Evaluating Ambiguous Random Variables and Updating by Proxy

Faruk Gul

and

Wolfgang Pesendorfer

Princeton University

February 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 two characterization theorems: the first relates our rule to standard Bayesian updating; the second characterizes the dynamic behavior of an agent who adopts our rule.

1. Introduction

Consider the following Ellsberg-type experiment: a ball is drawn from an urn consisting of blue and green balls of unknown proportion. After the ball is revealed, the decision maker flips a coin; if it comes up heads, he wins if the ball is blue; if it comes up tails, he wins if the ball is green. In this situation, standard models of ambiguity (Gilboa and Schmeidler (1989), Schmeidler (1989), Klibanoff, Marinacci and Mukherji (2005)) imply that the coin flip hedges the uncertainty of the draw. By contrast, consider the situation in which the order is reversed; the outcome of the coin flip is revealed before the ball is drawn from the urn. Most ambiguity models interpret this as a situation in which the coin flip does not hedge the uncertain draw from the urn. In particular, all axiomatic models set in the Anscombe-Aumann framework implicitly assume that a coin flip after drawing the ball hedges the ambiguity since this notion of hedging is built into Schmeidler's definition of uncertainty aversion, while a coin toss prior to drawing the ball does not.[1] Thus, the standard notion of hedging developed in ambiguity models implicitly assumes that the manner in which uncertainty resolves affects the value of prospects.

The main contribution of this paper is a model of belief updating consistent with the description above; that is, we formulate an updating rule that ensures that rolling back payoffs in a tree in which the coin toss occurs first yields values consistent with no hedging while the corresponding calculation in a tree in which the ball is drawn first yields values consistent with 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 does not, in general, satisfy the Law of Iterated Expectation; that is, decision makers that adopt our updating rule will not be indifferent to the timing of resolution of uncertainty. This sensitivity to the order in which uncertainty resolves is evident in

[1] If it did, agents could eliminate ambiguity by randomizing over acts. However, some models (for example, Ellis (2016)) do assume that the decision maker can hedge with a coin flip even if the coin flip is revealed before the unambiguous state.

the hedging discussion of the previous paragraph: consider the random variable, X, that yields 1 if the ball is green and the coin comes up heads or if the ball is blue and the coin comes up tails; in all other cases, the random variable takes the value 0. The coin flip after the draw from the urn hedges and hence the "value" of the random variable is $1/2$ if uncertainty resolves in this order. If the coin toss is resolved first, it does not hedge the ambiguity associated with the draw from the urn and hence the value of X is strictly less than $1/2$.

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 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. To put it differently, in Kreps-Porteus how much time elapses between the coin toss and the draw from the urn matters. 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. For this reason, our model is particularly amenable to a subjective interpretation. An agent who believes randomization hedges ambiguity behaves *as if* the ball is drawn first while an agent who does not believe randomization helps against ambiguity behaves as if the coin is tossed first. Intermediate situations between these two extremes can be modelled with the general information structures and the associated general models described below.[2]

Our primitive is an *evaluation* that associates a real number with random variables. We write $E(X)$ for the evaluation of the random variable X. Evaluations should be viewed as expectation operators for ambiguity theory. To study updating, we define compound evaluations: let $\mathcal{P} = \{B_1, \ldots, B_k\}$ be a partition of the state space. Assume that uncertainty resolves in two-stages; the first stage reveals the partition element $B \in \mathcal{P}$ while the second stage reveals the state $s \in B$. We let $E_{\mathcal{P}}(X)$ denote the evaluation of the random

[2] 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 subjective sequencing of resolution serves a role similar to Saito's parameter.

variable X when uncertainty resolves in this fashion and call it a *compound evaluation.* Let $E(X \mid \mathcal{P})$ denote 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 the first part of 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 $B_i \in \mathcal{P}$, we set $Y(s)$ equal to the conditional expectation $E(X \mid B)$ of X and then, we evaluate Y.

The main challenge is coming up with a definition of $E(X \mid B)$, the conditional evaluation of X given $B \in \mathcal{P}$. The following two examples illustrate the notion of conditional evaluation under two commonly used updating rules and how we depart from those rules.

1.1 Learning from Ambiguous Information

In this example, nature draws, consecutively, two balls from separate urns. The agent bets on the color of the first draw after observing the second draw. Nature draws the first ball from urn O consisting of one white and one black ball. If nature picks the white ball, it draws the second ball from urn W; if nature chooses the black ball, it draws the second ball from urn B. Urns W and B contain 12 balls, each either green or red. Urn W contains at least 4 red and at least 2 green balls while urn B contains least 4 green and at least 2 red 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 nature's first draw was white, and 0 otherwise.

Assume the agent translates the description of urn W into the following set of priors: the probability of a red draw from urn W is in the interval $[1/3, 5/6]$ and the probability of a red draw from urn R is in the interval $[1/6, 2/3]$; urn O translates to a unique prior of .5 for a white or black ball. The agent is ambiguity averse and uses the maxmin criterion to evaluate random variables. Thus, before observing nature's second draw, the value of X is $1/2$.

One commonly used updating rule is prior-by-prior updating. Pires (2002) provides an axiomatization of this rule in the Anscombe-Aumann setting. With this rule, the agent updates every prior according to Bayes' rule and then applies the maxmin criterion to the set of updated priors. If p, q are the probabilities of a red draw from urns W and R respectively, then the posterior probability of W given a red draw is $\frac{p}{p+q}$ and the probability of W given a green draw is $\frac{1-p}{2-p-q}$. Thus, the set of posteriors is

$$\Pr(W|r) \in [1/3, 5/8]$$
$$\Pr(W|g) \in [1/6, 2/3]$$

With the maxmin criterion applied to the set of updated beliefs, the bet is worth 1/3 after observing red and 1/6 after observing green, both less the ex ante value. Thus, if a maxmin agent updates prior-by-prior, then *all news is bad news.* Following a red draw, the agent evaluates X *as if* urn W has 4 red balls and 8 green balls and urn R has of 2 red balls and 10 green balls. Following a green draw, the agent evaluates X as if urn W consisted of 2 green balls and 10 red balls while urn R consisted of 8 green balls and 4 red balls.[3]

Intuitively, one might expect that despite the ambiguity, a red draw constitutes good news about X while a green draw is bad news. This is indeed what happens with our updating rule. When updating, our agents create proxy urns R^* and W^* that render the information (red or green) unambiguous. In this example, the proxy urn W^* has 7 green balls and 5 red balls while W^* has 7 green balls and 5 red balls. Conditional on observing a red draw, the value of X is then 7/12 and conditional on a green draw the value of X is 5/12. In Theorem 1, below, we provide a characterization of this proxy rule.

A second commonly used rule is the Dempster-Shafer rule (Dempster 1967, Shafer 1978). To illustrate it, consider the following modified version of the example above. As above, nature first draws a ball from urn O that consists of one white and one black ball. If the ball is white, nature draws a second ball from urn $\bar{W}$; if the ball is black, nature draws a second ball from urn $\bar{B}$. Urn $\bar{W}$ contains 4 red or green balls, at least 1 is red, at least 1 is green. Urn $\bar{B}$ contains 2 red balls and 2 green balls. As above, the decision

[3] 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.

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 nature's first draw was white, and 0 otherwise.

Assume the agent translates the description of urn $\bar{W}$ into the following set of priors: the probability of a red draw from urn $\bar{W}$ is in the interval $[1/4, 3/4]$; urns O and $\bar{B}$ each correspond to a unique prior of $1/2$. As above, the ex ante value of the X is $1/2$.

According to the Dempster-Shafer rule, the agent picks a "maximum likelihood" urn consistent with the description of urn $\bar{W}$ that generates a particular draw. Thus, upon observing a red ball, the agent updates *as if* the urn $\bar{W}$ consists of 3 red balls and 1 green ball. Thus, $p(W|r) = 3/5$. Similarly, upon observing a green ball, the agent updates as if urn $\bar{W}$ consists of 3 red balls and 1 green ball. Thus, $p(W|g) = 3/5$. In this case, irrespective of the draw, the agent's value of the random variable X increases above its ex ante value.

Intuitively, one might expect that observing a red or green draw provides no information about X. While the two urns differ, the two colors play a symmetric role in both and, as a result, it seems plausible that the agent leaves her valuation of X unchanged after observing either color. Again, this is what happens with our updating rule. More specifically, in the above example, an agent following our proposed rule would use the proxy urn $\bar{W}^*$ consisting of 2 red balls and 2 green balls. Urn $\bar{B}$ would stay unchanged and, therefore, the color draw provides no information about the value of X.

1.2 Evaluations and Proxy Updating

Our starting point is an evaluation that is characterized as a Choquet integral with respect to a totally monotone capacity. Our updating formula can be described as follows: first, we associate a proxy evaluation, $E^{\mathcal{P}}$, with every evaluation E and information partition $\mathcal{P}$. The proxy evaluation $E^{\mathcal{P}}$ is similar to E but has less ambiguity than E; in particular, every element of the information partition is $E^{\mathcal{P}}$-unambiguous. In the examples above, the proxy evaluation is the expectation of the random variable if the urns we called proxy urns are substituted for the original urns.

Once a decision maker learns $B \in \mathcal{P}$, B becomes unambiguous. Hence, we assume that the conditional evaluations of E are the same as the corresponding conditional evaluations of $E^{\mathcal{P}}$. To put it another way, since every element of $\mathcal{P}$ is $E^{\mathcal{P}}$-unambiguous, we compute the conditionals of $E^{\mathcal{P}}$ easily by applying Bayes' Law. Then, we *define* $E(\cdot \mid B)$ by setting it equal to $E^{\mathcal{P}}(\cdot \mid B)$.

For every totally monotone capacity π, there is a probability μ on the set of all nonempty subsets of S such that

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

This function μ is called the Möbius transform of π. The proxy capacity $\pi^{\mathcal{P}}$ of the original capacity, π, is defined as follows:

$$\begin{aligned}
\mu^{\mathcal{P}}(A) &= \sum_{B \in \mathcal{P}} \sum_{\{C : B \cap C = A\}} \frac{|A|}{|C|} \cdot \mu(C) \\
\pi^{\mathcal{P}}(A) &= \sum_{C \subset A} \mu^{\mathcal{P}}(C) \\
\pi(A \mid B) &= \frac{\pi^{\mathcal{P}}(A \cap B)}{\pi^{\mathcal{P}}(B)}
\end{aligned} \tag{1}$$

The first equation defines how the agent forms her proxy. When she conditions on an event B and this event is ambiguous, she modifies the capacity to ensure B becomes unambiguous. To do so, she adjusts the weight of any set $A = B \cap C$ in the Möbius transform of the proxy. In particular, she distributes the weight of C to A in proportion to the number of elements A shares with C. In section 3, we offer a characterization of our updating rule and proxy mapping and establish the validity of the updating formula above.

As we show in Corollary 2 below, we can also describe the proxy rule as a modified prior-by-prior update. The Choquet integral of a random variable with respect to a totally monotone capacity can equivalently be written as the expected value of that random variable with respect to the least favorable prior in the core of the capacity.[4] The naive prior-by-prior revision rule simply updates every prior in the core of the capacity and uses the least favorable updated prior for the conditional evaluation. Our proxy rule selects a

[4] The core of the capacity consists of all probabilities that assign to each event a value at least as large as the capacity.

subset of priors, updates each prior in the subset and then uses the least favorable prior. The subset depends on the information structure in the following way: Let ρ_π be the Shapley value of the totally monotone capacity π. Hence, ρ_π is an element of the core of π. Then, the subset of priors used for updating are those that agree with ρ_π on every element of the information partition.

1.3 Preview of Results

Section 2 provides axioms that ensure that the simple evaluation is a Choquet integral with a totally monotone capacity. In section 3, we derive our updating rule from a set of requirements that facilitate a comparison to the standard Bayesian model. In particular, we impose a weak version of the Law of Iterated Expectation which we call the "not all news can be bad news" property: if $E(X \mid B) \leq E(X)$ for all E-nonnull $B \in \mathcal{P}$, then $E_{\mathcal{P}}(X) := E(E(X \mid \mathcal{P})) = E(X)$. The interpretation of this property is as follows: conditioning reduces ambiguity and because of this reduction, all news can be good news. However, reduction of uncertainty cannot be uniformly bad. Section 5 applies our updating rule to a standard inference problem enriched by the presence of ambiguity. Section 4 establishes properties of conditional and compound evaluations and compares proxy updating to other rules.

In sections 3-5, we assume that the information that agents receive is a partition on the "payoff relevant" state space; that is, on the domain of the random variables in question. Hence, we have analyzed compound evaluations $E_{\mathcal{P}}$ given some partition $\mathcal{P}$. In section 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. This more general informational set up yields generalized compound evaluations which are better suited for situations in which information is subjective and unobservable to the analyst.

In Theorem 2, we show that every generalized compound evaluation is as a maxmin evaluation for some set of priors; in Theorem 3 we show that 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 information aquisition.

1.4 Related Literature

As is typical for ambiguity models, our theory does not satisfy the law of iterated expectations; that is, $E(X)$ need not equal $E(E(X|\mathcal{P}))$. Epstein and Schneider (2003) ensure that the law of iterated expectations is satisfied by restricting to conditioning events.[5] In contrast to Epstein and Schneider (2003), we consider all conditioning events. Restricting the possible conditioning events renders ambiguity theory incapable of analyzing problems 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) interprets the difference between $E(X)$ and $E(E(X|\mathcal{P})$ as the result of *changing preferences*; the former expression represents the ex ante evaluation while $E(X|B)$ represents the second period's valuation if event B occurs. Siniscalchi's agent faces a problem analogous to Strotz' (1955) sophisticated agent and, following Strotz, solves it with consistent planning. As a result, Siniscalchi's agents may seek commitment. Hanany and Klibanoff (2009) posit a non-recursive updating rule. As in Siniscalchi (2011), the way the random variable evolves does not affect its ex ante value but, in contrast to Siniscalchi (2011), the conditional value $E(X|B)$ changes according to the ex ante optimal plan. Thus, Hanany and Klibanoff (2009) give up recursivity and enforce equality of $E(X)$ and $E(E(X|\mathcal{P})$. By contrast, we follow Kreps and Porteus (1978) and attribute the difference between $E(X)$ and $E(E(X|\mathcal{P}))$ to differences in the timing of resolution of uncertainty.[6] As a result, ambiguity does not generate a preference for commitment in our model.

The two standard 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).

---

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

[6] 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 with value 1 if $s \in A$ and 0 otherwise. The restriction to positive random variables is for convenience. Allowing for random variables that take on negative values complicates the notation but creates no novel issues.

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 $E(X + Y) = E(X)$; Y is *E-unambiguous* if $E(X+Y) = E(X) + E(Y)$ for all 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. An event A is E-unambiguous if $E(X_A) + E(X_{A^c}) = E(X)$ for all X.

Our 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. 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 of uncertainty aversion and ensures that the capacity associated with the Choquet-integral representation of the evaluation is a totally monotone capacity. Given any collection of random variable $\mathcal{X} = (X_1, \ldots, X_n)$, $\alpha \geq 0$ and $A \subset S$, $K_\alpha(\mathcal{X}, A) = |\{i \,|\, X_i(s) \geq \alpha \text{ for all } s \in A\}|$. We say that the collection of random variables $\mathcal{X} = (X_1, \ldots, X_n)$ set-wise dominates the collection $\mathcal{Y} = (Y_1, \ldots, Y_n)$ if $K_\alpha(\mathcal{X}, A) \geq K_\alpha(\mathcal{Y}, A)$ for all α, A.

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

We let $E, \tilde{E}$ and $\hat{E}$ denote generic simple evaluations. A (simple) capacity is a function $\pi : 2^S \to [0, 1]$ such that (i) $\pi(\emptyset) = 0$, (ii) there exist a finite D such that $\pi(D) = 1$ and $\pi(A) = \pi(A \cap D)$ for all A. (iii) $\pi(A) \leq \pi(B)$ whenever $A \subset B$. Let S_π denote the minimal support of the capacity π. We say that $\mu : 2^{S_\pi} \to \mathbb{R}$ is the Möbius transform 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 $A \subset S_\pi$. Totally 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, define the Choquet integral of X with respect to the capacity π 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 and hence, we call this π the capacity of E.

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

Proof: See Appendix

The set D is a support for π if $\pi(D \cap A) = \pi(A)$ for all B. 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 $S_X = \{s \mid X(s) > 0\}$, 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.

3. 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 evaluation of the two stage random variable $(X, \mathcal{P})$. The main object of interest for our analysis is the notion of an *updating rule*: for any E and E-nonnull event B, let $E(\cdot \mid B)$ 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 $E(\cdot \mid B)$ be unspecified. For every X and finite partition $\mathcal{P}$ of S, let $E(X \mid \mathcal{P})$ denote the random variable Y such that $Y(s) = E(X \mid B)$ whenever $s \in B$ and B is E-nonnull. The agent evaluates the two stage random variable $(X, \mathcal{P})$ by iterated expectation; that is,

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

Note that we do not define Y at $s \in B \in \mathcal{P}$ if B is E-null. We are able to avoid specifying $E(\cdot \mid B)$ for E-null B because $E(\cdot \mid B)$ is multiplied by 0 when computing $E(E(X \mid \mathcal{P}))$.

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 $\pi^{\mathcal{P}}$ and $\mu_\pi^{\mathcal{P}}$ for π's proxy capacity and its Möbius transform. Equation (6), below, defines the Möbius transform of the proxy: for all events A,

$$\mu_\pi^{\mathcal{P}}(A) = \sum_{B\in\mathcal{P}} \sum_{\{D:D\cap B=A\}} \frac{|A|}{|D|} \cdot \mu_\pi(D) \tag{6}$$

The proxy capacity, $\pi^{\mathcal{P}}$, is then $\pi^{\mathcal{P}}(A) = \sum_{C\subset A} \mu_\pi^{\mathcal{P}}(C)$. We refer to the evaluation corresponding to $\pi^{\mathcal{P}}$ as the proxy evaluation $E^{\mathcal{P}}$, that is, $E^{\mathcal{P}}(X) = \int X d\pi^{\mathcal{P}}$. If π is a probability, then $\mu(A) = 0$ for all non-singleton events, and therefore, the proxy coincides with the original capacity. More generally, the proxy and the original capacity coincide whenever every $B \in \mathcal{P}$ is $E-$unambiguous. This follows from the fact that there can be no set C that intersects both B and its complement such that $\mu_\pi(C) > 0$ if B is unambiguous. (For a proof of this assertion, see Fact 1 in the appendix.) Equation (6) then implies that $\mu_\pi^{\mathcal{P}} = \mu_\pi$. Conversely, by equation (6), $\mu_\pi^{\mathcal{P}}(A) > 0$ implies $A \subset B$ for some $B \in \mathcal{P}$. Thus, the Möbius transform of $\mu_\pi^{\mathcal{P}}$ assigns zero weight to any event that intersects more than one information set. Fact 1 in the appendix shows that this, in turn, implies that every element of the information partition is $E^{\mathcal{P}}$-unambiguous.

Given the proxy capacity $\pi^{\mathcal{P}}$, the agent updates according to Bayes' rule:

$$\pi(A \,|\, B) := \frac{\pi^{\mathcal{P}}(A\cap B)}{\pi^{\mathcal{P}}(B)} \tag{7}$$

Note that $\pi(\cdot \,|\, B)$ does not depend on the information partition $\mathcal{P}$ as long as $B \in \mathcal{P}$. This follows from the fact that $\pi^{\mathcal{P}}(A) = \pi^{\mathcal{P}'}(A)$ for all $A \subset B \in \mathcal{P} \cap \mathcal{P}'$.

Let π be the capacity of E. We say that the conditional evaluation $E(\cdot \,|\, B)$ is the *proxy update of* E if

$$E(X \,|\, B) = \int X d\pi(\cdot|B)$$

where $\pi(\cdot|B)$ is as defined in equations (6) and (7).

To derive the proxy rule, we impose 4 conditions on the conditional evaluation. The first condition considers the simple class of evaluations for which there is some finite set D such that $E(X) = \min_{s\in D} X(s)$ for all random variables X. We call those *elementary evaluations*. Their capacity has the form

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

We write π_D for an elementary capacity with set D and E_D for the corresponding evaluation. 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 intersection 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 in 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. 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 Bayesian Law (i.e., when calculation 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 $A' = \{h(s) : s \in A\}$. For any evaluation E, let S_E denote the support of the capacity π of E.

C3, like C1 and C2 is unobjectionable and satisfied by all known updating rules. Nevertheless, providing a version of our updating rule that does not impose C3 is not difficult.

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

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

Proof: See Appendix.

Proxy updating is related to the Shapley values. Specifically, it can be interpreted as the ratio of Shapley values of appropriately defined games. For any capacity π, let $\rho_\pi(s)$ denote the Shapley value of 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 2 shows that the updated capacity is the ratio of Shapley values:

Corollary 1: *For any nonnull $B \in \mathcal{P}$ and $D := A \cup B^c$,*

$$\pi(A \,|\, B) = \frac{\rho_\pi^D(A \cap B)}{\rho_\pi(B)}$$

Next, we show that proxy updates can be interpreted as a modified prior-by-prior rule. First, note that we can represent any simple evaluation E as a maxmin evaluation

[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 $\{1, \ldots, |S_\pi|\}$ where $|A|$ denote the cardinality of any set A. 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:

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

for a particular set of priors. 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. For example, consider the elementary evaluation E_D with $D = \{1, 2, 3\}$. In that case, the core of E_D's capacity is the set of all probabilities on D.

A frequently used rule updates every prior in the core of the capacity and uses the least favorable updated prior for the conditional evaluation. As we noted in the introduction, this rule may lead to implausible behavior, in particular, to instances in which all news is bad news. Proxy updating resolves these issues by restricting the set of priors used for updating. For the three state example, consider the following information partition: $\{\{1, 2\}, \{3\}\}$. If the agent learns that the state is 1 or 2 she considers the subset of priors that assign probability 2/3 to the event $\{1, 2\}$ and updates each. She then evaluates random variables conditional on $\{1, 2\}$ with the least favorable among those updated priors.

More generally, recall that ρ_π is the Shapley value of π and let

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

The set of priors $\Delta^*(\pi)$ is the subset of $\Delta(\pi)$ such that each prior assigns $\rho_\pi(B)$ to every cell B of the information partition. As we noted above, our model draws a distinction between the agents ex ante and ex post perceptions of ambiguity. Once an information cell is revealed, it is no longer deemed ambiguous and the agent uses the Shapley value ρ_π to judge its likelihood. For updating purposes, she restricts the set of priors to those 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*

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

for any nonnull $B \in \mathcal{P}$.

4. Properties of Proxy Updating

We consider four properties for compound evaluations and discuss three distinct models of updating within the context of these properties. The first property is sometimes called consequentialism; it requires that the value of a random variable X conditional on an event B depend only on the value X takes on B:

$$E(X \mid B) = E(X_B \mid B) \text{ for all } E\text{-nonnull } B \text{ and } X \tag{c}$$

Proxy updates satisfy consequentialism since $\pi(\cdot \mid B)$ has support contained in B. The second property is recursivity:

$$E_{\mathcal{P}}(X) = E(E(X \mid \mathcal{P})) \text{ for all } X \tag{r}$$

Recursivity means that the compound evaluation is computed by rollback (i.e., backward induction). Our compound evaluations satisfy recursivity by definition; that is, we defined the compound evaluation $E_{\mathcal{P}}$ recursively. The third property is indifference to timing:

$$E_{\mathcal{P}}(X) = E(X) \text{ for all } X \tag{it}$$

Under updating by proxy, compound evaluations, in general, do not satisfy indifference to timing. This is the property that we have abandoned to avoid well-known impossibility results on updating under ambiguity.

In the literature, the terms dynamic consistency and recursivity are used in conflicting and overlapping ways. A complete catalogue of the various definitions of these terms and the relationships among these definitions is not feasible. Much of the discussion of dynamic consistency (or inconsistency) does not consider the possibility of relaxing indifference to timing. We will refer to the conjunction of recursivity and indifference to timing as dynamic consistency or, equivalently, the *law of iterated expectation*:

$$E(X) = E_{\mathcal{P}}(X) = E(E(X \mid \mathcal{P})) \text{ for all } X, \mathcal{P} \tag{lie}$$

The notion of dynamic consistency above is stronger than the corresponding desirata in Epstein and Schneider (2003) who only impose it for some fixed $\mathcal{P}$. Siniscalchi (2011) abandons the law of iterative expectation and assumes that the difference between the ex ante evaluation and the two-stage evaluation (i.e., $E(X) \neq E(E(X \,|\, \mathcal{P}))$) leads to a preference for commitment. In general, no model of updating under ambiguity satisfies dynamic consistency.

Our compound evaluations satisfy consequentialism and recursivity; they do not satisfy indifference to timing or the law of iterated expectation. Instead, we have imposed, (C2), weak dynamic consistency a equivalently "not all news can be bad news" property:

$$E(X|\,B) \leq \alpha \text{ for all } B \in \mathcal{P} \text{ implies } E(X) \leq \alpha \tag{wdc}$$

While weak dynamic consistency rules out the possibility that all news is bad news, it allows for the possibility that all news might be good news. The example below illustrates why ambiguity can lead to situations in which every realization of a signal provides good news about a random variable. Consider the following Ellsberg-type experiment: a ball is drawn from an urn consisting of blue and green balls of unknown proportion and a fair coin is flipped. The random variable X yields 1 if the coin comes up heads and the ball is blue; or if the coin comes up tails and the ball is green. The state space is $S = \{hb, hg, tb, tg\}$. Let $B = \{hb, tb\}, G = \{hg, tg\}, T = \{tb, tg\}, H = \{hb, hg\}$. The evaluation considers draws from the urn to be ambiguous but flips of a coin to be unambiguous. For example, assume that $E(1_B) = E(1_G) = E(X) = 1/4$ and $E(1_T) = E(1_H) = 1/2$. Consider the information partition $\mathcal{P} = \{B, G\}$. In that case, conditional in each possible cell of the partition, the random variable X presents the agent with a coin toss; heads wins if the signal is B, and tails wins if the signal is G. Updating by proxy implies that the value of X conditional on either realization is 1/2, greater than the original 1/4. Note that the "all news is good news" feature in this example follows immediately from a standard assumption implicit in the Anscombe-Aumann model: if an objective randomization occurs after the ambiguous draw is realized, then it serves as a hedge. By contrast, no such hedging occurs if the sequence is reversed or remains unspecified, as in the original situation.

The best-known updating rules for ambiguity models are prior-by-prior updating for maxmin expected utility and the Dempster-Shafer (DS) for totally monotone capacities. To facilitate comparisons between these two alternatives and the current theory, first we will define the simple evaluation rules associated with these models and then specify the compound evaluations implied by the corresponding updating rules.

For any set of prior Δ and non-null event B, let $\Delta^B = \{p(\cdot \mid B) \mid p \in \Delta \text{ and } p(B) > 0\}$ be the set of updated priors conditional on B where $p(\cdot \mid B)$ is the conditional of p given B. Given a simple evaluation E and E-nonnull set B, the prior-by-prior updating rule is yields the conditional evaluation $E^m(\cdot \mid B)$ such that

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

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:

$$\pi^{ds}(A \mid B) = \frac{\pi(A \cup B^c) - \pi(B^c)}{1 - \pi(B^c)} \tag{8}$$

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

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

As the example in the introduction shows, prior-by-prior updating does not satisfy weak dynamic consistency. The same is true for the Dempster-Shafer rule. As an example, consider the state space $S = \{1, 2, 3, 4\}$ and assume that $\{1\}, \{4\}$ are unambiguous with $\pi(\{1\} = \pi(\{4\}) = 1/4$. By contrast, 2 and 3 are ambiguous with $\pi(\{2\}) = \pi(\{3\}) = 0$ and $\pi(\{2, 3\}) = 1/2$.[8] Let $\mathcal{P} = \{\{1, 2\}, \{3, 4\}\}$ and consider the random variable X yields

[8] Additivity holds for all other sets, that is, $\pi(\{1, 2\}) = 1/4 = \pi(\{3, 4\})$ and $\pi(\{1, 2, 4\}) = 1/2 = \pi(\{1, 3, 4\})$.

1 if $s \in \{1, 4\}$ and 0 otherwise: then $E(X) = 1/2 > E^{ds}(X|B) = 1/3$ for both cells in the information partition.

Epstein and Schneider (2003) identify a necessary and sufficient condition (rectangularity) on $\mathcal{P}$ that ensures that the law of iterated expectation is satisfied for maxmin compound evaluations with prior-by-prior updating. The proposition below shows that the same condition, applied to our setting, is also necessary and sufficient for compound evaluations with updating by proxy and for 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 $A = C$. Thus, a partition is π-extreme if each of its elements, B_i, is either unambiguous or totally ambiguous in the sense that there is a unique positive probability element of the Möbius transform of π that intersects B_i and this element contains all nonnull states of B_i. Proposition 4, below, shows that $\pi-$extremeness is necessary and sufficient for the law of iterated expectation under any of the three updating rules.

Proposition 4: *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 4 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 each 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 (2003) 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 $T := \{0,1\}^n$ be sequences of 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 $S^n = \Theta \times T$.

The decision maker observes a sequence of signals $y = (y_1, \ldots, y_n) \in T$ and draws inferences about the parameter $\theta \in \Theta$. Specifically, the decision maker's goal is to evaluate the random variable X such that $X(s) = 1$ if $s \in \{\theta_1\} \times T$ and 0 otherwise. The information partition $\mathcal{P}$ is induces by the signal; that is, $\mathcal{P} = \{\Theta \times \{y\} \mid y \in T\}$. Let $\#y = \sum_{i=1}^n y_i$.

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

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

If $\epsilon = 0$, then this capacity corresponds to the standard setting 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 familiar Bernoulli source generates the signals and with probability ϵ the signal is maximally ambiguous.

Let $E(X|y)$ be the conditional evaluation of X given $y \in \{0,1\}^n$. It is straightforward to verify that $E(X|y) = E(X|y')$ 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 $n \to \infty$.

Proposition 5: *Assume* $0 < \epsilon < 1$ *and* $\alpha := \lim_{n \to \infty} \#y/n$. *Then,*

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

To illustrate the result in Proposition 5, assume that $\theta_1 = 1/3, \theta_2 = 2/3$. If $\epsilon = 0$, then, $\lim E(X \mid y) = 0$ for $\alpha < 1/2$ and $\lim E(X \mid y) = 1$ for $\alpha > 1/2$. By contrast, if $\epsilon > 0$, then

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

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 5 to the corresponding results for prior-by-prior updating and the Dempster-Shafer rule. Let $E^m(X|y)$ and $E^{ds}(X|y)$, respectively be the conditional evaluations with prior-by-prior and Dempster-Shafer updating. As for the proxy rule, it is straightforward to verify that $E^m(X|y)$ and $E^{ds}(X|y)$ depend on y only through $\#y$.

Proposition 6: *If* $\lim \#y/n$ *exists, then* $\lim E^m(X|y) = 0$ *and* $\lim E^{ds}(X|y) = p_i$.

For large n, the Dempster-Shafer rule predicts that the agent deems the signal uninformative irrespective of its realization. The understand the reason for this result, first note that we can interpret the Dempster-Shafer rule as basing the conditional value of X on the "maximum likelihood prior" among the set of priors in the core of the capacity. For large n, the maximum likelihood prior is the one that, conditional on either value of the parameter, places an ϵ mass on the realized sequence. 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 the updated priors in the core of the capacity. One such prior assumes that, 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 X converges to zero.

In the example above, there is no prior ambiguity regarding the parameter θ; that is, $\pi(A \cup B) = \pi(A) + \pi(B)$ for $A \subset \{\theta_1\} \times T, B \subset \{\theta_2\} \times T$. In such situations, the proxy rule (like the Dempster-Shafer 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: Assume $S = \{1, 2, 3\}$, state 1 is 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 recursive value $E_{\mathcal{P}} = E(E(\cdot|\mathcal{P}))$ for this example has the following maxmin representation: let Δ_S be the set of probabilities on S and let $\Delta = \{p \in \Delta_S : p_1 = p_2, 0 \leq p_3 \leq 2/3\}$. Then, it is easy to verify that the recursive value has the following form:

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

Note that the right-hand side of the above equation does not have a Choquet integral representation. Thus, recursive evaluations encompass behavior that is more general than simple evaluations. As we show in this section, the features of this example generalize: recursive values always have a maxmin representation and, conversely, for every maxmin evaluation there is a recursive value 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 recursive values. Therefore, we extend information structures to include signals that do not correspond to a partition of the payoff relevant states.

Let the set Ω 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 E^e on $S \times \Omega$ is an *extension of* E if its capacity, π^e, satisfies $\pi^e(A \times \Omega) = \pi(A)$ for all $A \subset S$. For any random variable X on S, define the extended random variable X^e on $S \times \Omega$ as follows: $X^e(s, \omega) = X(s)$ for all $(s, \omega) \in S \times \Omega$.

Note that Ω defines a partition $\mathcal{P}_\Omega$ of $S \times \Omega$:

$$\mathcal{P}_\Omega := \{S \times \{\omega)\} \,|\, \omega \in \Omega\}$$

Let $I = \{S \times \Omega, E^e\}$ be the information space. Then, the corresponding general compound evaluation is defined as follows:

$$E_I(X) := E^e_{\mathcal{P}_\Omega}(X^e) = E^e(E^e(X^e \,|\, \mathcal{P}_\Omega))$$

where $E^e(X^e \,|\, \mathcal{P}_\Omega)$ 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 information space I.

Theorem 2 characterizes all generalized compound evaluations:

Theorem 2: *If $I = \{S \times \Omega, \pi^e\}$ is an information space for E, then there exists a compact, convex set of simple probabilities Δ, each with support contained in S_π, such that*

$$E_I(X) = \min_{p \in \Delta} \int X dp$$

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. Let $\mathcal{V} = \{X \,|\, X(s) \leq 1\}$.[9]

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

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

[9] 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. The behavior consistent with compound evaluations is more permissive and corresponds to the set of *all* maxmin evaluations.

Alternatively, we may interpret the information structure as *subjective* rather than an observable feature of the environment. In other words, the information structure may reflect the sequence in which the agent envisages the uncertainty resolving. Under this interpretation, two agents facing the same random variable may perceive different information structures. Theorems 2 and 3 then characterize the observable implications of that model.

7. Appendix

7.1 Proof of Proposition 1

Verifying that E defined by $E(X) = \int X d\pi$ for some finite-support capacity satisfies P1 and P2 straightforward 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 $0 = \alpha_0$, and let $\alpha_1 < \ldots < \alpha_k$ be values of $\mathcal{X}$ and $\mathcal{Y}$ in increasing order. 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 $E(\gamma_A) > \gamma\mu(A)$ 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 contradiction. A symmetric argument for the $E(\gamma_A) < \gamma$ cases completes the yields another contradiction. Hence, $E(\gamma_A) = \gamma\mu(A)$.

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

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

where $\alpha_i, \ldots, \alpha_n$ be all of the nonzero values that X 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 where 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 $\alpha \in (0,1]$, $K_\alpha(\mathcal{X}, A) - K_\alpha(\mathcal{Y}, A) = 1 - 0 = 1$. Verifying $K_\alpha(\mathcal{X}, A) - K_\alpha(\mathcal{Y}, A) = 0$ 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,

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

Next, consider the case where 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. $\square$

7.2 Proofs of Theorem 1, Fact 1, and Corollaries

Define $\mathcal{A}_\pi = \{A \subset S \,|\, \mu_\pi(A) \neq 0\}$.

Fact 1: The following conditions are equivalent: (i) B is unambiguous. (ii) $A_\pi(s) \subset B$ for all $s \in B$. (iii) $\pi(B) + \pi(B^c) = 1$.

Proof: 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. It is well-known 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. $\square$

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 uniqueness part of Theorem 1, we will show that $\pi^B = \pi(\cdot|B)$ where $\pi(\cdot \,|\, B)$ is as defined by (6) and (7) 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. $\square$

For Lemma 2, we consider evaluations E, E_1 and E_D be such that $E = \frac{1}{2}E_1 + \frac{1}{2}E_D$. Let π, π_1 and π_D respectively, be the capacities of E, E_1 and E_D respectively. Clearly, $\pi = \frac{1}{2}\pi_1 + \frac{1}{2}\pi_2$. 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,*

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

Proof: Lemma 1, C1 and C2 imply that there is $a_i \geq 0$ such that $a_i > 0$ if and only if $B_i\cap D \neq \emptyset$ and

$$\pi^{B_i} = \frac{a_i}{1+a_i}\pi_{B\cap D} + \frac{1}{1+a_i}\delta_i \qquad (A1)$$

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.

First, consider the case in which $B_i\cap D$ is a singleton and each B_i has the same cardinality. Then, C3 and equation (A1) imply $\pi^{B_i} = \frac{a}{1+a}\pi_{B\cap D} + \frac{1}{1+a}\delta_i$. Since $\sum_{i=1}^{k} a_i =$

k, we have $a = 1$. By C1, $\pi^B = \pi^{B'}$ if $B \cap S_\pi = B' \cap S_\pi$, and, therefore, $\pi^{B'} = \frac{1}{2}\pi_{B'\cap D} + \frac{1}{2}\delta_i$ for all B' such that $B' \cap S_\pi = \{s_i, s_i'\}$ for $s_i \in C$ and $s_i' \in D$.

Next, consider B_1 such that $|B_1 \cap D| = m$ 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 $\pi_2 = \pi_D$ and let $s \in B, s \notin D$. If $\pi = \alpha\pi_2 + (1-\alpha)\delta_s$, 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 $\pi_1 = \frac{1}{k}\sum_{s\in C}\delta_s$. Let

$$\begin{aligned}
\pi_{12} &= \frac{1}{2}\pi_1 + \frac{1}{2}\pi_2 \\
\pi_{13} &= \alpha\pi_1 + (1-\alpha)\delta_s \\
\pi_{23} &= \alpha\pi_2 + (1-\alpha)\delta_s \\
\pi_0 &= \frac{\alpha}{1+\alpha}\pi_1 + \frac{\alpha}{1+\alpha}\pi_2 + \frac{1-\alpha}{1+\alpha}\delta_s
\end{aligned}$$

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

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

By C2

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

for some $\alpha_1, \alpha_2 \in (0,1)$ and $\gamma_i \in (0,1)$ such that $\sum\gamma_i = 1$.

Note that $\{s\} \neq D \cap B$ and let $r = |D \cap B|$. The first two equations then imply that $\gamma_1 = (1-\alpha)k/((1-\alpha)r+\alpha(k+1))$ and $\gamma_3 = \alpha r/(\alpha r+\alpha(k+1))$. By C2, there is $\alpha_3 \in (0,1)$ such that

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

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

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

as desired. □

Lemma 4: *Let $\pi_i = \pi_{D_i}$ for $i = 1, \ldots, n$ 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 $|A| > 0$ and $k_i = 0$ otherwise. If B is $\pi-$nonnull, then $\pi^B = \sum_{i=1}^n \alpha_i k_i m_{A_i} / \sum_{i=1}^n \alpha_i k_i$.*

Proof: Let $B' = B \cup \{s\}$ for some $s \in B^c$. Let $\pi^* = \frac{1}{2}\sum_{i=1}^n \alpha_i \pi_i + \frac{1}{2}\delta_s$ Iterative application of C2 implies that there are $w_i, i = 1, \ldots, n$ with $\sum w_i > 0$ such that

$$\begin{aligned}\pi^{*B'} &= \beta_0 \delta_s + (1-\beta_0)\sum_{i=1}^n \frac{w_i}{\sum_{j=1}^n w_j} m_{A_i} \\ &= \beta_i \frac{\alpha_i \frac{|A_i|}{|D_i|} m_{A_i} + \delta_s}{\alpha_i \frac{|A_i|}{|D|} + 1} + (1-\beta_i)\sum_{j\neq i} \gamma_j^i m_{A_j}\end{aligned}$$

for some (γ_j^i) such that $\sum_{j\neq i} \gamma_j^i = 1$. The second equality follows from Lemma 3 and C2. First assume that for all i, j such that $A_i \neq \emptyset$, $A_i \neq A_j$. Then, straightforward algebra implies that for j such that $|A_j| > 0$ (and therefore $w_j > 0$)

$$w_i/w_j = \alpha_i \frac{|A_i|}{|D_i|} \Big/ \alpha_j \frac{|A_j|}{|D_j|} = \alpha_i k_i / \alpha_j k_j$$

as desired. If $A_i = A_j$ for some non null A_i, let $I = \{t : A_t = A_i\}$. Let $A_k \neq A_i$. Then, for $J = \{t : A_t = A_k\}$, the display equation above implies that

$$\sum_{i\in I} w_i / \sum_{j\in J} w_i = \sum_{i\in I} \alpha_i \frac{|A_i|}{|D_i|} \Big/ \sum_{j\in J} \alpha_j \frac{|A_j|}{|D_j|}$$

which completes the proof of the lemma. □

Proof of Theorem 1: To prove the uniqueness part of the Theorem, let π be the capacity of E , 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, $\mathcal{A}_\pi = \{D_1, \ldots, D_n\}$. Let $\alpha_i = \mu(D_i)$. Then, $\pi = \sum_{i=1}^n \alpha_i \pi_{D_i}$. If B is π-nonull, then, by Lemma 4, for all $C \subset B$,

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

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

To see that the proxy updating rule satisfies C1-C4, first note that C1, C2 and C3 are immediate. It remains to prove property C4. Define $E^{\mathcal{P}}$ to be the evaluation with capacity $\pi^{\mathcal{P}}$. 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, $\pi = \sum_{i=1}^n \alpha_i m_{D_i}$ 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 $E^{\mathcal{P}}(X) \geq E(X)$ for all X. Now $E^{\mathcal{P}}(X) = \sum_{B \in \mathcal{P}} \pi^{\mathcal{P}}(B) E(X|B)$ implies $E(X|B) \geq E(X)$ for some $B \in \mathcal{P}$. $\square$

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 defined as in equation (6).

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 if any $s \in S$ can be expressed in terms of the Möbius transform as follows:

$$\rho_\pi(s) = \sum_{A \ni s} \frac{\mu_\pi(A)}{|A|} \tag{A21}$$

Equation (A21) follows easily from the definition of the Shapley value and the definition of the Möbius transform. Part (i) follows from equation (A21).

Each $B \in \mathcal{P}$ is $E^{\mathcal{P}}$-unambiguous and hence, by Fact 1 and equation (A21), $\rho_\pi(B) = \pi^{\mathcal{P}}(B)$ 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 (A21) 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). $\square$

Proof of Corollary 2: We first prove the result for an elementary evaluation E_D. Let π_D be the corresponding capacity and let ρ_D such that

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

be 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 $\mathcal{P} = \{B_1, \ldots, B_k\}$, we have

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

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 $\mathcal{P} = \{B_1, \ldots, B_k\}$ and let π be the capacity for E. Then, by the linearity of the Shapley value,

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

and, therefore, the corollary follows. $\square$

7.3 Proof of Propositions 3-5

Proof of Proposition 3: First, 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, $\pi = \sum_{i=1}^n \alpha_i m_{D_i}$ for some finite collection $\{D_i\}$ and coefficients $\alpha_i \geq 0$ such that $\sum_{i=1}^n \alpha_i = 1$. Since $E = \sum \alpha_i E_{D_i}$ and $E^{\mathcal{P}} = \sum \alpha_i E_{D_i}^{\mathcal{P}}$ it follows that $E^{\mathcal{P}}(X) \geq E(X)$ for all X. Now $E^{\mathcal{P}}(X) = \sum_{B \in \mathcal{P}} \pi(B) E(X|B)$ implies $E(X|B) \geq E(X)$ for some $B \in \mathcal{P}$. $\square$

Let Δ be a nonempty, compact and convex set of probabilities with supports contained in S. For a given information partition $\mathcal{P} = \{B_1, \ldots, B_n\}$, let $\Delta_i = \{q(\cdot \,|\, B_i) \,|\, p \in \Delta, B_i \in \mathcal{P}\}$. For any $p \in \Delta$ and $q_i \in \Delta_i$ for $i = 1, \ldots, n$, let $p[q_1, \ldots, q_n] = \sum_i p(B_i) \cdot q_i$. Finally, let

$$\Delta^2 = \{p[q_1, \ldots, q_n] \,|\, p \in \Delta, q_i \in \Delta_{B_i} \text{ for all } i = 1, \ldots, n\}$$

The pair $(\Delta, \mathcal{P})$ is *rectangular* if $\Delta^2 = \Delta$. Epstein and Schneider (2003) introduce rectangularity and show that E maxmin compound evaluations with prior-by-prior updating satisfies Law of Iterated Expectation if and only if $(\Delta, \mathcal{P})$ is rectangular. In our setting, the formal statement of this result is as follows:

Lemma 6: *(i) $E_{\mathcal{P}}^m(X) \leq E(X)$ for all X. (ii) $E_{\mathcal{P}}^m(X) = E(X)$ for all X if and only $(\Delta, \mathcal{P})$ is rectangular.*

Proof: Let $\Delta := \Delta(\pi)$. Then, it is easy the verify that $E_{\mathcal{P}}^m(X) = \min_{p \in \Delta^2} X dp$. Thus, $\Delta = \Delta^2$ implies $E_{\mathcal{P}}^m(X) = E(X)$ and, observing that $\Delta \subset \Delta^2$ proves (i); to prove that $E_{\mathcal{P}}^m(X) = \min_{p \in \Delta^2} X dp$ implies that $\Delta = \Delta^2$ let $p \in \Delta^2, p \notin \Delta$. Then, there is $A \subset S$ such that $p(A) < \min_{\Delta} q(A)$ and, therefore, $E(1_A) > E_{\mathcal{P}}^m(1_A)$. $\square$

Proof of Proposition 4: 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 $A = C$. Thus, a partition is π-extreme if each of its elements, B_i, is either unambiguous or totally ambiguous in the sense that there is a unique positive probability element of the Möbius transform of π that intersects B_i and this element contains all nonnull states of B_i. To prove that the Law of Iterated Expectation holds for updating by proxy whenever it holds for prior-by-prior updating, we will show that Law of Iterated Expectation holds for prior-by-prior updating implies that $\mathcal{P}$ is π-extreme. Showing that the latter condition 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, A, C such that $\mu_\pi(A) \cdot \mu_\pi(C) > 0$, $(A \backslash C) \cap B_i \neq \emptyset$, and $(A \cup C) \cap B_i^c \neq \emptyset$. Assume $i = 1$ and let X^z be such that $X^z(s) = 1$ for $s \in B_1^c$, $X^z(s) = z$ for $s \in (C \cap B_1)$ and $X^z(s) = 0$ otherwise.

$$E(X^z) = \begin{cases} [\pi(C \cup B_1^c) - \pi(B_1^c)]z + \pi(B_1^c) & \text{if } z < 1 \\ \pi(C \cap B_1)(z-1) + \pi(C \cup B_1^c) & \text{if } z \geq 1 \end{cases} \tag{$*$}$$

Note that $\pi(C \cup B_1^c) - \pi(B_1^c) > 0$ since $\mu_\pi(C) > 0$. Observe that under all three updating rules the value of 1_C conditional on B_1 is less than 1. That is, $E(1_C|B_1) < 1, E^{ds}(1_C|B_1) < 1$ and $E^m(1_C|B_1) < 1$. This is easy to verify using the fact that $(A \backslash C) \cap B_1 \neq \emptyset$ and $\mu(A) > 0$. Moreover, it is straightforward to verify that $E(1_C|B_1) > 0$ and $E^{ds}(1_C|B_1) > 0$. If $E^m(1_C|B_1) = 0$, then, for $0 \leq z < 1$, $E^m_{\mathcal{P}}(X^z)$ does not depend on z and, therefore, (*) implies that $E^m_{\mathcal{P}}(X^z) \neq E(X^z)$ for some $z \in (0,1)$. Thus, assume that $E^m(1_C|B_1) > 0$. Let $\delta \in (0,1)$ be the conditional value of 1_C given B_1 under one of the updating rules. Then, the recursive value of X^z under that updating rule is

$$\begin{cases} [1 - \pi(B_1^c)](\delta z) + \pi(B_1^c) & \text{if } \delta z < 1 \\ \pi(B_1)(\delta z) + (1 - \pi(B_1)) & \text{if } \delta z \geq 1 \end{cases}$$

Note that $\pi(B_1) + \pi(B_1^c) < 1$ since $(A \cup B) \cap B_1^{\mathcal{C}} \neq \emptyset$. As a result, $1 - \pi(B_1^c) > \pi(B_1)$. Expression (*) and the fact that $\delta < 1$ imply that the law of iterated expectation must fail for some $z > 1$. $\square$

Proof of Proposition 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,

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

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 $\alpha > (<) r_2$. Substituting these facts and taking $n \to \infty$ then proves Proposition 5. □

7.4 Proof of Theorems 2 and 3

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

For each $\omega \in \Omega$, let $E(\cdot\,|\,\omega)$ be the conditional evaluation of $E^e_{\mathcal{P}_\Omega}$ conditional on $\{\omega\}$. By definition, these evaluation are totally monotone and hence supermodular. Then, there is a convex, compact set of probabilities Δ_ω on $S \times \{\omega\}$ such that

$$E(Z\,|\,\omega) = \min_{\tilde{p}_\omega \in \Delta_\omega} \int Z dp$$

for all random variables Z on $S \times \Omega$. Identify each $\tilde{p}_\omega \in \Delta_\omega$ with a probability p_ω on S such that $p_\omega(A) = p(A \times \{\omega\})$.

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

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

for all random variables W on Ω.

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

$$\Delta = \left\{ p = \sum_{\omega \in \Omega} q(\omega) \cdot p_\omega \,\middle|\, q \in \Delta^*, p_\omega \in \Delta_\omega \text{ for all } \omega \in \Omega \right\}$$

Clearly, for any Ω-measurable Z on $S \times \Omega$, we have

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

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

$$\begin{aligned}
E^e(E^e(X^e \,|\, \Omega)) &= \min_{q\in\Delta^*} \sum_{\omega\in\Omega} q(\omega) \min_{p_\omega\in\Delta_\omega} \sum_{s\in S} X^e(s,\omega)\tilde{p}_\omega(s,\omega) \\
&= \min_{q\in\Delta^*} \sum_{\omega\in\Omega} q(\omega) \min_{p_\omega\in\Delta_\omega} \sum_{s\in S} X(s)p_\omega(s) \\
&= \min_{p\in\Delta} \sum X(s)p(s)
\end{aligned}$$

Hence, $E_I(X) = \min_{p\in\Delta} \int X dp$ 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_I such that

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

This result will establish the Theorem for every polytope $\Delta \subset \Delta_{\hat{S}}$ with rational extreme points. Then, continuity of the mapping $(\Delta, X) \to \min_{p\in\Delta} \sum_{s\in\hat{S}} X(s)p(s)$ (when the set of all closed, convex subsets of $\Delta_{\hat{S}}$ is endowed with the Hausdorff metric, $\mathcal{V}$ is endowed with the Euclidian metric and the product of the two sets is endowed with the corresponding product metric) together with the compactness of the set $\mathcal{V}$ and the denseness of polytopes with rational extreme points in the set of all compact, convex subsets of $\Delta_{\hat{S}}$ will yield the desired result.

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

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 $\Omega = \Delta_0$ and define $A_i^e \subset \hat{S} \times \Omega$ 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

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

Then, let $\mu_\pi(A) = \frac{|\{i \mid A = A_i\}|}{n}$. Let E be the evaluation that has π as its capacity. Clearly, $I = (S \times \Omega, E^e)$ is an information space for E.

By construction, $\pi^e(\cdot \mid p_j) = p_j$ for all p_j. Note also that $\pi^e(A^e) = 0$ unless A contains some A_j^e, which means the marginal π_Ω^e of π^e on Ω satisfies $\pi_\Omega^e(C) = 0$ unless $C = \Omega$. Hence,

$$E_I(X) = \min_{p_i \in \Delta_0} \int X dp_i$$

The equation above still holds when Δ_0 is replaced with Δ, its convex hull. $\square$

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(3): 865–895

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

Epstein L. G. and Schneider M. 2003. "Recursive multiple-priors," J. Econ. Theory 113: 1–31

Epstein L. G. and Schneider M. 2007. "Learning under ambiguity," Review of Economic Studies 74: 1275–1303

Ghirardato, P., F. Maccheroni and M. Marinacci (2004) "Differentiating Ambiguity and Ambiguity Attitude's *Journal of Economic Theory,* 118, pp. 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, 1993.

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, 11441152.

Klibanoff and Hanay Updating Ambiguity Averse Preferences, The B.E. Journal of Theoretical Economics Vol. 9, 1 (Advances), Article 37 (2009).

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: 137152.

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*, 21,3,1139–54.

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, Dynamic Choice under Ambiguity, Theoretical Economics, vol. 6 n.3, September 2011

Wasserman, L.A. and Kadane J. (1990), Bayes theorem for Choquet capacities, *Annals of Statistics*, 18, 13281339.