \min_{p \in \Delta} \int X dp$ as desired. $\square$

Proof of Theorem 3: We will show that for any finite nonempty $\hat{S}$ and collection of probabilities, $\Delta_0 = \{p_1, \ldots, p_m\} \subset \Delta_{\hat{S}}$ such that (i) no p_i is in the convex hull of the remaining elements $\{p_1, \ldots, p_{i-1}, p_{i+1}, \ldots, p_m\}$ and (ii) $p_i(s)$ is a rational number for every i and s, there exists a general compound evaluation $E_{\mathcal{T}}$ such that

$$E_{\mathcal{T}}(X) = \min_{p_i \in \Delta_0} \sum_{s \in \hat{S}} X(s) p_i(s)$$

Let $\Delta_0 = \{p_1, \ldots, p_m\} \subset \Delta_{\hat{S}}$ be a set with the properties described above. Then, there are integers n, $k_i(s)$ for all j, s such that $p_j(s) = \frac{k_j(s)}{n}$. We will define E^e by defining the Möbius transform μ_{π^e} of its capacity π^e on $S \times T$.

For all $j = \{1, \ldots, m\}$, let $F_j : \{1, \ldots, n\} \to \hat{S}$ be a function such that $|F_j^{-1}(s)| = k_j(s)$. Then, set $T = \Delta_0$ and define $A_i^e \subset \hat{S} \times T$ as follows:

$$A_i^e = \{(F_j(i), p_j) \,|\, j = 1, \ldots, m\}$$

Then, let

$$\mu_{\pi^e}(A^e) = \frac{|\{i \,|\, A^e = A_i^e\}|}{n}$$

and let E^e be the evaluation that has π^e as its capacity. Next, we will define a capacity π with support $\hat{S}$ by defining its Möbius transform μ_π: let

$$A_i = \{s \,|\, (s, p_j) \in A_i^e \text{ for some } j\}$$

Then, let $\mu_\pi(A) = \frac{|\{i \,|\, A = A_i\}|}{n}$. Let E be the evaluation that has π as its capacity.

By construction, $\pi^e(\cdot \,|\, p_j) = p_j$ for all p_j. Note also that $\pi^e(A^e) = 0$ unless A contains some A_j^e, which means the marginal π_T^e of π^e on T satisfies $\pi_T^e(R) = 0$ unless $R = T$. Hence,

$$E_{\mathcal{T}}(X) = \min_{p_i \in \Delta_0} \int X dp_i$$

The equation above still holds when Δ_0 is replaced with its convex hull.

The preceding argument establishes the desired result whenever Δ is a polytope with rational extreme points. Let $\mathcal{C}$ be the set of all nonempty, compact convex subsets of $\Delta_{\hat{S}}$. Define the mapping $f : \mathcal{C} \times \mathcal{V} \to I\!R_+$ as follows:

$$f(\hat{\Delta}, X) = \left| \min_{p \in \Delta} \int X dp - \min_{p \in \hat{\Delta}} \int X dp \right|$$

Endow $\mathcal{C}$ with the Hausdorff metric, $\mathcal{V}$ with the Euclidian metric, $\mathcal{C} \times \mathcal{V}$ with the corresponding product metric and note that the function f is continuous. Since $\mathcal{V}$ is compact, the function f^* defined by

$$f^*(\hat{\Delta}) = \max_{X \in \mathcal{V}} f(\hat{\Delta}, X)$$

is also continuous. Since the set of all polytopes in $\mathcal{C}$ is dense in $\mathcal{C}$ if follows that for $\epsilon > 0$, there is some polytope $\hat{\Delta}$ such that $f^*(\hat{\Delta}) < \epsilon$. Then, the preceding argument yields $\mathcal{T}$ such that $E_{\mathcal{T}}(X) = \min_{p \in \hat{\Delta}} \int X dp$ and completes the proof.

□

References

Dempster, A. P. (1967) “Upper and Lower Probabilities Induced by a Multivalued Mapping,” *The Annals of Mathematical Statistics*, 38, 325–339.

Cerreia-Vioglio, S, F. Maccheroni and M. Marinacci (2015) “Put-Call Parity and Market Frictions,” *Journal of Economic Theory*, 157, 730–762.

Ellis, A. (2016) “Condorcet Meets Ellsberg,” *Theoretical Economics*, 11, 865–895

Ellsberg, D. (1961) “Risk, Ambiguity and the Savage Axioms,” *Quarterly Journal of Economics*, 75, 643–669.

Epstein L. G. and M. Schneider (2003) “Recursive Multiple-Priors,” *Journal of Economic Theory*, 113, 1–31

Epstein L. G. and M. Schneider (2007) “Learning under Ambiguity,” *Review of Economic Studies*, 74, 1275–1303

Ghirardato P. and M. Marinacci (2001) “Risk, Ambiguity, and the Separation of Utility and Beliefs,” *Mathematics of Operations Research*, 26, 864–890.

Ghirardato, P., F. Maccheroni and M. Marinacci (2004) “Differentiating Ambiguity and Ambiguity Attitude’s *Journal of Economic Theory*, 118, 133–173.

Gilboa, I, and D. Schmeidler (1989) “Maxmin Expected Utility with a Non-Unique Prior,” *Journal of Mathematical Economics*, 18, 141–153.

Gilboa, I, and D. Schmeidler (1993) “Updating Ambiguous Beliefs,” *Journal of Economic Theory*, 59, 33–49.

Gilboa, I. and D. Schmeidler (1994) “Additive representations of non-additive measures and the Choquet integral,” *Annals of Operations Research*, 52, 43–65.

Jaffray, J.Y. (1992), “Bayesian Updating and Belief Functions,” *IEEE Transactions on Systems, Man and Cybernetics*, 22, 1144–1152.

Kalai, E. and D. Samet (1987) “On Weighted Shapley Values,” *International Journal of Game Theory*, 16, 205–222.

Klibanoff, P and E. Hanany “Updating Ambiguity Averse Preferences,” *The B.E. Journal of Theoretical Economics*, 9.

Kreps, D. and E. Porteus (1978), “Temporal Resolution of Uncertainty and Dynamic Choice Theory,” *Econometrica*, 46, 185–200.

Pires, C. P. (2002) “A Rule for Updating Ambiguous Beliefs,” *Theory and Decision*, 53: 137–152.

Raiffa, H. (1961) “Risk, ambiguity and the Savage axioms: Comment,” *Quarterly Journal of Economics*, 75, 690–694.

Saito, K., (2015) “Preferences for Flexibility and Randomization under Uncertainty,” *American Economic Review*, 105, 1246–1271.

Seidenfeld T. and L. Wasserman (1993) “Dilations for Sets of Probabilities,” *Annals of Statistics*, 21, 1139–1154.

Savage, L. J. (1954) The Foundations of Statistics, Wiley, New York.

Schmeidler, D. (1989) “Subjective Probability and Expected Utility without Additivity,” *Econometrica*, 57, 571–587.

Shafer, G. A Mathematical Theory of Evidence, Princeton University Press, 1976.

Siniscalchi, M. (2011) “Dynamic Choice under Ambiguity,” *Theoretical Economics*, 6, 379–421.

Wald, A. (1945) “Sequential Tests of Statistical Hypotheses,” *The Annals of Mathematical Statistics*, 16, 117–186

Wasserman, L. A. and J. Kadane (1990), “Bayes’ Theorem for Choquet Capacities,” *Annals of Statistics*, 18, 1328–1339.