Relative Maximum Likelihood Updating of Ambiguous Beliefs*

Xiaoyu Cheng†

June 9, 2020

Abstract

This paper proposes and axiomatizes a new updating rule: Relative Maximum Likelihood (RML) for ambiguous beliefs represented by a set of priors (C). This rule takes the form of applying Bayes' rule to a subset of the set C. This subset is a linear contraction of C towards its subset assigning a maximal probability to the observed event. The degree of contraction captures the extent of willingness to discard priors based on likelihood when updating. Two well-known updating rules of multiple priors, Full Bayesian (FB) and Maximum Likelihood (ML), are included as special cases of RML.

An axiomatic characterization of conditional preferences generated by RML updating is provided when the preferences admit Maxmin Expected Utility representations. The axiomatization relies on weakening the axioms characterizing FB and ML. The axiom characterizing ML is identified for the first time in this paper, addressing a long-standing open question in the literature.

JEL: D81, D83

Keywords: updating, ambiguity, maximum likelihood, full Bayesian, contingent reasoning, dynamic consistency

*I am grateful to Peter Klibanoff and Marciano Siniscalchi for invaluable guidance and discussions throughout the completion of this paper. I thank Eddie Dekel, Eran Hanany, Yucheng Liang, Jingyi Xue, especially Modibo Camara and Lorenzo Stanca for insightful discussions. I thank all participants at Northwestern Strategy Bag Lunch and Theory Lunch for comments. All remaining errors are my own.

†Department of Managerial Economics and Decision Sciences, Kellogg School of Management, Northwestern University, Evanston, IL, USA. E-mail: xiaoyu.cheng@kellogg.northwestern.edu

1 Introduction

For decisions under uncertainty, when information is not sufficient to pin down a unique distribution over the states, the decision maker (DM)'s revealed preference sometimes is not consistent with any single probabilistic belief, but it could be consistent with a set of priors (Ellsberg, 1961; Machina and Schmeidler, 1992). When this DM learns additional information, the updating of such a set of priors may actually involve two steps. First, she could use this information to make inferences about the plausibility of each prior and discard the ones deemed to be implausible. Second, she applies Bayes' rule to update every prior in this refined set conditional on the information received.

The two most popular updating rules for multiple priors, **Full Bayesian (FB)**[1] and **Maximum Likelihood (ML)**, are both examples of such an updating procedure. FB takes the form of applying Bayes' rule to the entire set of priors. In other words, the DM does not discard any prior under FB updating. On the other hand, under ML the DM discards all the priors that do not ascribe maximal probability (among all priors in the set) to the observed event and updates the remaining priors according to Bayes' rule. Therefore, FB and ML can be regarded as two polar extremes in terms of discarding priors based on the likelihood of the observed event in updating.

However, sometimes, a DM may not be willing to be as extreme as either case. For example, she may be willing to discard priors assigning a very small probability to the observed event, but keep priors assigning an almost maximal probability. Even though this type of intermediate updating seems to be as reasonable as the two extremes, in fact, characterizations of behaviors given by such updating are largely missing from the literature[2]. In other words, the preference behaviors corresponding to such non-extreme updating rules have not been extensively studied and identified.

In this paper, I propose a new updating rule, **Relative Maximum Likelihood (RML)** updating, which takes the form of applying Bayes' rule to a subset of the set of priors. The specific subset is determined by a linear contraction of the set of priors towards its subset assigning a maximal probability to the observed event. Consequently, RML is able to provide a means of intermediate updating between FB and ML. As the degree of contraction varies, it captures the entire range of attitudes towards discarding priors based on likelihood. As a result, FB and ML are included as extreme special cases.

I provide foundations for RML updating when the preferences admit Maxmin Expected Utility (MEU) representations (Gilboa and Schmeidler, 1989), a leading theory of decision making under ambiguity. Under MEU, the DM evaluates prospects according to the worst expected utility generated by the set of priors.

For MEU preferences, Pires (2002) characterizes FB by a simple behavioral axiom. On the other hand, Gilboa and Schmeidler (1993) axiomatizes ML when preference admit both MEU and Choquet Expected Utility (CEU) representations, which is a strict special case of the MEU preferences. Their axiomatization, however, does not extend to the more general case. In fact, there has been no axiomatization of ML under general MEU preferences in the literature. The first main result of this paper (Theorem 2.4 and more generally Theorem B.1) addresses this long-standing open question by providing a characterization of ML for MEU preferences. Moreover, this result

[1] Some may also refer to it as prior-by-prior updating

[2] See section 6 for more detailed references.

also proves to be instrumental for the characterization of RML.

RML is characterized by weakening the axioms leading to FB and ML. In turn, those axioms are relaxations of the well-known *dynamic consistency* principle[3]. The characterization results not only identify the behavioral foundations for the type of non-extreme updating specified by RML, but they also suggest the common behavioral patterns under FB and ML, two seemingly orthogonal updating rules.

Motivated by these characterizations, in this paper, I further identify a key issue in the applications of ambiguity and showcase how RML can be applied to address it.

Many recent applications of the MEU model assume the players update according to FB[4]. In settings such as mechanism design or information design, these applications find that the introduction of ambiguity is strictly beneficial for the principal. However, there is clearly a caveat that this finding may hinge on the specific assumption of FB updating. RML, as a larger family of updating rules including FB, provides a useful tool for examining this issue.

As an illustration, section 5 analyzes an example in the context of ambiguous persuasion studied by Beauchêne et al. (2019). In this example, they construct an ambiguous persuasion scheme granting the sender strictly more payoff than the optimal Bayesian persuasion under the assumption that the receiver uses FB updating. I first show that this scheme becomes strictly worse than the optimal Bayesian persuasion whenever the receiver slightly deviates from FB in the direction of ML. The level of such deviation can be easily captured by RML. This finding shows that the particular construction used by Beauchêne et al. (2019) can be non-robust with respect to updating.

Nonetheless, I construct an alternative ambiguous persuasion scheme, under which the receiver would behave exactly the same no matter what attitude he has towards discarding priors based on likelihood[5]. Such a scheme is thus *likelihood-robust*. Furthermore, I show that the sender can do strictly better using likelihood-robust ambiguous persuasion compared to Bayesian persuasion in this example. In other words, the strict gain from using ambiguous strategies does not rely on the specific assumption of FB updating in this case.

The remainder of the introduction gives the definition of RML updating and an overview of the characterization results.

1.1 Relative Maximum Likelihood Updating

Let a closed and convex[6] set C denote the set of priors over a state space Ω. For each conditioning event $E \subseteq \Omega$, let $C^*(E)$ denote the subset of the set of priors assigning maximal probability to the event E, i.e. $C^*(E) \equiv \arg\max_{p \in C} p(E)$.

Definition 1.1 (RML updating)**.** *Under RML updating, when the event E occurs, each of the priors in the set $C_\alpha(E)$ will be updated according to Bayes' rule. $C_\alpha(E)$ is defined as in the following, for some $\alpha \in [0,1]$:*

$$C_\alpha(E) \equiv \alpha C^*(E) + (1-\alpha)C = \{\alpha p + (1-\alpha)q : p \in C^*(E) \text{ and } q \in C\}.$$

[3]See Ghirardato (2002) for an example.

[4]For example, Bose and Renou (2014), Kellner and Le Quement (2018) and Beauchêne et al. (2019)

[5]This applies to RML updating, but also to any updating rule discarding priors based on likelihood. An example of such an updating rule not captured by RML is minimum likelihood updating where the receiver discard all the priors not ascribing minimum probability to the observed event.

[6]Restricting the set of priors to be closed and convex is without loss of generality under MEU preferences.

Namely, $C_\alpha(E)$ is an element-wise linear mixture of the two sets $C^*(E)$ and C. As $C^*(E) \subseteq C$, for all $\alpha \in [0,1]$ one has $C^*(E) \subseteq C_\alpha(E) \subseteq C$.

Geometrically, illustrated in Figure 1, the set $C_\alpha(E)$ is a linear contraction of the set C towards $C^*(E)$. In Figure 1, the triangle represents the simplex of probability distributions over three states ω_1, ω_2 and ω_3. The larger hexagon represents the set of priors C. For conditioning event $E = \{\omega_1, \omega_2\}$, the bottom line of the hexagon represents $C^*(E)$. The blue shaded area represents $C_\alpha(E)$ for some $\alpha \in [0,1]$.

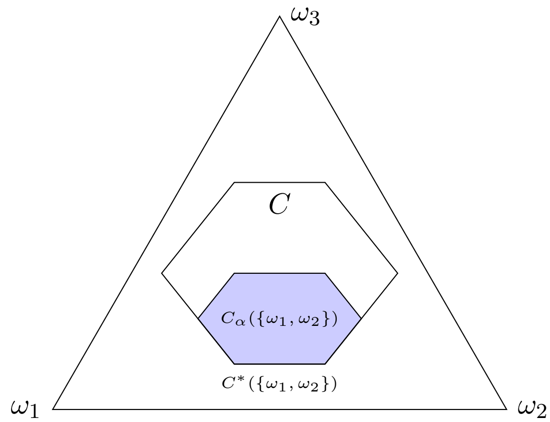

Figure 1: Graphical Illustration of $C_\alpha(E)$

Observe that when $\alpha = 0$, $C_\alpha(E)$ coincides with C and RML reduces to FB; when $\alpha = 1$, $C_\alpha(E)$ becomes $C^*(E)$ and RML reduces to ML. All the other $\alpha \in (0,1)$ capture intermediate behaviors between FB and ML. Moreover, the set $C_\alpha(E)$ shrinks as α increases, reflecting the fact that a larger α corresponds to a DM being more willing to discard priors. In this sense, the parameter α captures the inclination towards ML *relative* to FB.

To better illustrate how RML works and its relation with FB and ML, consider a version of Ellsberg's three-colored urn problem: An urn contains 30 red (R) balls and 60 other balls that can be either black (B) or yellow (Y). Nature randomly draws a ball from the urn. Let $\Omega = \{R, B, Y\}$ be the state space defined by the possible colors of the drawn ball. Suppose the DM's preference is represented by a set of priors coinciding with the information of the urn:

$$C = \{p \in \Delta(\{R, B, Y\}) : p(R) = 1/3, 0 \leq p(B) \leq 2/3\}$$

Then let $E = \{R, B\}$ be the observed event, i.e. the DM later learns that the drawn ball is not yellow. Under FB, the DM updates every prior in C conditional on E. In other words, she does not discard any prior in her initial set of priors after learning this information. Then the DM applies Bayes' rule to C to get the following set of posteriors:

$$\Pi_C \equiv \{p \in \Delta(\{R, B\}) : 1/3 \leq p(R) \leq 1\}$$

Under ML, the maximal probability of event E is ascribed by the prior $p \in C$ such that $p(B) = 2/3$:

$$C^*(E) = \{p \in \Delta(\{R, B, Y\}) : p(R) = 1/3, p(B) = 2/3\}$$

Namely, once the DM learns the drawn ball is not yellow, she further infers that there is no yellow ball in this urn and discards all the priors assigning positive probability to the event that the drawn ball is yellow. The set of posteriors under ML is then given by:

$$\Pi_{C^*(E)} \equiv \{p \in \Delta(\{R, B\}) : p(R) = 1/3\}$$

Under RML, for some $\alpha \in [0, 1]$, the updated set by definition is given by

$$\begin{aligned} C_\alpha(E) &= \alpha C^*(E) + (1-\alpha)C \\ &= \{p \in \Delta(\{R, B, Y\}) : p(R) = 1/3, 2\alpha/3 \leq p(B) \leq 2/3\} \end{aligned}$$

The set $C_\alpha(E)$ represents the DM's inference upon learning the drawn ball is not yellow as the urn is now believed to contain less than $60(1-\alpha)$ yellow balls. Notice that the maximum possible number of yellow balls in this refined set is decreasing with respect to α. The set of posteriors updated from this set is then:

$$\Pi_{C_\alpha(E)} \equiv \{p \in \Delta(\{R, B\}) : 1/3 \leq p(R) \leq 1/(2\alpha + 1)\}$$

Intuitively, given the observed event $\{R, B\}$, the DM may infer that there cannot be too many yellow balls in the urn, otherwise, such an event may not actually occur. Then a DM who is more confident about this type of inference would be willing to make a sharper conclusion about the possible number of yellow balls in the urn by discarding more priors. This results in a smaller refined set of priors. In this sense, the parameter α calibrates exactly the extent of willingness to discard priors based on likelihood.

1.2 Two Representation Theorems

The central contribution of this paper is identifying the axioms that characterize conditional preferences generated by RML updating. In particular, I provide two different representation theorems that differ in the extent to which the parameter α is allowed to vary across events (Theorem 3.3, Theorem 3.4).

Theorem 3.3 characterizes conditional preferences generated by RML updating when the parameter α is allowed to vary across events. Formally, the conditional preference upon observing event E is represented by the set of posteriors given by RML updating with parameter $\alpha[E]$. I call this form of conditional preference is represented by the **Contingent RML** updating. Namely, contingent on different events, the specific RML updating rule being applied might differ.

In contrast, Theorem 3.4 characterizes conditional preferences given by RML updating with the same parameter α across all the events. This simply requires adding an axiom to the characterization of Contingent RML updating.

Both representation theorems are important in their own right. A constant α can be more convenient for applications as it provides a sharper prediction of updating behaviors. On the other hand, a parameter $\alpha[E]$ varying across events allows for more flexibility in terms of incorporating updating behavior.

For example, some experimental results in Liang (2019) can be interpreted from the perspective of Contingent RML updating. The results in Table 4.2 therein correspond with a conditional preference represented by Contingent RML updating with $\alpha[E] < 1/2$ when $E =$"good news" and

$\alpha[E] > 1/2$ when $E =$"bad news". Details of this observation can be found in section 4. In the same section, I provide a full list of behavioral predictions that RML and Contingent RML are able to offer in this specific environment studied by Liang (2019) and some other recent experimental papers.

The remainder of this paper is organized as follows. Section 2 setups the environment and provides a characterization of ML updating. Section 3 is the main section, which gives the formal definition of preferences represented by Contingent RML and RML updating, and also provides the foundations for them. Section 4 looks at a special environment with ambiguous signals and illustrates the predictions and interpretations that Contingent RML and RML are able to offer. Section 5 applies RML to the example of ambiguous persuasion and addresses the issue of robustness to updating. Section 6 talks about a connection between RML updating and a relative likelihood ratio test and also the related literature. Section 7 concludes. All the proofs are collected in appendix A.

2 Preliminaries

2.1 Set up

Let Ω be the set of *states* with at least three elements[7], endowed with a sigma-algebra Σ of *events*. Denote a generic event by E. Let X be the set of all simple (i.e. finite-support) lotteries over a set of consequences Z and let x denote a generic element of X. Let $\mathcal{F}$ denote the set of *bounded acts*, meaning that each $f \in \mathcal{F}$ is a bounded Σ-measurable function from Ω to X. With conventional abuse of notation, denote a constant act which maps all states $\omega \in \Omega$ to x simply by x.

The primitive is a family of preferences $\{\succsim_E\}_{E\in\Sigma}$ over all acts $f \in \mathcal{F}$. Let $\succsim_\Omega \equiv \succsim$ denote the ex-ante preference. For all the other non-empty $E \in \Sigma$, let $\succsim_E$ denote the conditional preference when event E occurs.

First of all, assume that the ex-ante preference $\succsim$ admits a Maxmin Expected Utility (MEU) representation, i.e. it is represented by a closed and convex set $C \subseteq \Delta(\Omega)$ and an affine utility function u such that for all $f, g \in \mathcal{F}$:

$$f \succsim g \Leftrightarrow \min_{p\in C} \int_\Omega u(f)dp \geq \min_{p\in C} \int_\Omega u(g)dp$$

where $u(f)$ denotes a random variable $Y : \Omega \to \mathbb{R}$ such that $Y(\omega) = u(f(\omega))$ for all $\omega \in \Omega$. Denote the ex-ante evaluation of an act f according to this MEU representation by $U(f)$. In addition, assume that C is a polytope in the space $\Delta(\Omega)$ meaning it has finitely many extreme points[8].

The finitely many extreme points assumption is extremely useful as it helps simplify the statements of the axioms to a great extent meanwhile conveying the same intuitions. All characterizations can be achieved without this assumption by similar axioms as shown in appendix B. On the other hand, this assumption is also arguably reasonable as it still covers a large number of popular cases of multiple priors, for example the ϵ-contamination.

[7] When there are only two states, the conditional preferences are trivial.

[8] This assumption is studied in Siniscalchi (2006). Under his terminology, it says that the preference can be represented by finitely many plausible priors.

Assume that X is unbounded from above under the ex-ante preference[9]. Namely, $u(X) = \{u(x) : x \in X\}$ is unbounded from above. Without loss of generality, normalize $u(\cdot)$ such that $u(X) = (\underline{u}, \infty)$ for some $\underline{u} \in \mathbb{R}_{<0} \cup \{-\infty\}$.

For each $E \in \Sigma$, for any $f, h \in \mathcal{F}$, let $f_E h$ denote an act mapping all $\omega \in E$ to $f(\omega)$ and all $\omega \in E^c$ to $h(\omega)$. An event E is $\succsim$-null if for all $f, g, h \in \mathcal{F}$, one has $f_E h \sim g_E h$. Under MEU, an event E is $\succsim$-null if and only if $p(E) = 0$ for all $p \in C$. Thus, E is $\succsim$-nonnull if there exists $p \in C$ such that $p(E) > 0$.

Furthermore, define an event E to be strict $\succsim$-nonnull if for all $x, x' \in X$ such that $x \succ x'$, one has $x_E x' \succ x'$. Under MEU, an event E is strict $\succsim$-nonnull if and only if $p(E) > 0$ for all $p \in C$. Therefore, an event E is $\succsim$-nonnull but is not strict $\succsim$-nonnull if there exists $p, p' \in C$ such that $p(E) = 0$ and $p'(E) > 0$.

For each $\succsim$-nonnull $E \in \Sigma$, assume that the conditional preference $\succsim_E$ also admits a MEU representation. It is represented by a set $C_E \subseteq \Delta(\Omega)$ and the same utility function u such that for all $f, g \in \mathcal{F}$:

$$f \succsim_E g \Leftrightarrow \min_{p \in C_E} \int_\Omega u(f) dp \geq \min_{p \in C_E} \int_\Omega u(g) dp$$

Meanwhile, the conditional preference $\succsim_E$ for $\succsim$-null E is unrestricted.

Finally, assume that the conditional preferences satisfy *consequentialism*: first, for all $p \in C_E$, $p(E) = 1$, i.e. the complement of the conditioning event is irrelevant for the conditional preference; and second, ex-ante preference and conditioning event E completely determine the conditional preference $\succsim_E$, which rules out the possibility that the updating rule may depend on the context of decision problems (e.g. the menu of available acts).

All properties assumed for the primitive have axiomatize foundations:

- Maxmin Expected Utility representation: axioms from Gilboa and Schmeidler (1989).
- C has finitely many extreme points: no local hedging axiom from Siniscalchi (2006).
- X is unbounded from above: unboundedness axiom from Maccheroni et al. (2006)
- u is independent of E: state independence axiom from Pires (2002) or unchanged tastes axiom from Hanany and Klibanoff (2007).
- Consequentialism: null complement axiom and the independence axioms in section 3.2 of Hanany and Klibanoff (2007).

Setting up in this way allows me to establish a clear equivalence between the key axioms relating the conditional preferences to the ex-ante preference and the updating rules. Hence the key behavioral foundations characterizing different updating rules under MEU preferences can be highlighted in the representation theorems.

[9] It suffices to assume X to be unbounded, however if X is unbounded from below, the axioms need to modified in order to accommodate such a case. For simplicity, this paper assumes it away.

2.2 Full Bayesian Updating

For dynamic choices, the well-known dynamic consistency principle relate the conditional preferences to the ex-ante preference in two different directions (see Ghirardato (2002) for an example). To further distinguish between these two directions, they will be referred to separately as **Contingent Reasoning (CR)** and **Dynamic Consistency (DC)** in this paper[10]:

Axiom CR (**C**ontingent **R**easoning).
For all $f, h \in \mathcal{F}$ and $x \in X$, if $f \sim_E x$ then $f_E h \sim x_E h$.

Axiom DC (**D**ynamic **C**onsistency).
For all $f, h \in \mathcal{F}$ and $x \in X$, if $f_E h \sim x_E h$ then $f \sim_E x$.

Intuitively, CR emphasizes that the ex-ante preference could be recovered by considering conditional preferences contingent on whether or not the event E occurs. Namely, the ex-ante comparison between the acts $f_E h$ and $x_E h$ could be done by the following procedure: contingent on event E occurring, the DM is conditionally indifferent. On the other hand, contingent on E not occurring, she receives exactly the same thing. Thus, she further concludes that she is indifferent ex-ante.

DC emphasizes the other direction, that the conditional preference should be consistent with the ex-ante preference when it does not matter what pays on the event E^c. In other words, if two acts only differ within the event E, then the DM's ex-ante comparison should remain the same after learning event E occurring.

As MEU preferences violate Savage's P2 (sure thing principle), whether or not the statement $f_E h \sim x_E h$ is true would also depend on the act h. As a result, it is no longer meaningful to state CR and DC in terms of all the acts, as a different h may result in a different conclusion. Instead, the claims about CR and DC should also specify under which act it holds.

For example, a consequence $x \in X$ is said to be the **conditional certainty equivalent** of an act f given event E if $x \sim_E f$. Consider the following statement of CR, in which the act h is explicitly fixed to be the conditional certainty equivalent:

Axiom CR-C (**C**ontingent **R**easoning given **C**onditional certainty equivalent).
For all $f \in \mathcal{F}$ and $x \in X$, if $f \sim_E x$, then $f_E x \sim x$.

CR-C is exactly the A9 axiom in Pires (2002) and FB updating is defined in Pires (2002) by the following[11]:

Definition 2.1 (FB). *The primitive $\{\succsim_E\}_{E \in \Sigma}$ is represented by FB updating if the following holds*

[10] The statements of the axioms are equivalent to the statements of dynamic consistency adopted by Ghirardato (2002).

[11] The definition of FB restricts behaviors only for strict $\succsim$-nonnull events. It is because for events that are $\succsim$-nonnull but not strict $\succsim$-nonnull, there exists $p \in C$ such that $p(E) = 0$, which cannot be updated via Bayes' rule. Pires (2002) shows, in those cases, her axioms are not able to pin down the exact conditional preferences. Notice that, such a problem does not apply to RML when $\alpha \neq 0$, since $p(E) > 0$ for all $p \in C_\alpha(E)$ even when E is not strict $\succsim$-nonnull. Yet for uniformity, I impose the same restriction to all characterizations.

for all strict $\succsim$-nonnull $E \in \Sigma$ and for all $f \in \mathcal{F}$:

$$\min_{p \in C_E} \int_\Omega u(f) dp = \min_{p \in C} \int_E u(f) \frac{dp}{p(E)}.$$

Under the current framework, her representation theorem can be simplified to the following:

Theorem 2.2 (Pires, 2002)**.** *$\{\succsim_E\}_{E \in \Sigma}$ is represented by FB updating if and only if CR-C holds for all strict $\succsim$-nonnull events $E \in \Sigma$.*

This theorem highlights that CR-C is the key behavioral foundation for FB updating under MEU preferences.

2.3 Maximum Likelihood Updating

Another popular updating rule for MEU preferences is the Maximum Likelihood (ML) updating rule proposed in Gilboa and Schmeidler (1993). The conditional preferences represented by ML updating are defined as in the following:

Definition 2.3 (ML)**.** *The primitive $\{\succsim_E\}_{E \in \Sigma}$ is represented by ML updating if the following holds for all strict $\succsim$-nonnull $E \in \Sigma$ and for all $f \in \mathcal{F}$:*

$$\min_{p \in C_E} \int_\Omega u(f) dp = \min_{p \in C^*(E)} \int_E u(f) \frac{dp}{p(E)}$$

where $C^*(E) = \arg\max_{p \in C} p(E)$.

Namely, the conditional preference $\succsim_E$ for strict $\succsim$-nonnull E is represented by the set of posteriors that were updated from the subset of the set of priors assigning a maximal probability to the observed event E.

Gilboa and Schmeidler (1993) provide an axiomatic characterization of ML updating when the preferences admit both MEU and CEU representations. This class of preferences is a strict subset of preferences admitting MEU representations.

In that case, it is without loss of generality to let X be bounded and denote the best consequence in X by x^*. Gilboa and Schmeidler (1993) show that ML is characterized by the following axiom:

Axiom CR-B (**C**ontingent **R**easoning given **B**est consequence).
For all $f \in \mathcal{F}$ and $x \in X$, if $f \sim_E x$, then $f_E x^* \sim x_E x^*$.

CR-B claims that, the DM's ex-ante evaluation of the act $f_E x^*$ can be recovered by applying contingent reasoning with respect to the event E. Gilboa and Schmeidler (1993) offer a "pessimistic" interpretation of this behavior: the DM's conditional preference of an act f comes from the consideration that the best consequence would have been received had the complement event happened. In other words, her conditional evaluation of the act f reflects a disappointment that event E actually occurs. Under MEU preferences, the best consequence is believed to be received only with the minimum probability ex-ante. Consequently, the disappointment reflected in the conditional preference is in accordance with such an ex-ante belief.

As one enlarges the domain of preferences to the more general case of MEU. Although this intuition may sill seem to be applicable, the exact intuition captured by CR-B becomes imprecise. The following example illustrates a case where the ex-ante preference admits MEU but not CEU representation, and CR-B is violated under ML updating:

Example 1. *Consider a state space $\Omega = \{\omega_1, \omega_2, \omega_3\}$, denote any $p \in \Delta(\Omega)$ by a vector with three coordinates $p = (p_1, p_2, p_3)$ such that $p_1 + p_2 + p_3 = 1$. Let set $C \subseteq \Delta(\Omega)$ be the closed convex hull of the following three points: $(1/2, 0, 1/2)$, $(0, 1/2, 1/2)$ and $(1/3, 1/3, 1/3)$.*

Let $X = [0, 1]$ with utility function $u(x) = x$. Consider an act f pays one at ω_1, zero at ω_2 and is undetermined on state ω_3. Let the conditioning event E be the following set: $\{\omega_1, \omega_2\}$.

First of all, one can verify that a MEU preference represented by this C does not admit a CEU representation[12]. If the conditional preference is given by ML updating, as $C^*(E)$ contains only the extreme point $(1/3, 1/3, 1/3)$, it implies that

$$f \sim_E 1/2$$

CR-B claims that $f_E x^*$ should be ex-ante indifferent to $1/2_E x^*$ for $x^* = 1$. However, it is actually the case that

$$f_E 1 \prec 1/2_E 1$$

i.e. CR-B is false.

One might notice that the breakdown of CR-B results from the fact that $f_E 1$ and $1/2_E 1$ are evaluated at different extreme points under MEU. Especially, the act $f_E 1$ is not evaluated at the extreme point $(1/3, 1/3, 1/3)$. Namely, her ex-ante evaluation of the act $f_E 1$ is not given by the prior assigning a maximal probability to the event E. Moreover, the graphical illustration given in figure 2 suggests that it is exactly because the consequence $x^* = 1$ is not good enough.

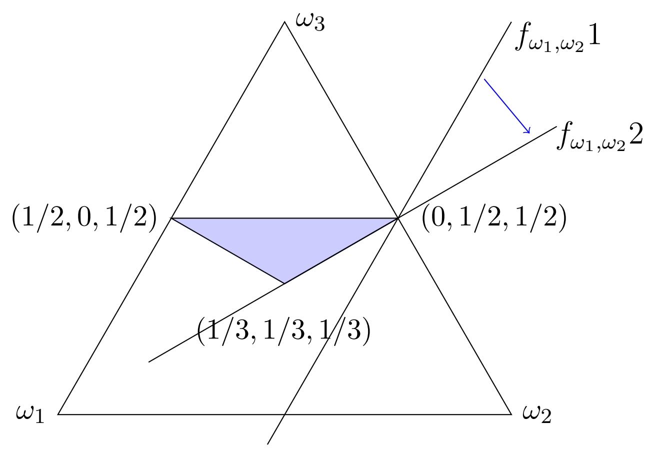

Figure 2: Graphical Illustration of Example 1

In figure 2, the arrow indicates how the indifference curve of act $f_{\omega_1,\omega_2} x$ under expected utility would be changing (in angle) if one increases the value of consequence x. When $x = 1$,

[12]This C is not a core of any convex capacity.

the act $f_{\omega_1,\omega_2}1$ is evaluated at the extreme point $(0, 1/2, 1/2)$ according to MEU. When $x = 2$, the act $f_{\omega_1,\omega_2}2$ is evaluated at the segment between the two extreme points $(1/3, 1/3, 1/3)$ and $(0, 1/2, 1/2)$. If one keeps increasing x, apparently for all $x \geq 2$, the act $f_{\omega_1,\omega_2}x$ will be evaluated at the extreme point $(1/3, 1/3, 1/3)$ (When x goes to infinity, the act eventually will be parallel to the bottom line of this triangle).

Therefore, for this act f and conditioning event $E = \{\omega_1, \omega_2\}$, the consequence x paid on the event E^c needs to be at least better than $x = 2$ for the ex-ante evaluation of it to be given by some prior assigning a maximal probability to E. Consequently, the assumption that X is bounded by $x^* = 1$ would prevent that from happening. In fact, one can also verify that $f_E x \sim 1/2_E x$ for all $x \geq 2$ in this example.

In summary, Gilboa and Schmeidler (1993)'s intuition about the behavioral foundation of ML updating is qualitatively correct, i.e. the conditional evaluation of an act f given by ML reflects a disappointment that event E occurs while some sufficiently good consequence would have been paid had event E not occur. However, under general MEU, the best consequence x^* sometimes is not good enough. Example 1 shows exactly a case where the threshold $(x = 2)$ for sufficiently good consequence is strictly better than the upper bound $(x^* = 1)$. Furthermore, notice that this threshold depends on the act being considered as well. Obviously, when f pays 2 at ω_1 and 0 at ω_2, the threshold will become $x = 4$. Hence, to accommodate all these findings, X needs to be unbounded from above.

In fact, the behavioral foundation of ML updating under MEU is more precisely that for all acts there exists a threshold of sufficiently good consequence such that CR holds. Formally,

Axiom CR-S (**C**ontingent **R**easoning given **S**ufficiently good consequences).
For all $f \in \mathcal{F}$ there exists $\bar{x}_{E,f} \in X$ such that, for all $x, x^* \in X$ with $x^* \succsim \bar{x}_{E,f}$, if $f \sim_E x$, then $f_E x^* \sim x_E x^*$.

For the necessity of CR-S under ML updating, the existence of the threshold $\bar{x}_{E,f}$ for every E and f is guaranteed by assumptions on the primitives: X is unbounded from above and C has finitely many extreme points.

The reason why X being unbounded from above is necessary has been explained. For the other assumption, when C contains infinitely many extreme points, for example, it is a circle in the three-dimensional simplex. Consider the act f in example 1. $f_E x$ is evaluated at the prior assigning a maximal probability to the event E only if x goes to infinity. In that case, a finite threshold $\bar{x}_{E,f}$ will not exist. In contrast, when C contains finitely many extreme points, such a threshold always exists[13].

The following theorem shows that CR-S is the key behavioral foundation for ML updating under MEU:

Theorem 2.4 (ML). *$\{\succsim_E\}_{E\in\Sigma}$ is represented by ML updating if and only if CR-S holds for all strict $\succsim$-nonnull events $E \in \Sigma$.*

In appendix B, I provide an axiomatization of ML without the finitely many extreme points assumption. The alternative axiom there accommodates the possibility that a finite threshold $\bar{x}_{E,f}$

[13]Proved by Lemma A.3.

may not exist. In the case of non-existence, it suffices to consider a sequence of increasing thresholds such that the difference between $f_E x^*$ and $x_E x^*$ vanishes for all x^* better than the threshold. The formal characterization is given by Theorem B.1.

3 Relative Maximum Likelihood Updating

3.1 Representations

Recall Definition 1.1, RML updating applies Bayes' rule to the following subset of the set of priors C for some $\alpha \in [0, 1]$:

$$C_\alpha(E) \equiv \alpha C^*(E) + (1-\alpha)C = \{\alpha p + (1-\alpha)q : p \in C^*(E) \text{ and } q \in C\}$$

As RML features a parameter representing the extent of inclination towards ML relative to FB, it creates the possibility that a DM may not have the same inclination across different events. However, this is impossible under either FB or ML, as the parameters in those cases, by definition, are fixed to be a constant ($\alpha \equiv 0$ or 1).

For this reason, two definitions for conditional preferences generated by RML updating differ in the extent to which the parameter α is allowed to vary across events are provided. The weaker definition captures the case in which the conditional preferences are updated by RML with parameter $\alpha[E]$. In this case, $\alpha[E]$ explicitly depends on the observed event E capturing the different inclinations across events. Such conditional preferences are defined to be represented by the Contingent RML updating:

Definition 3.1 (Contingent RML)**.** *The primitive $\{\succsim_E\}_{E\in\Sigma}$ is represented by Contingent RML updating if for all strict $\succsim$-nonnull $E \in \Sigma$ there exists $\alpha[E] \in [0, 1]$ such that for all $f \in \mathcal{F}$:*

$$\min_{p\in C_E} \int_\Omega u(f)dp = \min_{p\in C_{\alpha[E]}(E)} \int_E u(f)\frac{dp}{p(E)}$$

where $C_{\alpha[E]}(E) \equiv \alpha[E]C^*(E) + (1-\alpha[E])C$.

By definition, Contingent RML does not require a DM to have a consistent updating rule across events. For example, it allows for the following updating behavior. A DM updates her belief according to FB updating ($\alpha[E] = 0$) for some event E. Meanwhile, she updates according to ML updating ($\alpha[E'] = 1$) for another event E'.

Such inconsistency may be caused by the different maximal or minimal probability of events, different contexts, or even just different labels of events. Although one can argue whether this kind of consistency in updating is desirable or not, by offering such flexibility, Contingent RML is able to accommodate a broader class of behaviors. Particularly, some experimental evidence also reflects such inconsistency[14].

On the other hand, undoubtedly, Contingent RML may not be strong enough in applications when sharper predictions of a DM's updating behaviors are needed. For example, in the case where

[14] See discussion in section 4.

a DM's conditional preference is observed only under one of the events. Without a constant parameter α, the analyst cannot draw any meaningful conclusion about this DM's updating behavior under the other events.

Therefore, the stronger definition captures exactly the case in which the conditional preferences are updated from RML with the same parameter α across all the events. Such conditional preferences will simply be defined to be represented by RML updating:

Definition 3.2 (RML)**.** *The primitive $\{\succsim_E\}_{E\in\Sigma}$ is represented by RML updating if there exists $\alpha \in [0,1]$ such that for all strict $\succsim$-nonnull $E \in \Sigma$ and for all $f \in \mathcal{F}$:*

$$\min_{p\in C_E} \int_\Omega u(f)dp = \min_{p\in C_\alpha(E)} \int_E u(f)\frac{dp}{p(E)}$$

where $C_\alpha(E) \equiv \alpha C^*(E) + (1-\alpha)C$.

3.2 Preference Foundation

Clearly, RML is a more general class of updating rules that includes both FB and ML as special cases. Hence, the set of conditional preferences generated by RML is also a superset of the conditional preferences admitting FB or ML updating. Consequently, the axioms characterizing FB or ML may sometimes be violated by this larger set of conditional preferences. The characterization provided in this paper help pin down exactly the relaxations of those axioms to accommodate these behaviors.

Recall that FB is characterized by CR-C ($f \sim_E x$ implies $f_E x \sim x$). Meanwhile, it is immediate that FB also satisfies DC-C: $f_E x \sim x$ implies $f \sim_E x$. Moreover, DC-C can be equivalently written as[15]:

Axiom DC-C (**D**ynamic **C**onsistency given **C**onditional certainty equivalent).
For all $f, g \in \mathcal{F}$ and for all $x \in X$ with $x \sim_E f$, if $f_E x \sim g_E x$, then $f \sim_E g$.

Notice that the quantifier $x \sim_E f$ is simply a definition of the constant act x needed for the statement of this axiom. The existence of such an x for any act f under the conditional preference is guaranteed by the MEU representation.

Similarly, ML is characterized by CR-S ($f \sim_E x$ implies $f_E x^* \sim x_E x^*$ for all $x^* \succsim \bar{x}_{E,f}$). It also satisfies DC-S: $f_E x^* \sim x_E x^*$ for all $x^* \succsim \bar{x}_{E,f}$ implies $f \sim_E x$, which can further be equivalently written as:

Axiom DC-S (**D**ynamic **C**onsistency given **S**ufficiently good consequences).
For all $f, g \in \mathcal{F}$ there exists $\bar{x}_{E,f,g} \in X$ such that, for all $x^* \in X$ with $x^* \succsim \bar{x}_{E,f,g}$, if $f_E x^* \sim g_E x^*$, then $f \sim_E g$.

[15]To see why they are equivalent, the necessity of the original statement of DC-C is immediate by letting f to be a constant act, i.e. $f = x$. For sufficiency, first notice that for each f there exists x_f such that $f_E x_f \sim x_f$ holds: There exists $\underline{x}$ and $\overline{x}$ such that $f_E \underline{x} \succsim \underline{x}$ and $f_E \overline{x} \prec \overline{x}$, meanwhile $U(f_E x)$ is continuous and increasing in x. Then the original DC-C implies $f \sim_E x_f$. For any g with $f_E x_f \sim g_E x_f \sim x_f$, the last indifference further implies $g \sim_E x_f$ by the original DC-C.

The characterization of conditional preferences generated by RML updating relies on weakening exactly these four axioms: CR-C, CR-S, DC-C, and DC-S. In other words, all the axioms characterizing RML restrict behaviors only when the act pays on the complement event E^c is either the conditional certainty equivalent or any of the sufficiently good consequences.

Consider the following two axioms:

Axiom CR-UO (**C**ontingent **R**easoning with **U**ndershooting and **O**vershooting).

For all $f \in \mathcal{F}$ and $x \in X$ there exists $\bar{x}_{E,f} \in X$ such that, for all $x^* \in X$ with $x^* \succsim \bar{x}_{E,f}$, if $f \sim_E x$, then (i) $f_E x \precsim x$ and (ii) $f_E x^* \succsim x_E x^*$.

Axiom DC-CS (**D**ynamic **C**onsistency given **C**onditional certainty equivalent and **S**ufficient good consequences).

For all $f, g \in \mathcal{F}$ there exists $\bar{x}_{E,f,g} \in X$ such that, for all $x, x^* \in X$ with $x \sim_E f$ and $x^* \succsim \bar{x}_{E,f,g}$, if (i) $f_E x \sim g_E x$ and (ii) $f_E x^* \sim g_E x^*$, then $f \sim_E g$.

Notice that CR-UO is a relaxation of CR-C and CR-S, whereas DC-CS is a relaxation of DC-C and DC-S. However, the relaxations take different forms.

CR-UO keeps the common premise of CR-C and CR-S ($f \sim_E x$) but relaxes the conclusion of the two axioms. Instead of implying indifferences, CR-UO claims that only weak inequalities in term of the ex-ante preference can be concluded. Namely, the ex-ante preference recovered by CR is either undershooting or overshooting.

DC-CS keeps the common conclusion of DC-C and DC-S ($f \sim_E g$) but strengthens the premise by requiring the premises of both DC-C and DC-S to hold. In other words, DC can be concluded only when both ex-ante indifferences hold, which is a stronger requirement than either DC-C or DC-S.

Before providing intuitions for the two axioms, the following theorem claims that these two axioms are indeed the key behavioral foundations for Contingent RML updating (as none of the axioms restrict behaviors across events):

Theorem 3.3 (Contingent RML)**.** *$\{\succsim_E\}_{E\in\Sigma}$ is represented by Contingent RML if and only if CR-UO and DC-CS hold for all strict $\succsim$-nonnull events $E \in \Sigma$. Furthermore for every such E, $\alpha[E]$ is unique if there exists $p, p' \in C$ s.t. $p(E) \neq p'(E)$.*

For a very high-level intuition, CR-UO suggests that the updating rule needs to update some intermediate set $C(E)$ such that $C^*(E) \subseteq C(E) \subseteq C$. Given this restriction, DC-CS further pins down the set $C(E)$ to be exactly the functional form given by RML with some $\alpha[E]$.

To be more specific, notice that if the DM does not want to behave as extreme as either FB or ML and update some set $C(E)$ such that $C^*(E) \subseteq C(E) \subseteq C$. Her conditional evaluation of any act f would be weakly higher than under FB and weakly lower than under ML. Let x, x^{FB} and x^{ML} denote the conditional certainty equivalent of an act f under different updating rules, then one has $x^{ML} \succsim x \succsim x^{FB}$.

First, notice that CR-C implies $f_E x^{FB} \sim x^{FB}$. If one replaces x^{FB} by x on both sides of this indifference, $x \succsim x^{FB}$ would further imply $f_E x \precsim x$. This is because the better consequence x is received only on event E^c on the left hand side, yet it will be received in all states on the right hand side. Thus the increase in terms of utility is higher on the right hand side, i.e. $f_E x \precsim x$.

Intuitively, this DM's ex-ante comparison between the acts $f_E x$ and x can no longer be recovered exactly from contingent reasoning. Instead, because she discards some priors for conditional evaluation of the act f, her ex-ante evaluation of $f_E x$ would be **undershooting**, meaning that she is now only confident in the fact that $f_E x$ should be weakly worse than x.

On the other hand, as CR-S implies $f_E x^* \sim x^{ML}_E x^*$ for all sufficiently large x^*, replacing x^{ML} by the worse consequence x would imply that $f_E x^* \succsim x_E x^*$. Similarly, the DM's ex-ante comparison between $f_E x^*$ and $x_E x^*$ cannot be recovered exactly by contingent reasoning as well. Since now her conditional evaluation of the act f depends on more priors than under ML, her ex-ante evaluation of $f_E x^*$ would be **overshooting**. It means that instead of knowing that she should be indifferent, now she is only confident that $f_E x^*$ should be weakly better than $x_E x^*$.

From the above discussion, CR-UO captures exactly the intuition about a DM who is not willing to be as extreme as either FB or ML. However, the axiom per se does not bring anything more than that. Actually, it is not hard to show that the CR-UO axiom is equivalent to updating any intermediate set $C(E)$[16]. Therefore, to further pin down a specific set to represent the DM's conditional preference, more restrictions need to impose on the DM's behaviors.

DC-CS restricts the DM's conditional preferences by imposing DC when the ex-ante preference satisfies both $f_E x \sim g_E x$ and $f_E x^* \sim g_E x^*$. Notice that, under MEU preferences, these two indifferences do not always hold at the same time. Instead, it is actually quite demanding for two acts to be indifferent in both cases. In other words, under DC-CS, the DM concludes that she is conditionally indifferent between the two acts f and g, only if they are sufficiently similar in the sense that she finds f and g to be indifferent in both types of ex-ante comparisons.

Therefore, DC-CS itself is not a strong requirement on preferences, since it restricts behaviors only when this rather strong premise is true. Nonetheless, it is sufficient to pin down the functional form of conditional preferences, which is given exactly by the Contingent RML.

In the following, I am going to provide a sketch of the proof of this result:

Sketch of proof of Theorem 3.3. For sufficiency of CR-UO and DC-CS, fix any strict $\succsim$-nonnull $E \in \Sigma$, the proof proceeds by the following four steps:

Step 1. For any $f \in \mathcal{F}$, find any x^* that is better than the following three thresholds: (i) x, (ii) $\bar{x}_{E,f}$ and (iii) $\hat{x}_{E,f}$, where $\hat{x}_{E,f}$ denotes the consequence such that for all $x^* \succsim \hat{x}_{E,f}$ one has

$$\min_{p \in C} \int_\Omega u(f_E x^*) dp = \min_{p \in C^*(E)} \int_\Omega u(f_E x^*) dp$$

Fix such an x^*, CR-UO implies that $f_E x^* \succsim x_E x^* \succsim x \succsim f_E x$. Then it can be easily seen from Figure 3 that there always exists an $\alpha[E, f] \in [0, 1]$ such that the following equation holds:

$$\alpha[E, f] U(f_E x^*) + (1 - \alpha[E, f]) U(f_E x) = \alpha[E, f] U(x_E x^*) + (1 - \alpha[E, f]) U(x) \tag{3.1}$$

Furthermore, this $\alpha[E, f]$ is unique if either $f_E x \prec x$ or $f_E x^* \succ x_E x^*$ holds.

[16] The proof of this statement relies on a separating hyperplane argument, similar to the proof of Proposition 12 in Hanany and Klibanoff (2007).

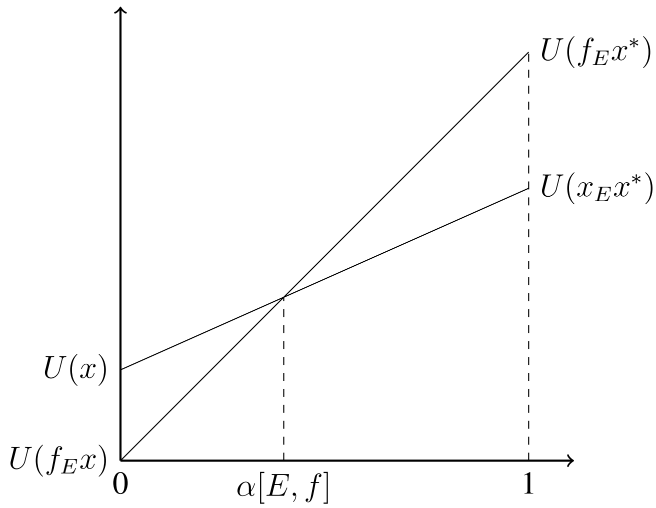

Figure 3: Graphical Illustration of Step 1

Step 2. Fix any $f \in \mathcal{F}$ such that $\alpha[E, f]$ is unique (if it does not exist then $\alpha[E]$ is not unique). First, show that for the act $f_\lambda y = \lambda f + (1 - \lambda) y$ with any $\lambda \in (0, 1]$ and $y \in X$, one has $\alpha[E, f_\lambda y] = \alpha[E, f]$. Namely, $\alpha[E, f]$ is unchanged under the transformation of f by taking mixtures with constant acts. This is given by equation (3.1) and the certainty independence axiom in Gilboa and Schmeidler (1989).

Then for any act $g \in \mathcal{F}$ that cannot be obtained by taking mixtures of f with constant acts, show that there always exists such transformations of f and g that the pair of transformed acts satisfies the premises of DC-CS.

Step 3. Given the pair of transformed acts satisfying the premises of DC-CS. DC-CS further implies that $\alpha[E, f]$ needs to be equal to $\alpha[E, g]$. Since the construction in step 2 works for arbitrary $g \in \mathcal{F}$, it thus concludes that $\alpha[E, f]$ needs to be a constant across all $f \in \mathcal{F}$. In other words, equation (3.1) holds with $\alpha[E]$ independent of f:

$$\alpha[E]U(f_E x^*) + (1 - \alpha[E])U(x) = \alpha[E]U(x_E x^*) + (1 - \alpha[E])U(x) \tag{3.2}$$

Step 4. By plugging into each term in equation (3.2), the DM's conditional evaluation of any act $f \in \mathcal{F}$ can be shown to be represented by

$$\min_{p \in C_{\alpha[E]}(E)} \int_E u(f) \frac{dp}{p(E)}$$

i.e. it is represented by RML updating with $\alpha[E]$.

□

Theorem 3.3 shows that CR-UO and DC-CS characterize conditional preferences that are given by the Contingent RML updating rule where the parameter $\alpha[E]$ may be different across events. Clearly, it is because both CR-UO and DC-CS restrict behaviors only within each conditioning event E, yet does not explicitly restrict behaviors across events. Notice that imposing either CR-C or CR-S for all events actually also restricts behaviors across events, since the resulting updating rule takes the form $\alpha[E] \equiv 0$ or $\alpha[E] \equiv 1$. However, the two axioms characterizing the Contingent RML do not have this additional power.

To further restrict behaviors across events, consider the following axiom. It features a similar statement as DC-CS but across different events E_1 and E_2:

Axiom EC (**E**vent **C**onsistency).
For all $f, g \in \mathcal{F}$ and $E_1, E_2 \in \Sigma$ there exists $\bar{x}_{E_1,E_2,f,g} \in X$ such that, for all $x, x_1^*, x_2^* \in X$ with $x \sim_{E_1} f$, $x_1^* \succsim \bar{x}_{E_1,E_2,f,g}$, $x_2^* \succsim \bar{x}_{E_1,E_2,f,g}$ and $x_{E_1}x_1^* \sim x_{E_2}x_2^*$, if (i) $f_{E_1}x \sim g_{E_2}x$ and (ii) $f_{E_1}x_1^* \sim g_{E_2}x_2^*$, then $g \sim_{E_2} x$.

First, notice that EC is also a relaxation of DC-C and DC-S, since under DC-C, the first premise implies $g \sim_{E_2} x$, and under DC-S, the second premise implies $g \sim_{E_2} x$. It is a relaxation of DC-C and DC-S different from the relaxation in DC-CS. In particular, EC and DC-CS are orthogonal in the sense that they do not imply each other.

EC, as suggested by its name, uses constant acts to calibrate the conditional preferences. More specifically, two acts have the same conditional certainty equivalent, one under events E_1 and the other under E_2, if they are indifferent under the ex-ante preference when some specific act is paid on E_1^c and E_2^c respectively. The behavioral intuition of DC-CS can be adapted here to provide a similar explanation for this axiom.

The following theorem shows that EC achieves exactly its goal: calibrating a constant α to represent conditional preferences that are given by RML updating:

Theorem 3.4 (RML). *$\{\succsim_E\}_{E\in\Sigma}$ is represented by RML if and only if CR-UO, DC-CS and EC hold for all strict $\succsim$-nonnull events $E \in \Sigma$. Furthermore, α is unique if there exists such an E that $p(E) \neq p'(E)$ for some $p, p' \in C$.*

4 Updating Ambiguous Signals

In this paper, the state space Ω is the grand state space, which could include both payoff-relevant states and payoff-irrelevant signals. To facilitate discussions, in this section, Ω will be explicitly written as the Cartesian product of states and signals: $\Theta \times S$ with generic element $\theta \times s$. Then the event in which signal s realizes is given by $E = \Theta \times s$.

Given a set of priors $C \subseteq \Delta(\Theta \times S)$, its marginal distribution over the states[17] Θ is ambiguous if there exists $p, p' \in C$ such that $p(A \times S) \neq p'(A \times S)$ for some $A \times S \in \Sigma$. Conversely, it is probabilistic over the states if $p(A \times S) = p'(A \times S)$ for all $A \times S \in \Sigma$ and $p, p' \in C$. In the latter case, ambiguity could only arise through the signals. This section illustrates the implications of Contingent RML and RML in this special environment.

Though special, this environment has been adopted in various applications, for example, Bose and Renou (2014), Beauchêne et al. (2019), Kellner and Le Quement (2018). In addition, several recent experimental papers also focus on understanding subjects' response to ambiguous signals when

[17] I will refer payoff-relevant states as simply states hereafter, as it should cause no confusion.

there is no other source of ambiguity, for example, Liang (2019), Shishkin and Ortoleva (2019), Kellner et al. (2019). Therefore, understanding the implications of RML in this specific environment will be helpful not only for the comparison with other updating rules, but also for illustrating the theoretical contributions of RML for future applications.

In the following, I will use a stylized example to illustrate the theoretical predictions RML is able to offer here.

Example 2. *There are two payoff-relevant states: $\Theta = \{\theta_1, \theta_2\}$ with unambiguous marginal prior $p(\theta_1) = \beta \in [0, 1]$. The DM is evaluating a bet f that pays one at θ_1 and nothing at θ_2. There are two signals: $S = \{s_1, s_2\}$ and the signaling structure is ambiguous, meaning that there are two possible correlations between signals and states:*

$$p(s_i|\theta_i) = \lambda_1$$
$$p(s_i|\theta_i) = \lambda_2$$

and which correlation generates the signal is unknown. Without loss of generality, assume $(\lambda_1 + \lambda_2)/2 \geq 1/2$ and $\lambda_1 \geq \lambda_2$. Namely, the signal s_1 is "on average" an informative signal for the state θ_1. Moreover, the first signaling device governed by λ_1 is more accurate than λ_2.

Suppose the DM's preference is represented by a set of priors coinciding exactly with the information given in this example. Her set of priors C can be easily parameterized by a parameter $\mu \in [0, 1]$ denoting the probability that the first signaling device governed by λ_1 is the one generating the signals. Namely, $C = \{p_\mu \in \Delta(\Theta \times S) : \mu \in [0, 1]\}$ where p_μ is defined as in the following:

$$\begin{aligned}
&p_\mu(\theta_1 \times \{s_1, s_2\}) = \beta \\
&p_\mu(\theta_2 \times \{s_1, s_2\}) = 1 - \beta \\
&p_\mu(\theta_1 \times s_1) = \beta[\mu \cdot \lambda_1 + (1 - \mu) \cdot \lambda_2] \\
&p_\mu(\theta_2 \times s_1) = (1 - \beta)[\mu \cdot (1 - \lambda_1) + (1 - \mu) \cdot (1 - \lambda_2)] \\
&p_\mu(\{\theta_1, \theta_2\} \times s_1) = p_\mu(\theta_1 \times s_1) + p_\mu(\theta_2 \times s_1)
\end{aligned}$$

To fully characterize the DM's behavior in this example, there are in total six different cases: $\beta > 1/2, \beta = 1/2$ and $\beta < 1/2$ combined with signal realization s_1 or s_2. Only one of the cases will be derived here in detail and a summary of behaviors in all these cases will be provided afterwards in Table 1.

Consider the case $\beta > 1/2$ and signal s_1 is realized. Notice that when $\beta > 1/2$, the likelihood of signal s_1, given by $p_\mu(\{\theta_1, \theta_2\} \times s_1)$ is maximized when $\mu = 1$. Moreover, thanks to the simple parametrization of the set C, for any $\alpha \in [0, 1]$ the set $C_\alpha(\{\theta_1, \theta_2\} \times s_1)$ is given by

$$C_\alpha(\{\theta_1, \theta_2\} \times s_1) = \{p_\mu \in C : \mu \in [\alpha, 1]\}$$

Under RML, the DM updates only the priors p_μ for $\mu \in [\alpha, 1]$ and for each p_μ the posterior about states after observing s_1 is given by

$$\pi_\mu(\theta_1|s_1) \equiv \frac{p_\mu(\theta_1 \times s_1)}{p_\mu(\{\theta_1, \theta_2\} \times s_1)}$$

Therefore, the DM's set of posteriors will be the following interval[18]: $[\pi_\alpha(\theta_1|s_1), \pi_1(\theta_1|s_1)]$. Clearly, under FB, the DM's set of posteriors will be $[\pi_0(\theta_1|s_1), \pi_1(\theta_1|s_1)]$, and under ML, the DM will end up with a single posterior: $\pi_1(\theta_1|s_1)$.

Furthermore under MEU, the DM's conditional evaluation of the bet f will be given by the lowest posterior of state s_1. Let x_f^{FB}, x_f^{RML} and x_f^{ML} denote the conditional certainty equivalent of the bet f under FB, RML and ML updating. Then one has $x_f^{FB} \leq x_f^{RML} \leq x_f^{ML}$ for all $\alpha \in [0,1]$. Especially, x_f^{RML} is increasing with respect to α and coincides with x_f^{FB} and x_f^{ML} when $\alpha = 0$ and $\alpha = 1$ respectively.

Table 1 summarizes the predictions of behaviors under RML in all six cases. The first two rows are the only cases where the DM's evaluation of f depends on the value of α. In other words, those are the cases where different updating rules affect one's conditional evaluation of f. Thus, RML is able to provide richer predictions than FB and ML.

The third and fourth row are the cases where the MEU evaluation of f is given by the posteriors updated from the maximum likelihood priors. Thus, even though the updated beliefs are different under FB and ML, MEU makes their evaluations the same. For this reason, although the updated belief under RML is also different from FB and ML for any $\alpha \in (0,1)$, the MEU evaluation under RML is still the same.

Finally, the fifth and sixth rows represent the same type of special case where every prior agrees on the likelihood of every signal. It is the case where FB coincides with ML, and therefore RML does not have an extra bite. All the priors will be updated under RML for all $\alpha \in [0,1]$, thus the updated belief will always be the same.

β	Signal	ML prior	Evaluation of f	Comparison with probabilistic signal
$> 1/2$	s_1	$\mu = 1$	$\pi_\alpha(\theta_1|s_1)$	$\alpha < 1/2$: lower $\alpha = 1/2$: equal $\alpha > 1/2$: higher
$> 1/2$	s_2	$\mu = 0$	$\pi_\alpha(\theta_1|s_2)$	$\alpha < 1/2$: lower $\alpha = 1/2$: equal $\alpha > 1/2$: higher
$< 1/2$	s_1	$\mu = 0$	$\pi_0(\theta_1|s_1)$	All α: lower
$< 1/2$	s_2	$\mu = 1$	$\pi_1(\theta_1|s_2)$	All α: lower
$= 1/2$	s_1	All μ	$\pi_0(\theta_1|s_1)$	All α: lower
$= 1/2$	s_2	All μ	$\pi_1(\theta_1|s_2)$	All α: lower

Table 1: Summary of Example 2 under RML

The last column of Table 1 presents a comparison highlighted in experiments by Liang (2019). It compares the DM's s conditional evaluations of f under this ambiguous signaling structure with a probabilistic signaling structure where the correlation equals to $(\lambda_1+\lambda_2)/2$. The latter is designed to reflect an average accuracy of the two signaling devices in the ambiguous signaling structure. For probabilistic signals, the DM simply follows Bayesian updating.

In the case where s_1 is realized, notice that the DM's posterior under the probabilistic signaling structure is exactly given by $\pi_{1/2}(\theta_1|s_1)$. Therefore, for $\beta \leq 1/2$, the DM's conditional evaluation

[18]Notice that $\pi_\mu(\theta_1|s_1)$ is increasing in μ.

under ambiguous signal is always **lower** than her conditional evaluation under probabilistic signal. On the other hand, when $\beta > 1/2$, this comparison will depend on the value of α. More specifically, $\alpha < 1/2$ implies that the conditional evaluation under ambiguous signal is lower, $\alpha > 1/2$ implies that the conditional evaluation under ambiguous signal is higher, and $\alpha = 1/2$ implies they are the same. Thus, from the perspective of RML, this type of comparison sometimes actually reflects the DM's attitudes towards discarding priors based on likelihood captured by α.

For the experimental results shown by Table 4.2 in Liang (2019), each row there can be categorized into some case in the Table 1 here[19]. As s_1 is "good news" and s_2 is "bad news" in his terminology. A rather interesting pattern emerged from the choice data is when $\beta > 1/2$, the subjects' comparison is lower after receiving good news (s_1) and higher after receiving bad news (s_2). Notice that from Table 1, the two directions of the comparisons are aligned with $\alpha < 1/2$ and $\alpha > 1/2$ respectively. This implies that the subjects may have different attitudes towards discarding priors when the signal realization is different, especially when they are associated with meanings such as "good news" and "bad news". This pattern can be captured by the **Contingent RML**, which is characterized exactly to reflect such flexibility in behaviors.

Moreover, the other rows of Table 4.2 correspond to the cases where α is irrelevant for this comparison. The choice data there is also aligned with behaviors under RML, i.e. the subjects find evaluations of f is lower under ambiguous signals[20].

Therefore, not only the Contingent RML with $\alpha[\{\theta_1,\theta_2\}\times s_1] < 1/2$ and $\alpha[\{\theta_1,\theta_2\}\times s_2] > 1/2$ offers an interpretation for the different directions of comparison across signals, but it also provides an approach to accommodate almost all the behavioral patterns in that experiment[21].

In summary, for this special case where a DM has probabilistic belief over states and ambiguous belief over signals, RML is able to provide richer predictions. Such richness, on the other hand, also proves to be useful for applications involving ambiguous signals.

5 Likelihood-robust Ambiguous Persuasion: An Example

In settings such as mechanism design or information design, many recent papers show that the designer could gain strictly more payoff by introducing ambiguity in their communications. These findings are mostly observed in the case where the agent/receiver is modeled by MEU preferences with FB updating. This section uses an example in the context of information design to explore both the robustness of such a finding once the FB assumption is relaxed to RML and how the generalization to RML may affect optimal persuasion.

Consider the persuasion environment studied by Kamenica and Gentzkow (2011), where the sender commits to a signaling device and the receiver takes actions contingent on the signal realizations. Bayesian persuasion describes the case where the available signaling devices for the sender are only probabilistic: each device is a mapping from the state space to the distributions over the signal space.

[19] The mentioned experimental result is included in appendix C for reference.

[20] Only row 8 of Table 4.2 in Liang (2019) is different from the prediction of RML. The last two rows do not specify the signals thus are ignored.

[21] The same conclusion can be drawn also from a within-subject comparison shown in Table B.5 of Liang (2019)

If the sender also has access to an ambiguous device, which specifies a set of probabilistic devices and the probability of using any one of them is unknown[22]. Beauchêne et al. (2019) show that when the receiver's preference is represented by MEU with FB updating, the sender is able to gain strictly more payoff by using an ambiguous device compared with the optimal probabilistic device.

Notice that once the sender commits to an ambiguous device, from the receiver's point of view, it is exactly the updating ambiguous signals situation extensively discussed in the previous section. Clearly, the FB assumption excludes the possibility that the receiver may be willing to make inferences about the devices based on the likelihood of the realized signal. This raises the possibility that the strict gain from using ambiguous signals may hinge on assumptions concerning such inferences.

Beauchêne et al. (2019) briefly talk about the scenario where the receiver updates beliefs using ML instead of FB. They show an example where the sender could strictly benefit from sending ambiguous signals. However, in that example, the sender cannot benefit from sending ambiguous signals if the receiver were assumed to update using FB. Thus, their additional discussion does not resolve the issue addressed here, as it still relies on the specific updating rule.

RML, as a unifying updating rule including both FB and ML, provides a convenient tool to address this issue. More specifically, with the parametrization of RML, one can check for which values of the parameter α the same conclusion holds. In particular, if the strict gain from using ambiguous signals can be guaranteed for all possible values of α, then such a finding is *likelihood-robust*.

To give an example of such an analysis, I take the illustrative example from Beauchêne et al. (2019) and relax their FB assumption to RML. For this particular example, first of all, the ambiguous device they construct is more beneficial than the optimal probabilistic device only when $\alpha = 0$, i.e. only when RML reduces to FB. In other words, their finding of strict benefits from using ambiguous signals seems to crucially depend on the specific assumption of FB.

Nonetheless, for this example, I further show that there exists a *likelihood-robust* ambiguous device. It induces the same actions from the receiver for all possible values of α under RML. Moreover, the sender gains strictly more payoff from this ambiguous device than from the optimal probabilistic device. Namely, the strict benefit from using ambiguous signals in this example is robust to the possibility that the receiver may make likelihood-based inferences from the realized signals.

Example 3. *There are two states $\{\omega_l, \omega_h\}$ with a uniform marginal prior, and the receiver has three feasible actions: $\{a_l, a_m, a_h\}$. Payoff of the sender and receiver for each state and action is given as follows:*

	ω_l	ω_h
a_l	*(-1,3)*	*(-1,-1)*
a_m	*(0,2)*	*(0,2)*
a_h	*(1,-1)*	*(1,3)*

where in each cell, the first number is the sender's payoff and the second is the receiver's.

[22] The sender can choose a set of probabilistic devices and delegate the choice from this set to a third party or to the draw from an Ellsberg urn to make the signals ambiguous.

Notice the sender always prefers the receiver to take higher action yet the receiver prefers to choose the action that matches the state. Moreover, the sender is modeled to optimize her ex-ante payoff under MEU preferences.

First, consider the setting with FB assumption. In the following, I replicate the construction of the ambiguous device from Beauchêne et al. (2019). For more in-depth explanations, readers are advised to refer to the original paper.

Let $\{m_l, m_h\}$ be the set of signals and consider the following two probabilistic devices π_1 and π_2, where the cell (m_l, ω_l) denotes the probability of generating signal m_l given state ω_l.

$\pi_1(m\|\omega)$	ω_l	ω_h
m_l	$2/3$	0
m_h	$1/3$	1

$\pi_2(m\|\omega)$	ω_l	ω_h
m_l	$3/4$	$1/4$
m_h	$1/4$	$3/4$

If the sender's ambiguous device is a set containing π_1 and π_2, then the receiver would form the following set of posteriors:

$$p(\omega_h|m_l) = \{0, 1/4\}$$
$$p(\omega_h|m_h) = \{3/4, 3/4\}$$

where the first and second posterior in each set is updated from π_1 and π_2 respectively. Given these posteriors, the receiver with MEU preference would take action a_m when signal m_l is realized and takes action a_h when signal m_h is realized. The posterior $p(\omega_h|m_l) = 1/4$ is crucial since any posterior assigning a smaller probability on ω_h would induce the receiver to take action a_l, which makes the sender worse off.

Given the receiver's action for each signal, the sender's ex-ante evaluation of the ambiguous device may also be affected by the existence of ambiguity. If the sender uses an ambiguous device which contains exactly these two devices, then under MEU her ex-ante payoff is given by (u_s stands for the sender's utility function)

$$\begin{aligned}
&\min_{\pi\in\{\pi_1,\pi_2\}} \left(\frac{1}{2}\pi(m_l|\omega_l) + \frac{1}{2}\pi(m_l|\omega_h)\right) u_s(a_m) + \left(\frac{1}{2}\pi(m_h|\omega_l) + \frac{1}{2}\pi(m_h|\omega_h)\right) u_s(a_h)\\
&= \left(\frac{1}{2}\pi_2(m_l|\omega_l) + \frac{1}{2}\pi_2(m_l|\omega_h)\right) u_s(a_m) + \left(\frac{1}{2}\pi_2(m_h|\omega_l) + \frac{1}{2}\pi_2(m_h|\omega_h)\right) u_s(a_h)\\
&= \frac{1}{2}u_s(a_m) + \frac{1}{2}u_s(a_h)
\end{aligned}$$

This is exactly the payoff from using the optimal probabilistic device. Thus, the sender needs to hedge against this ambiguity to get a higher payoff.

Consider the following construction: First, increase the number of signals to four such that now the signals are $\{m_l, m_h, m_l', m_h'\}$. Second, consider a probabilistic device generated by a mixture of π_1 and π_2: $\pi_1' = \lambda\pi_1 \oplus (1-\lambda)\pi_2$. This represents a device sending signals $\{m_l, m_h\}$ with probability λ according to the device π_1 and sending signals $\{m_l', m_h'\}$ with probability $(1-\lambda)$ according to π_2. Given this device, the receiver's posteriors coincide with π_1 when $m \in \{m_l, m_h\}$ and coincides with π_2 when $m \in \{m_l', m_h'\}$.

Moreover, consider the following two probabilistic devices constructed in the same manner $\pi_1' = \lambda\pi_1 \oplus (1-\lambda)\pi_2$ and $\pi_2' = (1-\lambda)\pi_2 \oplus \lambda\pi_1$:

$\pi_1'(m\|\omega)$	ω_l	ω_h
m_l	$\lambda \cdot 2/3$	0
m_h	$\lambda \cdot 1/3$	λ
m_l'	$(1-\lambda)\cdot 3/4$	$(1-\lambda)\cdot 1/4$
m_h'	$(1-\lambda)\cdot 1/4$	$(1-\lambda)\cdot 3/4$

$\pi_2'(m\|\omega)$	ω_l	ω_h
m_l	$(1-\lambda)\cdot 3/4$	$(1-\lambda)\cdot 1/4$
m_h	$(1-\lambda)\cdot 1/4$	$(1-\lambda)\cdot 3/4$
m_l'	$\lambda \cdot 2/3$	0
m_h'	$\lambda \cdot 1/3$	λ

When $m \in \{m_l, m_h\}$, the posterior of π_1' coincides with π_1 and the posterior of π_2' coincides with π_2, so that the set of posteriors generated by the ambiguous device $\Pi' = \{\pi_1', \pi_2'\}$ remains the same as the ambiguous device containing π_1 and π_2. Then signals m_l and m_l' would induce the receiver to take action a_m and signals m_h and m_h' would induce action a_h.

Furthermore, notice that the difference between these two probabilistic devices is *only* the labels of the signals. Thus, given the receiver's action is the same across signal m_l and m_l' as well as across signals m_h and m_h'. The two probabilistic devices induce each action with the same frequency, hence, the sender's ex-ante payoff will be the same across these two devices. Therefore, using an ambiguous device containing π_1' and π_2' will not generate any ambiguity for the sender's ex-ante payoff.

More specifically, the sender's ex-ante payoff from this ambiguous device is given by

$$\left[\frac{1}{2}(1-\lambda) + \frac{1}{3}\lambda\right] u_s(a_m) + \left[\frac{1}{2}(1-\lambda) + \frac{2}{3}\lambda\right] u_s(a_h) \tag{5.1}$$

which is strictly higher than using the optimal probabilistic device when $\lambda > 0$. Furthermore, it is increasing in λ.

Therefore, the optimal ambiguous persuasion can be approached by letting $\lambda \to 1$, notice that λ cannot be exactly one.

So far, Beauchêne et al. (2019)'s construction of the optimal ambiguous device under the assumption of FB is provided. Notice that the likelihood of generating the signals by each device depends on λ. If letting $\lambda \to 1$, the likelihood of generating the same signal by each device might be severely different.

Let $l_i(m)$ denote the likelihood of generating signal m under device π_i':

π_1'	$l_1(m)$	$p(\omega_h\|m)$
m_l	$\lambda \cdot 1/3$	0
m_h	$\lambda \cdot 2/3$	3/4
m_l'	$(1-\lambda)\cdot 1/2$	1/4
m_h'	$(1-\lambda)\cdot 1/2$	3/4

π_2'	$l_2(m)$	$p(\omega_h\|m)$
m_l	$(1-\lambda)\cdot 1/2$	1/4
m_h	$(1-\lambda)\cdot 1/2$	3/4
m_l'	$\lambda \cdot 1/3$	0
m_h'	$\lambda \cdot 2/3$	3/4

When $\lambda \to 1$, notice that the likelihood of generating signal m_l by device π_2' goes to 0, whereas the likelihood of m_l by device π_1' goes to $1/3$. Intuitively, knowing $\lambda \to 1$, whenever signal m_l is observed, the receiver should be almost sure that it is generated by device π_1'. However, the crucial posterior $1/4$ inducing action a_m is in fact generated by the other device π_2'.

Indeed, for RML with any $\alpha > 0$, since the device π_2' has the minimum likelihood of generating the signal m_l, the crucial posterior $1/4$ is always excluded from the set of posteriors. Then, the receiver would find a_l to be strictly better than a_m, thus the desired action a_m can no longer be induced. One can also verify that the sender's payoff is equivalent to some non-optimal probabilistic device. Thus she instead becomes strictly worse off using ambiguous signals compared with using the optimal probabilistic device when the receiver slightly deviates from FB in the direction of ML.

Nonetheless, when $\lambda = 3/5$, the likelihood of generating signal m_l by device π'_1 and π'_2 becomes the same and it is also true for the signal m'_l. In this case, the receiver cannot use the likelihood of the realized signal to make any inference about the devices. It is the case where FB and ML coincide, and the receiver will always update with respect to both devices under RML no matter what his α is. As the crucial posterior $1/4$ will always be updated, the sender is able to induce the action a_m from the receiver when signal m_l or m'_l realizes.

Furthermore, even though the likelihood of generating m_h and m'_h is not the same across the two devices, as the corresponding posteriors are the same, these signals can always induce the same action a_h from the receiver as well. That being said, the ambiguous device Π' with $\lambda = 3/5$ is able to always induce action a_m and a_h with corresponding signals regardless of the receiver's likelihood-based inferences. In particular, for all possible values of α in the case of RML updating.

In addition, the sender's payoff will be given by equation (5.1) with $\lambda = 3/5$. Thus the value is strictly higher than using the optimal probabilistic device.

Therefore, this ambiguous device guarantees the sender strictly more payoff than using probabilistic devices in a likelihood-robust way. As a result, it suggests that the strict benefit from using ambiguous signals is robust to the concern that the receiver may make likelihood-based inferences about the devices from the realized signals.

6 Related Literature

This paper adds to the literature on dynamic choice under ambiguity by proposing the RML updating rule and providing an axiomatic foundation for it. RML is motivated by the idea of using observed information to refine the initial belief in updating. For multiple priors, FB (Pires, 2002) and ML (Gilboa and Schmeidler, 1993) are the two extremes of refining the initial set of priors according to the likelihood of the observed information. RML is able to capture intermediate behaviors between these two.

6.1 Refining the Initial Belief in Updating

The idea of refining beliefs as new information arrives is essential in many different non-Bayesian updating rules. And it is certainly not an exclusive feature of updating multiple priors.

For example, when the initial belief is a single distribution over the states, Ortoleva (2012) characterizes the hypothesis testing updating rule. Under this rule, if the probability of the observed event is too small according to the initial belief, the DM will then find another prior as her revised belief and apply Bayes' rule to this new prior for updating. The revised prior is selected according to a likelihood-based criterion. As a result, the hypothesis testing updating rule emphasizes dealing with unexpected events, i.e. those assigned a probability of zero or almost zero under the initial belief.

Also when the initial belief is a singleton, Zhao (2017) characterizes the Pseudo-Bayesian updating rule when information takes the form "event A is more likely than event B". In the case where received information contradicts the DM's initial belief, she also finds another prior as her revised belief and applies Bayes' rule to update. This revised belief is chosen from the ones satisfying the constraint specified by the unexpected information such that it is closest to the initial belief in terms of Kullback-Leibler divergence.

When the initial belief is ambiguous yet the DM is ambiguity neutral represented by a single prior, Suleymanov (2018) characterizes the Robust Maximum Likelihood updating rule in which the DM revises her initial belief according to the maximum likelihood of the observed event. Namely, the DM is an expected utility maximizer both ex-ante and conditionally, yet the posterior is not updated from the prior. Thus, the main difference between the Robust Maximum Likelihood and RML (Relative Maximum Likelihood) is that the latter updating rule requires the posteriors to be updated from the subset of those priors representing the DM's ex-ante preference. In other words, RML necessarily reduces to Bayesian updating when the DM's ex-ante preference is represented by the expected utility, which is not true for Robust Maximum Likelihood updating.

In cases where the initial belief is a set of priors, one way of refining is discarding priors from the initial set according to some criterion. RML belongs to this category. An updating rule proposed by Epstein and Schneider (2007) uses likelihood ratio test in statistics as a criterion for discarding priors. More discussions about the differences between RML and their rule can be found in section 6.2

On the other hand, the dynamic consistent updating rule characterized in Hanany and Klibanoff (2007) also features discarding priors, and the criterion there is to maintain the optimality of the ex-ante optimal act.

Yet another way of refining is to consider a different set of priors, where it is possible that some priors are not presented in the initial belief. Ortoleva (2014) characterizes the hypothesis testing updating rule for multiple priors: if the likelihood of the observed event is too low under some prior in the initial belief, then the DM will revise her initial belief and change to another set of priors for updating.

Beyond the theoretical developments on this idea of refining the initial belief, De Filippis et al. (2018) also identify such behavior in a social learning lab experiment. Their finding suggests that the non-Bayesian behavior observed in the experiment is consistent with a generalized maximum likelihood updating rule where subjects revise their initial belief according to the information received.

6.2 Likelihood Ratio Test

A likelihood ratio test is commonly used in statistics to determine whether a statistical model with fewer parameters is "good enough" compared with a model with the maximum number of parameters. The former model is good enough if its likelihood of generating the observed data is sufficiently close to the likelihood under the latter model according to their ratio compared with some threshold.

In the literature on updating ambiguous beliefs, Epstein and Schneider (2007) explicitly use the likelihood ratio test as the criterion to determine whether a prior will be updated or not. For example, given a threshold $\lambda \in [0, 1]$, the DM applies Bayes' rule to update a prior p in the set of priors C if and only if the following inequality holds:

$$\frac{p(E)}{\max_{p\in C} p(E)} \geq \lambda \tag{6.1}$$

In this sense, the value of this threshold (λ) also reflects an extent of willingness to discard priors based on likelihood. For RML, its functional form suggests a similar criterion for discarding priors. Notice that, for all $p \in C_\alpha(E) \equiv \alpha C^*(E) + (1-\alpha)C$,

$$\frac{p(E)}{\max_{p\in C} p(E)} \geq \alpha + (1-\alpha)\frac{\min_{p\in C} p(E)}{\max_{p\in C} p(E)}$$

or equivalently

$$\frac{p(E) - \min_{p\in C} p(E)}{\max_{p\in C} p(E) - \min_{p\in C} p(E)} \geq \alpha$$

The second formula implies that the criterion suggested by RML considers a *relative* likelihood ratio test where $\min_{p\in C} p(E)$ becomes an additional benchmark. The threshold α determines the fraction of priors that are deemed to be plausible according to relative likelihood. Clearly, when the minimal probability of event E is zero, this criterion coincides with that in equation (6.1) when $\alpha = \lambda$.

Importantly, however, this relative likelihood ratio criterion is only a necessary condition for a prior to be updated by Bayes' rule according to RML. The linear contraction suggests that the shape of the set of priors C also plays a crucial role. In fact, the set $C_\alpha(E)$ sometimes is a strict subset of the set satisfying the relative likelihood ratio criterion with the same α.

Let $\hat{C}_\alpha(E)$ denote the set satisfying the relative likelihood ratio criterion with some α:

$$\hat{C}_\alpha(E) = \left\{ p \in C : \frac{p(E) - \min_{p\in C} p(E)}{\max_{p\in C} p(E) - \min_{p\in C} p(E)} \geq \alpha \right\}$$

Figure 4 shows a case where $C_\alpha(E) \subsetneqq \hat{C}_\alpha(E)$. For the same scenario as in figure 1, notice that the area below the red dashed line and in the set C is the set $\hat{C}_\alpha(E)$ with the same α as in $C_\alpha(E)$. Hence the area below the red dashed line yet is not in $C_\alpha(E)$ represents the priors satisfying the likelihood-based criterion yet are still discarded under RML updating.

In summary, though RML is in someways similar to rules, such as those in Epstein and Schneider (2007), using likelihood ratio thresholds, its reliance on the *relative* likelihood ratios and on the shape of the set of priors differentiate it from such approaches.

Further comparison of RML with updating rules using likelihood ratio thresholds may be seen through comparison of the characterization axioms. For example, both Kovach (2015) and Hill (2019) provide axiomatic characterizations for such updating rules in different frameworks. Their axioms leverage the objective lotteries to calibrate the DM's subjective belief about the conditioning events. In contrast, the axiomatization of RML does not rely on this special structure.

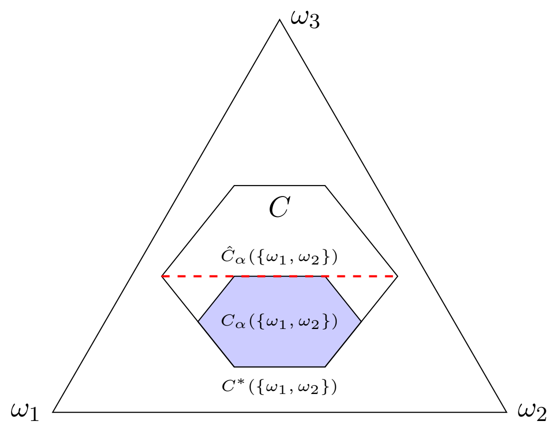

Figure 4: Graphical Illustration of $C_\alpha(E)$ and $\hat{C}_\alpha(E)$

6.3 Dynamic Choice under Ambiguity

A fundamental issue in dynamic choice under ambiguity is the fact that, for ambiguity sensitive choices, an updating rule cannot preserve both consequentialism and dynamic consistency at the same time (Hanany and Klibanoff, 2007; Siniscalchi, 2009).

According to the definition in Hanany and Klibanoff (2007), consequentialism means that conditional preferences should not depend on an event not occurring or the context of the decision problem (e.g. feasible acts, etc.). Dynamic consistency (which is weaker than the DC axiom in this paper) means that the ex-ante most preferred act should remain optimal to any other acts that are both feasible and agree with it on the event not occurring after updating. Any updating rule for ambiguous beliefs without other restrictions needs to relax either one of these two properties.

RML takes the consequentialist approach by requiring consequentialism and relaxing dynamic consistency. The essential implication of consequentialism is that the DM should update her belief in the same way regardless of the decision problem at hand. In other words, the updated belief is given by a function of only the ex-ante belief and the conditioning event.

As a result, dynamic consistency will sometimes be violated under RML. On one hand, the violation of dynamic consistency under ambiguity is commonly observed in experiments, e.g. Dominiak et al. (2012). On the other hand, consistent planning is proposed in the literature as a way to overcome this problem of consequentialist updating rules. With consistent planning, the DM is assumed to be sophisticated such that she is able to anticipate her future decisions and chooses an ex-ante optimal plan accordingly. A behavioral characterization of consistent planning is given by Siniscalchi (2011).

Along another route, Hanany and Klibanoff (2007) and Hanany and Klibanoff (2009) axiomatize updating rules that preserve dynamic consistency yet do not require consequentialism. More specifically, the dynamic consistent updating rules for multiple priors characterized in Hanany and Klibanoff (2007) explicitly depend on the feasible set of acts as well as the optimal acts. Therefore, different decision problems result in different updated beliefs according to this rule.

Instead of relaxing one of these two properties, Epstein and Schneider (2003) characterize the rectangularity condition for the set of priors such that dynamic consistency is preserved under FB updating, i.e. consequentialism also holds. Namely, dynamic consistency and consequentialism can both be satisfied when the ex-ante belief takes a specific form. The regularity condition, however, imposes a restriction on the possible conditioning events to which FB updating is applicable.

Another way of keeping both consequentialism and dynamic consistency is to relax reduction of compound evaluations[23], pursued by Li (2015) and Gul and Pesendorfer (2017). According to this approach, the DM's ex-ante evaluation of an act will also depend on the temporal resolution of uncertainty.

In addition, Gul and Pesendorfer (2017) observe a feature of both FB and ML, which is "all news is bad news". Namely, a DM sometimes finds that the ex-ante preferred alternative is dominated by another alternative no matter what the signal realization is. Assuming "not all news can be bad news", they use this axiom to characterize the updating by proxy rule for preferences admitting CEU representation with totally monotone capacities. This class of preferences is a strict subset of the preferences considered in this paper.

Their observation "all news is bad news" also applies to RML, since FB and ML are special cases of RML. However, requiring "not all news can be bad news" as in the updating by proxy rule sometimes implies that the beliefs updated from ambiguous signals are completely unambiguous[24]. Without imposing such an assumption, RML on the other hand allows for more flexibility in terms of how much ambiguity can be preserved when updating ambiguous signals.

7 Concluding Remarks

A general updating procedure for multiple priors proposed at the beginning of this paper specifies two steps. In the first step, the DM uses the observed information to refine the initial belief. And then in the second step, she applies Bayes' rule to update the remaining priors.

Bayesian updating of a single prior belief is a special case where the first step is absent. In the case where the true probability law governing the uncertainty is known, such absence is reasonable since one cannot further refine the belief but can only update it conditional on the information received.

However, in scenarios where the underlying probability law is unknown, the DM needs to form a conjecture about the uncertainty for decision making. Whether the conjecture is a singleton or a set of probabilities, it all seems too stringent to require the DM to always stick with her initial conjecture despite new information she might receive. Thus, applying Bayes' rule to the conjecture belief actually reflects confidence about her initial belief. Accordingly, updating rules that do not reflect such confidence and allow for revising the initial belief might also be reasonable.

Several different updating rules have been proposed in the literature to capture the situation in which initial conjecture is a singleton and it may be revised after seeing new information. For when the initial beliefs are multiple priors, this paper proposes RML updating rule, in which the initial beliefs are revised based on the likelihood of the information observed. More importantly, this paper pinpoints the behaviors that are equivalent to such an updating rule providing a preference foundation for using likelihood as a criterion in updating.

This paper also uses RML to address applications involving ambiguity in information design and mechanism design. Many existing results in these areas are derived based on the specific assumption of FB. I have shown RML to be useful for identifying the extent of deviation from FB allowed for those results to still hold. Last but not least, I illustrate through an example that

[23] Or the law of iterated expectation in Gul and Pesendorfer (2017)'s terminology

[24] See Example 1 in Gul and Pesendorfer (2017) for an example

adopting RML instead of FB as the model of updating helps robustify results from the literature meanwhile maintaining tractability.

Appendix A Proofs of the results

Throughout all the proofs, let C denote the convex and closed set of priors representing the ex-ante preference; let $C^*(E)$ denote the subset of C assigning a maximal probability to the event E: $C^*(E) \equiv \{p \in C : p(E) \geq p'(E) \, \forall p' \in C\}$ and let $p^*(E)$ denote the maximal probability of event E: $p^*(E) \equiv \max_{p \in C} p(E)$.

A.1 Proof of Theorem 2.4

Necessity. The necessity of CR-S for ML updating is proved via the following three lemmas:

Lemma A.1. *Suppose the conditional preferences are represented by ML. For all strict $\succsim$-nonnull $E \in \Sigma$, for all $f \in \mathcal{F}$ and for all $x, x^* \in X$ with $x^* \succsim x$, if $f \sim_E x$ and*

$$\min_{p \in C} \int_\Omega u(f_E x^*) dp = \min_{p \in C^*(E)} \int_\Omega u(f_E x^*) dp$$

then $f_E x^* \sim x_E x^*$.

Proof of Lemma A.1. Suppose the conditional preferences are represented by ML. For any strict $\succsim$-nonnull E, $f \sim_E x$ implies that $\min_{p \in C^*(E)} \int_E u(f) \frac{dp}{p(E)} = u(x)$, then one can further derive

$$\begin{aligned}
\min_{p \in C} \int_\Omega u(f_E x^*) dp &= \min_{p \in C^*(E)} \int_\Omega u(f_E x^*) dp \\
&= \min_{p \in C^*(E)} \left[\int_E u(f_E x^*) \frac{dp}{p(E)} \cdot p(E) + (1 - p(E)) u(x^*) \right] \\
&= p^*(E) \cdot \min_{p \in C^*(E)} \int_E u(f) \frac{dp}{p^*(E)} + (1 - p^*(E)) u(x^*) \\
&= p^*(E) u(x) + (1 - p^*(E)) u(x^*) \\
&= \min_{p \in C} \int_\Omega u(x_E x^*) dp
\end{aligned}$$

where the third equality follows from $p(E) = p^*(E)$ for all $p \in C^*(E)$, the last equality follows from the fact that $u(x^*) \geq u(x)$ as $x^* \succsim x$. □

Lemma A.2. *For all strict $\succsim$-nonnull $E \in \Sigma$, for all $f \in \mathcal{F}$, if there exists $\bar{x} \in X$ such that*

$$\min_{p \in C} \int_\Omega u(f_E \bar{x}) dp = \min_{p \in C^*(E)} \int_\Omega u(f_E \bar{x}) dp$$

then for all $x^* \succsim \bar{x}$ one has

$$\min_{p \in C} \int_\Omega u(f_E x^*) dp = \min_{p \in C^*(E)} \int_\Omega u(f_E x^*) dp$$

Proof of Lemma A.2. For any strict $\succsim$-nonnull $E \in \Sigma$ and any $f \in \mathcal{F}$. Suppose there exists $x \in X$ such that $\min_{p \in C} \int_\Omega u(f_E x)dp = \min_{p \in C^*(E)} \int_\Omega u(f_E x)dp$. Towards a contradiction suppose there also exists x' such that $x' \succsim x$ as well as

$$\min_{p \in C} \int_\Omega u(f_E x')dp < \min_{p \in C^*(E)} \int_\Omega u(f_E x')dp$$

Then this strict inequality further implies:

$$\min_{p \in C} \left[\int_E u(f)dp + u(x')(1 - p(E)) \right] < \min_{p \in C^*(E)} \left[\int_E u(f)dp + u(x')(1 - p(E)) \right]$$

$$\min_{p \in C} \left[\int_E u(f)dp + u(x')(1 - p(E)) \right] < \min_{p \in C^*(E)} \left[\int_E u(f)dp + u(x)(1 - p^*(E)) + (u(x') - u(x))(1 - p^*(E)) \right]$$

$$\min_{p \in C} \left[\int_E u(f)dp + u(x')(1 - p(E)) \right] - (u(x') - u(x))(1 - p^*(E)) < \min_{p \in C^*(E)} \int_\Omega u(f_E x)dp$$

$$\min_{p \in C} \left[\int_E u(f)dp + u(x)(1 - p(E)) + (u(x') - u(x))(p^*(E) - p(E)) \right] < \min_{p \in C^*(E)} \int_\Omega u(f_E x)dp$$

$$\min_{p \in C} \left[\int_\Omega u(f_E x)dp + (u(x') - u(x))(p^*(E) - p(E)) \right] < \min_{p \in C^*(E)} \int_\Omega u(f_E x)dp$$

Notice that for the LHS of the last inequality, the minimum of $\int_\Omega u(f_E x)dp$ can be obtained at some $p \in C^*(E)$ which also minimizes the second term $(u(x') - u(x))(p^*(E) - p(E))$, since for all $p \in C$, $p^*(E) - p(E) \geq 0$. Thus the minimum of LHS is obtained at some p with $p(E) = p^*(E)$ and it implies

$$\min_{p \in C} \int_\Omega u(f_E x)dp = \min_{p \in C^*(E)} \int_\Omega u(f_E x)dp < \min_{p \in C^*(E)} \int_\Omega u(f_E x)dp$$

which is a contradiction. □

Lemma A.3. *If $\succsim$ admits MEU representation with C having finitely many extreme points, then for all strict $\succsim$-nonnull $E \in \Sigma$, for all $f \in \mathcal{F}$, an $\bar{x}_{E,f} \in X$ such that both*

$$\min_{p \in C} \int_\Omega u(f_E \bar{x}_{E,f})dp = \min_{p \in C^*(E)} \int_\Omega u(f_E \bar{x}_{E,f})dp$$

and $\bar{x}_{E,f} \succsim_E f$ hold always exists.

Proof of Lemma A.3. First show that when C contains only finitely many extreme points, an $\bar{x}_{E,f}$ such that $f_E \bar{x}_{E,f}$ is evaluated at some extreme point in $C^*(E)$ always exists.

Let q be any extreme point of $C^*(E)$ and let p be any extreme point of C. The act $f_E \bar{x}_{E,f}$ is evaluated at q if for all extreme points p of C,

$$\int_\Omega u(f_E \bar{x}_{E,f})dq \leq \int_\Omega u(f_E \bar{x}_{E,f})dp$$

It can be further derived as

$$\int_E u(f)dq + (1 - q(E))u(\bar{x}_{E,f}) \leq \int_E u(f)dp + (1 - p(E))u(\bar{x}_{E,f}) \tag{A.1}$$

Notice that the first term of both LHS and RHS does not depend on $\bar{x}_{E,f}$, furthermore, $(1-q(E)) \leq (1-p(E))$ as $q \in C^*(E)$. When p is also in $C^*(E)$, the value of $\bar{x}_{E,f}$ does not matter and one can pin down the $q \in C^*(E)$ such that minimizes the evaluation of $f_E\bar{x}_{E,f}$ among all extreme points in $C^*(E)$. Denote the minimizing q by q^*.

Fix q^*, for an extreme point p not in $C^*(E)$, the following inequality becomes strict: $(1 - q^*(E)) < (1 - p(E))$. Then for all E, f and such an extreme point p, inequality (A.1) holds if

$$u(\bar{x}_{E,f,p}) \geq \frac{\int_E u(f)dq^* - \int_E u(f)dp}{q^*(E) - p(E)}$$

Since X is unbounded from above under the ex-ante preference, such an $\bar{x}_{E,f,p}$ always exist. Then it suffices to let $\bar{x}_{E,f} = \max\{\max_p \bar{x}_{E,f,p}, x\}$ for $x \sim_E f$ where both "max" are according to the ex-ante preference. Since there are only finitely many extreme points p in C, the maximum over a finite set always exists. □

In summary, for all strict $\succsim$-nonnull $E \in \Sigma$, for all $f \in \mathcal{F}$ and $x \in X$, Lemma A.2 and A.3 together show the existence of a threshold $\bar{x}_{E,f}$ such that for all $x^* \succsim \bar{x}_{E,f}$, the maxmin evaluation of the act $f_E x^*$ is given by an extreme point of $C^*(E)$, meanwhile $x^* \succsim x$. Then Lemma A.1 further implies that, when $\succsim_E$ is given by ML updating, for all $x^* \succsim \bar{x}_{E,f}$ one has $f_E x^* \sim x_E x^*$.

Sufficiency. For the sufficiency of CR-S, fix any strict $\succsim$-nonnull $E \in \Sigma$, consider the contra positive statement: not ML updating implies not CR-S.

Let C_E be the closed and convex set of posteriors representing the conditional preference $\succsim_E$. Not ML updating implies that $C_E \neq \{p/p(E) : p \in C^*(E)\}$. In other words, either there exists $\tilde{p} \in C^*(E)$ such that $\tilde{p}/\tilde{p}(E) \notin C_E$, or there exists $q \in C_E$ such that $q \notin \{p/p(E) : p \in C^*(E)\}$ or both.

Not CR-S means that there exists an $f \in \mathcal{F}$ and $x \in X$ with $f \sim_E x$ such that, for all $\bar{x} \in X$, there always exists $x^* \succsim \bar{x}$ with $f_E x^* \nsim x_E x^*$.

For the two different cases of not ML, since both C_E and $\{p/p(E) : p \in C^*(E)\}$ are convex and closed set, the same type of separating hyperplane argument can be applied to both cases. Thus the proof here only shows the implication of the first case, while the same argument applies to the other case.

Formally, in the first case, there exists $\tilde{p} \in C^*(E)$ such that $\tilde{p}/\tilde{p}(E) \notin C_E$. As $\{\tilde{p}/\tilde{p}(E)\}$ is compact and C_E is convex and closed, the geometric form of Hahn-Banach theorem implies that there exists an act $f \in \mathcal{F}$ such that[25]

$$\int_E u(f)\frac{d\tilde{p}}{\tilde{p}(E)} < \min_{p \in C_E} \int_E u(f)dp$$

Then first as $\tilde{p} \in C^*(E)$, one has

$$\min_{p \in C^*(E)} \int_E u(f)\frac{dp}{p(E)} \leq \int_E u(f)\frac{d\tilde{p}}{\tilde{p}(E)}$$

[25] See Theorem 1.7 of Brezis (2010) for a reference.

Second, for any $x \in X$, $f \sim_E x$ implies that $\min_{p \in C_E} \int_E u(f)dp = u(x)$. These inequalities and equality together imply that

$$\min_{p \in C^*(E)} \int_E u(f) \frac{dp}{p(E)} < u(x)$$

On the other hand by Lemma A.3, there always exists an $\bar{x}_{E,f}$ such that $\bar{x}_{E,f} \succsim x$ and for all $x^* \succsim \bar{x}_{E,f}$,

$$\min_{p \in C} \int_\Omega u(f_E x^*)dp = \min_{p \in C^*(E)} \int_\Omega u(f_E x^*)dp$$

Then for any $x^* \succsim \bar{x}_{E,f}$, the following is true:

$$\begin{aligned}
\min_{p \in C} \int_\Omega u(f_E x^*)dp &= \min_{p \in C^*(E)} \left[\int_E u(f) \frac{dp}{p(E)} \cdot p(E) + u(x^*)(1 - p(E)) \right] \\
&= p^*(E) \min_{p \in C^*(E)} \int_E u(f) \frac{dp}{p(E)} + u(x^*)(1 - p^*(E)) \\
&< p^*(E) \cdot u(x) + u(x^*)(1 - p^*(E)) \\
&= \min_{p \in C} \int_\Omega u(x_E x^*)dp
\end{aligned}$$

Namely, if $f \sim_E x$ then $f_E x^* \prec x_E x^*$ for all $x^* \succsim \bar{x}_{E,f}$. That is, for this $f \in \mathcal{F}$, for any $\bar{x} \in X$, there always exists $x^* \succsim \bar{x}$ such that $f_E x^* \prec x_E x^*$. Therefore, CR-S is not true.

The argument for the second case is analogously the same and combining both cases shows that not ML updating implies not CR-S. □

A.2 Proof of Theorem 3.3

The necessity of CR-UO is given in the main text. For DC-CS, reversing the arguments in step 4 of proving sufficiency shows that equation (A.10) is necessary under Contingent RML updating. It then immediately implies that DC-CS needs to be true.

For the sufficiency, fix any strict $\succsim$-nonnull $E \in \Sigma$, the proof proceeds by the following steps:

Step 1. Show that for all $f \in \mathcal{F}$ there exists $\alpha[E, f] \in [0, 1]$ such that, for all sufficiently large consequence $x^* \in X$ one has

$$\alpha[E, f]U(f_E x^*) + (1 - \alpha[E, f])U(f_E x) = \alpha[E, f]U(x_E x^*) + (1 - \alpha[E, f])U(x) \tag{A.2}$$

Furthermore, $\alpha[E, f]$ is unique if either $f_E x \prec x$ or $f_E x^* \succ x_E x^*$ hold.

When CR-UO is true, $f \sim_E x$ implies $f_E x \precsim x$ and there exists $\bar{x}_{E,f} \in X$ such that $f_E x^* \succsim x_E x^*$ for all $x^* \succsim \bar{x}_{E,f}$.

First consider all $x^* \in X$ such that $x^* \succsim \bar{x}_{E,f}$ and $x^* \succsim x$, then the following inequalities hold: $f_E x^* \succsim x_E x^* \succsim x \succsim f_E x$, which further implies that, *for each* x^*, there always exists an $\alpha[E, f] \in [0, 1]$ such that the following equation holds: (see figure 3)

$$\alpha[E, f]U(f_E x^*) + (1 - \alpha[E, f])U(f_E x) = \alpha[E, f]U(x_E x^*) + (1 - \alpha[E, f])U(x)$$

Notice that, $\alpha[E, f]$ here may depend on the value of x^* because the act $f_E x^*$ could be evaluated at different extreme points for different x^*. However, in the case where $f_E x^*$ is always evaluated at the extreme points in $C^*(E)$, as both $U(f_E x^*)$ and $U(x_E x^*)$ have the common term $u(x^*)(1 - p^*(E))$ which cancels out, $\alpha[E, f]$ does not depend on the value of x^* any more.

By Lemma A.2 and A.3, for all $f \in \mathcal{F}$, there exists another threshold $\hat{x}_{E,f}$ such that for all $x^* \succsim \hat{x}_{E,f}$, one has

$$\min_{p \in C} \int_\Omega u(f_E x^*) dp = \min_{p \in C^*(E)} \int_\Omega u(f_E x^*) dp$$

Therefore, there exists $\alpha[E, f]$ such that the following is true for all $x^* \succsim \max\{\bar{x}_{E,f}, x, \hat{x}_{E,f}\}$:[26]

$$\alpha[E, f]U(f_E x^*) + (1 - \alpha[E, f])U(f_E x) = \alpha[E, f]U(x_E x^*) + (1 - \alpha[E, f])U(x)$$

Furthermore, it is easy to see that this $\alpha[E, f]$ is unique if either $f_E x \prec x$ or $f_E x^* \succ x_E x^*$ hold.

Step 2. Construction of acts satisfying the premises of DC-CS.

Equivalence Class. First, for any two acts f, f', denote them by $f \equiv_E f'$ if $f(\omega) = f'(\omega)$ for all $\omega \in E$. Then an equivalence class of acts can be accordingly defined:

$$[f] = \{f' \in \mathcal{F} : f' \equiv_E f\}$$

By step 1, $\alpha[E, f] = \alpha[E, f']$ whenever $f' \in [f]$. Hereafter, I use f to denote the whole class of acts $[f]$, as it should cause no confusion.

When $\alpha[E]$ is not unique. If there does not exist any $f \in \mathcal{F}$ with $\alpha[E, f]$ being unique. Then by step 1, for all $f \in \mathcal{F}$ and $x \in X$ with $f \sim_E x$, it implies that both $f_E x \sim x$ and $f_E x^* \sim x_E x^*$ for all sufficiently large x^* hold.

Then it is the case in which both CR-C and CR-S hold at the same time. Namely, the conditional preference $\succsim_E$ can be represented by both FB and ML. As E is strict $\succsim$-nonnull, it implies that $C = C^*(E)$. In other words, $C \neq C^*(E)$ implies there exists at least an $f \in \mathcal{F}$ such that $\alpha[E, f]$ is unique.

When unique $\alpha[E, f]$ exists, fix some $f \in \mathcal{F}$ such that $\alpha[E, f]$ is unique.

$\alpha[E, f]$ is unchanged under mixture with constant acts. For any $y \in X$, for any $\lambda \in (0, 1]$ consider the act $f_\lambda y = \lambda f + (1 - \lambda)y$, which is an Anscombe-Aumann mixture of acts[27]. By certainty independence, $f \sim_E x$ implies that $f_\lambda y \sim_E \lambda x + (1 - \lambda)y$. Let $x_\lambda y \in X$ denote the consequence indifferent to $\lambda x + (1 - \lambda)y$. Then for all x^* that are sufficiently large for f and $f_\lambda y$, equation (A.2) for $f_\lambda y$ implies that

$$\alpha[E, f_\lambda y]U([f_\lambda y]_E x^*) + (1 - \alpha[E, f_\lambda y])U([f_\lambda y]_E x_\lambda y) = \alpha[E, f_\lambda y]U([x_\lambda y]_E x^*) + (1 - \alpha[E, f_\lambda y])U(x_\lambda y) \tag{A.3}$$

Moreover, $\alpha[E, f_\lambda y]$ is unique. Notice that

$$U([f_\lambda y]_E x_\lambda y) = U([f_E x]_\lambda y) = \lambda U(f_E x) + (1 - \lambda)U(y)$$

[26]Throughout this proof, whenever I say "for all x^* that are sufficiently good for some acts", it means that for each one of those acts, say f, $x^* \succsim \max\{\bar{x}_{E,f}, x, \hat{x}_{E,f}\}$ holds.

[27]The case $\lambda = 0$ is excluded since when $\lambda = 0$, $f_\lambda y$ coincides with y, and then $\alpha[E, y]$ is not unique.

where the first equality follows from Anscombe-Aumann mixture, the second equality follows from certainty independence.

Then one can write equation (A.3) as

$$\alpha[E, f_\lambda y] \min_{p\in C^*(E)} \int_E u(f_\lambda y)dp + (1-\alpha[E, f_\lambda y])[\lambda U(f_E x) + (1-\lambda)U(y)]$$
$$= \alpha[E, f_\lambda y]u(x_\lambda y)p^*(E) + (1-\alpha[E, f_\lambda y])[\lambda U(x) + (1-\lambda)U(y)]$$

which further implies that

$$\alpha[E, f_\lambda y]U(f_E x^*) + (1-\alpha[E, f_\lambda y])U(f_E x) = \alpha[E, f_\lambda y]U(x_E x^*) + (1-\alpha[E, f_\lambda y])U(x)$$

As $f \sim_E x$ one also has

$$\alpha[E, f]U(f_E x^*) + (1-\alpha[E, f])U(f_E x) = \alpha[E, f]U(x_E x^*) + (1-\alpha[E, f])U(x)$$

Since $\alpha[E, f]$ is unique, it has to be the case that $\alpha[E, f_\lambda y] = \alpha[E, f]$ for all $\lambda \in (0, 1]$ and $y \in X$.

Construction of acts satisfying premises of DC-CS. Recall that the utility function $u(\cdot)$ is normalized such that $u(X) = (\underline{u}, \infty)$ for some $\underline{u} \in \mathbb{R}_{<0} \cup \{-\infty\}$, as X is unbounded from above.

For the fixed $f \in \mathcal{F}$, for any $\epsilon > 0$, there always exists an act f_ϵ given by taking mixtures between f and some constant act such that $u(f_\epsilon(\omega)) \in [0, \epsilon]$ for all $\omega \in \Omega$. For example, taking mixtures with the constant act z such that $u(z) = \epsilon/2$. By previous argument, one has $\alpha[E, f_\epsilon] = \alpha[E, f]$.

Take any act $g \in \mathcal{F}$ that cannot be obtained from f by taking mixtures with constant acts. (If can, then $\alpha[E, g] = \alpha[E, f]$) For any $\lambda \in (0, 1]$ and $y \in X$, consider the act $g_\lambda y$, i.e. the λ mixture between g and y.

Then for the two premises of DC-CS[28], it suffices to find $\lambda \in (0, 1]$ and $y \in X$ such that the following two equations hold:

$$U(f_{\epsilon E} x) = U([g_\lambda y]_E x) \tag{A.4}$$

for $x \sim_E f_\epsilon$ and

$$U(f_{\epsilon E} x^*) = U([g_\lambda y]_E x^*) \tag{A.5}$$

for all x^* that is sufficiently good for f_ϵ and $g_\lambda y$. The threshold for x^* could be chosen after f_ϵ and $g_\lambda y$ are pinned down such that both $U(f_{\epsilon E} x^*)$ and $U([g_\lambda y]_E x^*)$ will both be evaluated at extreme points in $C^*(E)$.

Consider the extreme points of the set of posteriors of $C^*(E)$, let q_f and q_g denote the two of them evaluating f_ϵ and $g_\lambda y$ respectively, then from equation (A.5) one can derive

$$U(f_{\epsilon E} x^*) = U([g_\lambda y]_E x^*)$$
$$\min_{p\in C} \int_\Omega u(f_{\epsilon E} x^*)dp = \min_{p\in C} \int_\Omega u([g_\lambda y]_E x^*)dp$$
$$\min_{p\in C^*(E)} \int_\Omega u(f_{\epsilon E} x^*)dp = \min_{p\in C^*(E)} \int_\Omega u([g_\lambda y]_E x^*)dp$$

[28]Notice here f_ϵ takes the role of f and $g_\lambda y$ takes the role of g in the statement of this axiom.

$$\min_{p\in C^*(E)}\int_E u(f)dp+u(x^*)(1-p^*(E))=\min_{p\in C^*(E)}\int_E u(g_\lambda y)dp+u(x^*)(1-p^*(E))$$
$$\min_{p\in C^*(E)}\int_E u(f)dp=\min_{p\in C^*(E)}\int_E u(g_\lambda y)dp$$
$$\min_{p\in C^*(E)}\int_E u(f)\frac{dp}{p^*(E)}=\min_{p\in C^*(E)}\int_E u(g_\lambda y)\frac{dp}{p^*(E)}$$

i.e.

$$u(f_\epsilon)\cdot q_f=\lambda u(g)\cdot q_g+(1-\lambda)u(y)$$

Thus, for each $\lambda\in(0,1]$, the constant act y could be pinned down by letting

$$(1-\lambda)u(y)=u(f_\epsilon)\cdot q_f-\lambda u(g)\cdot q_g \tag{A.6}$$

and y is arbitrary if $\lambda=1$.

Next consider equation (A.4):

$$\min_{p\in C}\int_\Omega u(f_{\epsilon E}x)dp=\min_{p\in C}\left\{\lambda\int_E u(g)dp+(1-\lambda)u(y)p(E)+(1-p(E))u(x)\right\}$$

Consider the RHS of this equation and plug into $(1-\lambda)u(y)$ from equation (A.6) yields

$$\begin{aligned}&\min_{p\in C}\left\{\lambda\int_E u(g)dp+[u(f_\epsilon)\cdot q_f-\lambda u(g)\cdot q_g]p(E)+(1-p(E))u(x)\right\}\\&=\min_{p\in C}\left\{u(f_\epsilon)\cdot q_fp(E)+(1-p(E))u(x)+\lambda\left[\int_E u(g)dp-u(g)\cdot q_gp(E)\right]\right\}\end{aligned}$$

i.e.

$$\min_{p\in C}\int_\Omega u(f_{\epsilon E}x)dp=\min_{p\in C}\left\{u(f_\epsilon)\cdot q_fp(E)+(1-p(E))u(x)+\lambda\left[\int_E u(g)dp-u(g)\cdot q_gp(E)\right]\right\} \tag{A.7}$$

Therefore, equation (A.4) holds if there exists $\lambda\in(0,1]$ that solves equation (A.7). Notice that the LHS of equation (A.7) does not depend on λ, meanwhile the RHS is a continuous function of λ, which I further denoted it by $R(\lambda)$.

Given continuity, the existence of a solution to equation (A.7) can be proved by showing that $R(0)>$ LHS and $R(1)<$ LHS.

First consider $R(0)$:

$$\begin{aligned}R(0)&=\min_{p\in C}\{u(f_\epsilon)\cdot q_fp(E)+(1-p(E))u(x)\}\\&\geq\min_{p\in C}\{u(x)p(E)+(1-p(E))u(x)\}\\&=u(x)\\&>\min_{p\in C}\int_\Omega u(f_{\epsilon E}x)dp=\text{LHS}\end{aligned}$$

where the second and forth inequality comes from $f_{\epsilon E}x^* \succsim x_E x^*$ and $x \succsim f_{\epsilon E}x$ respectively. The forth inequality is strict because of the fact that $\alpha[E, f]$ is unique, thus one of the inequalities has to be strict.

Next consider $R(1)$:

$$\begin{aligned} R(1) &= \min_{p\in C}\left\{u(f_\epsilon)\cdot q_f p(E) + (1-p(E))u(x) + \int_E u(g)dp - u(g)\cdot q_g p(E)\right\} \\ &\leq \epsilon + \min_{p\in C}\left\{\int_E u(g)dp - u(g)\cdot q_g p(E)\right\} \\ &= \epsilon + \min_{p\in C}\left\{p(E)\cdot\left[\int_E u(g)\frac{dp}{p(E)} - u(g)\cdot q_g\right]\right\} \end{aligned}$$

where the inequality follows from $u(f_\epsilon) \leq \epsilon$.

For the second term, its minimum is 0 if it is the case $\min_{p\in C}\int_E u(g)\frac{dp}{p(E)} - u(g)\cdot q_g = 0$. In this case, notice that conditional evaluation of g coincides under FB and ML updating. That is, both CR-C and CR-S holds for g, then $\alpha[E, g]$ is not unique. It suffices to let $\alpha[E, g] = \alpha[E, f]$.

On the other hand, if it is the case $\min_{p\in C}\int_E u(g)\frac{dp}{p(E)} - u(g)\cdot q_g < 0$, the minimum of the second term is negative. Then it suffices to find $\epsilon > 0$ such that

$$\epsilon < -\min_{p\in C}\left\{p(E)\cdot\left[\int_E u(g)\frac{dp}{p(E)} - u(g)\cdot q_g\right]\right\}$$

Given this ϵ, it further implies that

$$R(1) < \epsilon - \epsilon = 0 \leq \min_{p\in C}\int_\Omega u(f_{\epsilon E}x)dp = \text{LHS}$$

Therefore, the existence of $\lambda \in (0, 1)$ that solves equation (A.7) is guaranteed.

Finally, once λ is solved, $u(y)$ is given by equation (A.6):

$$u(y) = \frac{u(f_\epsilon)\cdot q_f - \lambda u(g)\cdot q_g}{1-\lambda} \geq -\frac{\lambda}{1-\lambda}u(g)\cdot q_g$$

where the last inequality comes from $u(f_\epsilon(\omega)) \geq 0$. It remains to show that $u(y) > \underline{u}$ to guarantee the existence of this construction. It suffices to transform g by taking mixtures with constant acts before the construction to get $u(g)\cdot q_g \leq 0$, then it would imply $u(y) \geq 0 > \underline{u}$ as desired.

Step 3. DC-CS implies $\alpha[E, f]$ to be a constant across all $f \in \mathcal{F}$, and it is unique if $C \neq C^*(E)$.

In the following, I abuse notation to use f to denote f_ϵ and g to denote $g_\lambda y$ be the pair of acts constructed in the last step satisfying the two premises of DC-CS.

By step 1, $f \sim_E x$ implies that

$$\alpha[E, f]U(f_E x^*) + (1-\alpha[E, f])U(f_E x) = \alpha[E, f]U(x_E x^*) + (1-\alpha[E, f])U(x) \tag{A.8}$$

DC-CS implies $g \sim_E x$ and thus,

$$\alpha[E, g]U(g_E x^*) + (1-\alpha[E, g])U(g_E x) = \alpha[E, g]U(x_E x^*) + (1-\alpha[E, g])U(x) \tag{A.9}$$

Consider the LHS of equation (A.9) and denote it by L:

$$\begin{aligned} L &= \alpha[E,g]U(g_E x^*) + (1-\alpha[E,g])U(g_E x) \\ &= \alpha[E,f]U(g_E x^*) + (1-\alpha[E,f])U(g_E x) + [\alpha[E,g]-\alpha[E,f]](U(g_E x^*) - U(g_E x)) \\ &\equiv L' + [\alpha[E,g]-\alpha[E,f]]M_1 \end{aligned}$$

Meanwhile the RHS of equation (A.9) denoted by R can be further derived as

$$\begin{aligned} R &= \alpha[E,g]U(x_E x^*) + (1-\alpha[E,g])U(x) \\ &= \alpha[E,f]U(x_E x^*) + (1-\alpha[E,f])U(x) + [\alpha[E,g]-\alpha[E,f]](U(x_E x^*) - U(x)) \\ &\equiv R' + [\alpha[E,g]-\alpha[E,f]]M_2 \end{aligned}$$

Notice that by equation (A.8), R' also equals to $(1-\alpha[E,f])U(f_E x) + \alpha[E,f]U(f_E x^*)$. Thus $L' = R'$ as $f_E x \sim g_E x$ and $f_E x^* \sim g_E x^*$ hold.

Then the fact $L = R$ implies

$$\begin{aligned} & L - R = L' + [\alpha[E,g]-\alpha[E,f]]M_1 - R' - [\alpha[E,g]-\alpha[E,f]]M_2 = 0 \\ & \Rightarrow L' - R' = [\alpha[E,f]-\alpha[E,g]][M_1 - M_2] \end{aligned}$$

Further notice that

$$\begin{aligned} M_1 - M_2 &= [U(g_E x^*) - U(g_E x)] - [U(x_E x^*) - U(x)] \\ &= \left[\min_{p\in C}\int_\Omega u(g_E x^*)dp - \min_{p\in C}\int_\Omega u(g_E x)dp\right] - [u(x)p^*(E) + u(x^*)(1-p^*(E)) - u(x)] \\ &= \left[\min_{p\in C^*(E)}\int_\Omega u(g_E x^*)dp - \min_{p\in C}\int_\Omega u(g_E x)dp\right] - [u(x^*) - u(x)](1-p^*(E)) \\ &= \left[\min_{p\in C^*(E)}\int_\Omega u(g_E x^*)dp - \min_{p\in C^*(E)}\int_\Omega u(g_E x)dp\right] - [u(x^*) - u(x)](1-p^*(E)) \\ &+ \left[\min_{p\in C^*(E)}\int_\Omega u(g_E x)dp - \min_{p\in C}\int_\Omega u(g_E x)dp\right] \\ &= \min_{p\in C^*(E)}\int_\Omega u(g_E x)dp - \min_{p\in C}\int_\Omega u(g_E x)dp \end{aligned}$$

where the last equality follows from

$$\begin{aligned} & \min_{p\in C^*(E)}\int_\Omega u(g_E x^*)dp - \min_{p\in C^*(E)}\int_\Omega u(g_E x)dp \\ &= \left[\min_{p\in C^*(E)}\int_E u(g)\frac{dp}{p(E)} - \min_{p\in C^*(E)}\int_E u(g)\frac{dp}{p(E)}\right] p^*(E) + [u(x^*) - u(x)](1-p^*(E)) \end{aligned}$$

Next, consider the following lemma:

Lemma A.4. *For any $f \in \mathcal{F}$ such that $\alpha[E,f]$ is unique, if $f \sim_E x$ then*

$$\min_{p\in C}\int_\Omega u(f_E x)dp < \min_{p\in C^*(E)}\int_\Omega u(f_E x)dp$$

Proof of Lemma A.4. Since if $\alpha[E, f]$ is unique, it implies that for $f \sim_E x$, one of the inequalities $f_E x^* \succsim x_E x^*$ and $x \succsim f_E x$ is strict. Notice that the first inequality implies

$$\min_{p \in C^*(E)} \int_\Omega u(f) dp \geq u(x) p^*(E)$$

and the second implies that

$$u(x) \geq \min_{p \in C} \int_\Omega u(f_E x) dp$$

Add the term $u(x)(1 - p^*(E))$ to both sides of the first inequality, then combine both inequalities and recall that one of them has to be strict yield:

$$\min_{p \in C^*(E)} \int_\Omega u(f) dp + u(x)(1 - p^*(E)) > \min_{p \in C} \int_\Omega u(f_E x) dp$$

which is equivalent to $\min_{p \in C^*(E)} \int_\Omega u(f_E x) dp > \min_{p \in C} \int_\Omega u(f_E x) dp$. □

Define

$$\Delta f_E x \equiv \min_{p \in C^*(E)} \int_\Omega u(f_E x) dp - \min_{p \in C} \int_\Omega u(f_E x) dp$$

If $\alpha[E, g]$ is not unique, then it suffice to let $\alpha[E, g] = \alpha[E, f]$. If it is also unique, given Lemma A.4, $g \sim_E x$ implies that

$$\Delta g_E x > 0$$

Therefore the difference

$$L' - R' = [\alpha[E, f] - \alpha[E, g]][M_1 - M_2] = [\alpha[E, f] - \alpha[E, g]] \cdot \Delta g_E x$$

is 0 if and only if $\alpha[E, f] = \alpha[E, g]$.

Notice that the construction in step 2 can be applied to any $g \in \mathcal{F}$. It then implies that $\alpha[E, f]$ needs to be a constant across all $f \in \mathcal{F}$. Therefore, equation (A.2) now can be written as

$$\alpha[E] U(f_E x^*) + (1 - \alpha[E]) U(f_E x) = \alpha[E] U(x_E x^*) + (1 - \alpha[E]) U(x) \tag{A.10}$$

Step 4. Equation (A.10) implies that the DM's conditional evaluation of any $f \in \mathcal{F}$ can be represented by

$$\min_{p \in C_{\alpha[E]}(E)} \int_E u(f) \frac{dp}{p(E)} = u(x)$$

i.e. is given by RML updating with $\alpha[E]$.

For all sufficiently good x^*, the LHS of equation (A.10) can further be derived as:

$$\begin{aligned} &\alpha[E] U(f_E x^*) + (1 - \alpha[E]) U(f_E x) \\ = &\alpha[E] \min_{p \in C} \int_\Omega u(f_E x^*) dp + (1 - \alpha[E]) \min_{p \in C} \int_\Omega u(f_E x) dp \end{aligned}$$

$$
\begin{aligned}
&= \alpha[E] \min_{q \in C^*(E)} \int_\Omega u(f_E x^*) dq + (1-\alpha[E]) \min_{p \in C} \int_\Omega u(f_E x) dp \\
&= \alpha[E] \left[\min_{q \in C^*(E)} \int_E u(f) \frac{dq}{p^*(E)} \cdot p^*(E) + (1 - p^*(E)) u(x^*) \right] \\
&\quad + (1-\alpha[E]) \min_{p \in C} \int_\Omega u(f_E x) dp \\
&= \alpha[E] \left[\min_{q \in C^*(E)} \int_E u(f) \frac{dq}{p^*(E)} \cdot p^*(E) + (1 - p^*(E)) u(x) \right] \\
&\quad + \alpha[E][u(x^*) - u(x)][1 - p^*(E)] + (1-\alpha[E]) \min_{p \in C} \int_\Omega u(f_E x) dp
\end{aligned}
$$

where the second equality follows from $f_E x^*$ is evaluated at some $p \in C^*(E)$.

On the other hand, the RHS of equation (A.10) can also be derived as

$$
\begin{aligned}
&\alpha[E] U(x_E x^*) + (1-\alpha[E]) U(x) \\
&= \alpha[E] \min_{p \in C} \int_\Omega u(x_E x^*) dp + (1-\alpha[E]) u(x) \\
&= \alpha[E][u(x) p^*(E) + u(x^*)(1 - p^*(E))] + (1-\alpha[E]) u(x) \\
&= \alpha[E][u(x^*) - u(x)][1 - p^*(E)] + u(x)
\end{aligned}
$$

Observe that now equalizing the LHS and RHS of equation (A.10) implies

$$
\begin{aligned}
u(x) &= \alpha[E] \left[\min_{q \in C^*(E)} \int_E u(f) \frac{dq}{p^*(E)} \cdot p^*(E) + (1 - p^*(E)) u(x) \right] + (1-\alpha[E]) \min_{p \in C} \int_\Omega u(f_E x) dp \\
&= \alpha[E] \min_{q \in C^*(E)} u(f_E x) dq + (1-\alpha[E]) \min_{p \in C} \int_\Omega u(f_E x) dp \\
&= \min_{q \in C^*(E)} \min_{p \in C} \int_\Omega u(f_E x) d((\alpha[E]) q + (1-\alpha[E]) p) \\
&= \min_{p \in C_{\alpha[E]}(E)} \int_\Omega u(f_E x) dp
\end{aligned}
$$

where $C_{\alpha[E]}(E) = \alpha[E] C^*(E) + (1-\alpha[E]) C$.

From the last equality one can further derive

$$
\begin{aligned}
0 &= \min_{p \in C_{\alpha[E]}(E)} \int_\Omega u(f_E x) dp - u(x) \\
&= \min_{p \in C_{\alpha[E]}(E)} \int_\Omega [u(f_E x) - u(x)] dp \\
&= \min_{p \in C_{\alpha[E]}(E)} \int_E [u(f) - u(x)] dp \\
&= \min_{p \in C_{\alpha[E]}(E)} \left[\int_E u(f) dp - u(x) p(E) \right]
\end{aligned}
$$

When E is strict $\succsim$-nonnull, $p(E) > 0$ for all $p \in C_{\alpha[E]}(E)$, then the last equality further implies

$$0 = \min_{p \in C_{\alpha[E]}(E)} \left[\int_E u(f) \frac{dp}{p(E)} - u(x) \right]$$

i.e.

$$\min_{p \in C_{\alpha[E]}(E)} \int_E u(f) \frac{dp}{p(E)} = u(x)$$

which represents the conditional evaluation of f under $\succsim_E$ since $f \sim_E x$.

Therefore, the conditional preference $\succsim_E$ for each strict $\succsim$-nonnull event E is given by RML updating with $\alpha[E]$.

□

A.3 Proof of Theorem 3.4

Given Theorem 3.3, the only remaining proof here is to show that $\alpha[E]$ is a constant across all events if and only if the EC axiom holds.

The necessity of the EC axiom is also immediate when one plugs a constant α into equation (A.10).

In the following I show that EC implies $\alpha[E]$ to be a constant across all strict $\succsim$-nonnull events

First of all, consider the case that there does not exist any strict $\succsim$-nonnull $E \in \Sigma$ such that $C \neq C^*(E)$. Then it implies that all $p \in C$ agree with the probability of all strict $\succsim$-nonnull event E.

Thus, if there exists only one strict $\succsim$-nonnull event E such that $C \neq C^*(E)$, i.e. $\alpha[E]$ is unique, then it suffices to let this $\alpha[E]$ to be the constant α across all events.

Next, when there exists at least two strict $\succsim$-nonnull events, E_1 and E_2, such that both $\alpha[E_1]$ and $\alpha[E_2]$ are unique.

Similar to the construction in the proof of Theorem 3.3, fix any $f \in \mathcal{F}$. For any $\epsilon > 0$, let f_ϵ denote the act given by taking mixtures between f and some constant act such that $u(f_\epsilon(\omega)) \in [0, \epsilon]$ for all $\omega \in \Omega$.

Take any $g \in \mathcal{F}$, let $g_\lambda y$ denote the act given by taking mixtures between g and $y \in X$ with $\lambda \in (0, 1]$.

Then for the two premises of EC, it suffices to find λ and y such that the following equations hold:

$$U(f_{\epsilon E_1} x) = U([g_\lambda y]_{E_2} x) \tag{A.11}$$

for $x \sim_{E_1} f_\epsilon$ and

$$U(f_{\epsilon E_1} x_1^*) = U([g_\lambda y]_{E_2} x_2^*) \tag{A.12}$$

for all sufficiently large x_1^* and x_2^* with $x_{E_1} x_1^* \sim x_{E_2} x_2^*$.

As the threshold for x_1^* and x_2^* could be chosen after f_ϵ and $g_\lambda y$ are pinned down, thus both $U(f_{\epsilon E_1} x_1^*)$ and $U([g_\lambda y]_{E_2} x_2^*)$ can be guaranteed to be evaluated at an extreme point of $C^*(E_1)$ and an extreme point of $C^*(E_2)$ respectively.

Then from equation (A.12) one can further derive

$$\min_{p\in C}\int_{\Omega} u(f_{\epsilon E_1}x_1^*)dp = \min_{p\in C}\int_{\Omega} u([g_\lambda y]_E x_2^*)dp$$
$$\Rightarrow \min_{p\in C^*(E_1)}\int_{E_1} u(f_\epsilon)dp + u(x_1^*)(1-p^*(E_1)) = \min_{p\in C^*(E_2)}\int_{E_2} u(g_\lambda y)dp + u(x_2^*)(1-p^*(E_2))$$

Let q_f and q_g be the two extreme points in the set of posteriors of $C^*(E_1)$ and $C^*(E_2)$ that evaluate f_ϵ and $g_\lambda y$ respectively, then the last equality can be written as:

$$u(f_\epsilon)\cdot q_f + u(x_1^*)(1-p^*(E_1)) = u(g_\lambda y)\cdot q_g + u(x_2^*)(1-p^*(E_2))$$

Furthermore as the condition $x_{E_1}x_1^* \sim x_{E_2}x_2^*$ implies that

$$u(x)p^*(E_1) + u(x_1^*)(1-p^*(E_1)) = u(x)p^*(E_2) + u(x_2^*)(1-p^*(E_2))$$

i.e.

$$u(x_1^*)(1-p^*(E_1)) - u(x_2^*)(1-p^*(E_2)) = u(x)[p^*(E_2)-p^*(E_1)]$$

As E_1 and E_2 are chosen arbitrarily, without loss of generality, let $p^*(E_2)-p^*(E_1)\geq 0$ and also notice that $u(x)\geq 0$. In the following, let $M\equiv u(x)[p^*(E_2)-p^*(E_1)]$. Notice that $M\in[0,\epsilon]$.

Then equation (A.11) is further equivalent to

$$u(f_\epsilon)\cdot q_f + M = \lambda u(g)\cdot q_g + (1-\lambda)u(y) \tag{A.13}$$

Therefore, for each $\lambda\in(0.1]$, equation (A.12) holds if $(1-\lambda)u(y)$ is given by the following

$$(1-\lambda)u(y) = u(f_\epsilon)\cdot q_f + M - \lambda u(g)\cdot q_g \tag{A.14}$$

and y is arbitrary if $\lambda=1$.

Next consider the equation (A.11)

$$\min_{p\in C}\int_{\Omega} u(f_{\epsilon E_1}x)dp = \min_{p\in C}\left\{\int_{E_2}\lambda u(g)dp + (1-\lambda)u(y)p(E_2) + u(x)(1-p(E_2))\right\}$$

Plugging into $(1-\lambda)u(y)$ from equation (A.14) to the RHS and denote it by $R(\lambda)$:

$$\begin{aligned}R(\lambda) &= \min_{p\in C}\left\{\int_{E_2}\lambda u(g)dp + u(f_\epsilon)\cdot q_f p(E_2) + Mp(E_2) - \lambda u(g)\cdot q_g p(E_2) + u(x)(1-p(E_2))\right\}\\ &= \min_{p\in C}\left\{u(f_\epsilon)\cdot q_f p(E_2) + u(x)(1-p(E_2)) + Mp(E_2) + \lambda\left[\int_{E_2} u(g)dp - u(g)\cdot q_g p(E_2)\right]\right\}\end{aligned}$$

Again, $R(\lambda)$ is a continuous function of λ and the LHS of equation (A.11) is a constant of λ. Thus it suffices to show $R(0) >$ LHS and $R(1) <$ LHS.

For $R(0)$ one has,

$$\begin{aligned}R(0) &= \min_{p\in C}\{u(f_\epsilon)\cdot q_f p(E_2) + u(x)(1-p(E_2)) + Mp(E_2)\}\\ &\geq \min_{p\in C}\{u(f_\epsilon)\cdot q_f p(E_2) + u(x)(1-p(E_2))\}\end{aligned}$$

$$\geq \min_{p\in C} \{u(x)p(E_2) + u(x)(1-p(E_2))\}$$
$$= u(x)$$
$$> \min_{p\in C} \int_\Omega u(f_{\epsilon E_1} x)dp = \text{LHS}$$

On the other hand for $R(1)$,

$$R(1) = \min_{p\in C} \left\{ u(f_\epsilon)\cdot q_f p(E_2) + u(x)(1-p(E_2)) + Mp(E_2) + \left[\int_{E_2} u(g)dp - u(g)\cdot q_g p(E_2)\right]\right\}$$
$$\leq 2\epsilon + \min_{p\in C}\left\{\int_{E_2} u(g)dp - u(g)\cdot q_g p(E_2)\right\}$$

For the second term as $\alpha[E_2]$ is unique, its minimum is negative. Therefore, it suffices to find $\epsilon > 0$ such that

$$\epsilon < -\frac{1}{2}\min_{p\in C}\left\{\int_{E_2} u(g)dp - u(g)\cdot q_g p(E_2)\right\}$$

Then given this ϵ one has

$$R(1) < 2\epsilon - 2\epsilon = 0 \leq \text{LHS}$$

Therefore, the existence of $\lambda \in (0,1)$ that solves equation (A.14) is guaranteed.

Finally, once λ is solved, $u(y)$ is given by equation (A.14):

$$u(y) = \frac{u(f_\epsilon)\cdot q_f + M - \lambda u(g)\cdot q_g}{1-\lambda} \geq -\frac{\lambda}{1-\lambda}u(g)\cdot q_g$$

where the last inequality comes from $u(f_\epsilon(\omega)) + M \geq 0$. It remains to show that $u(y) \geq \underline{u}$ when $\underline{u}$ exists to guarantee the existence of this construction. It suffices to transform g by taking mixtures with constant acts such that $u(g)\cdot q_g \leq 0$.

From this point on, apply exactly the same argument in step 3 of the proof of Theorem 3.3 would imply that $\alpha[E_1] = \alpha[E_2]$. Therefore, $\alpha[E]$ needs to be a constant across all strict $\succsim$-nonnull $E \in \Sigma$. □

Appendix B ML with Infinitely Many Extreme Points

The characterization results in the main text rely on the assumption that the set of priors C has finitely many extreme points. As mentioned, such an assumption helps simplify the axioms while conveys the same intuition. This appendix provides an axiomatization of ML without this assumption. The axiomatizations of RML in this case could be achieved in a similar manner.

Without this assumption, the ex-ante preference $\succsim$ is only assumed to admit a MEU representation. The set of priors C could have infinitely many extreme points. Except for this, every other assumption on the primitive $\{\succsim_E\}_{E\in\Sigma}$ is the same as in Section 2.

By definition, $\{\succsim_E\}_{E\in\Sigma}$ is represented by ML updating if for all strict $\succsim$-nonnull $E\in\Sigma$ and for all $f\in\mathcal{F}$:

$$\min_{p\in C_E}\int_\Omega u(f)dp = \min_{p\in C^*(E)}\int_E u(f)\frac{dp}{p(E)}$$

For the current setting where C may contain infinitely many extreme points, consider the following axiom:

Axiom Approximate CR-S (Approximate **C**ontingent **R**easoning given **S**ufficiently good consequences).

For all $f\in\mathcal{F}$and for all $x,z,w\in X$ with $z\succ w$, there exists $\bar{x}_{E,f,z,w}$ such that for all $x^*\in X$ with $x^*\succsim\bar{x}_{E,f,z,w}$, if $f\sim_E x$, then

$$\frac{1}{2}f_Ex^*+\frac{1}{2}w\prec\frac{1}{2}x_Ex^*+\frac{1}{2}z$$

and

$$\frac{1}{2}f_Ex^*+\frac{1}{2}z\succ\frac{1}{2}x_Ex^*+\frac{1}{2}w$$

Notice that CR-S implies Approximate CR-S. On the other hand, in the case where CR-S is silent as $\bar{x}_{E,f}$ does not exist, Approximate CR-S imposes an additional restriction on the behaviors. It restricts that the difference between f_Ex^* and x_Ex^* should be arbitrarily small when x^* is sufficiently good.

Hence, Approximate CR-S conveys essentially the same intuition as CR-S. The following representation theorem shows that Approximate CR-S is equivalent to ML in the current setting.

Theorem B.1. *$\{\succsim_E\}_{E\in\Sigma}$ is represented by ML updating if and only if Approximate CR-S holds for all strict $\succsim$-nonnull events E.*

Proof of Theorem B.1. First consider the following lemma:

Lemma B.2. *For any strict $\succsim$-nonnull $E\in\Sigma$ and $f\in\mathcal{F}$, for any $\epsilon>0$ there exists $\bar{x}_{E,f,\epsilon}\in X$ such that*

$$\min_{p\in C^*(E)}\int_\Omega u(f_Ex^*)dp-\min_{p\in C}\int_\Omega u(f_Ex^*)dp<\epsilon$$

for all $x^*\succsim\bar{x}_{E,f,\epsilon}$.

Proof of Lemma B.2. For any strict $\succsim$-nonnull $E \in \Sigma$ and $f \in \mathcal{F}$, either there exists $\bar{x}_{E,f}$ such that

$$\min_{p \in C} \int_\Omega u(f_E x^*) dp = \min_{p \in C^*(E)} \int_\Omega u(f_E x^*) dp$$

for all $x^* \succsim \bar{x}_{E,f}$ or not. If it is the first case, then this lemma is trivially true.

Consider the case there does not exist $\bar{x}_{E,f}$ for some E and f. For each $x \in X$, let p_x be the probability measure in C that evaluates the act $f_E x$ according to MEU, i.e. $p_x \equiv \arg\min_{p \in C} \int_\Omega u(f_E x) dp$. Let q denote the probability measure in $C^*(E)$ that evaluates the act $f_E x$ and notice that it does not depend on the value of x.

Then one has

$$\begin{aligned} &\min_{p \in C^*(E)} \int_\Omega u(f_E x) dp - \min_{p \in C} \int_\Omega u(f_E x) dp \\ = &\int_E u(f) dq + u(x)(1 - p^*(E)) - \int_E u(f) dp_x - u(x)(1 - p_x(E)) \end{aligned}$$

Take derivative with respect to $u(x)$ and apply envelope theorem yields

$$\frac{d}{du(x)} \left[\min_{p \in C^*(E)} \int_\Omega u(f_E x) dp - \min_{p \in C} \int_\Omega u(f_E x) dp \right] = p_x(E) - p^*(E)$$

The current assumption $p_x \notin C^*(E)$ implies that $p_x(E) - p^*(E) < 0$, i.e. the difference is decreasing with respect to $u(x)$. Furthermore, since the difference is bounded below by zero, monotone convergence theorem implies that

$$\min_{p \in C^*(E)} \int_\Omega u(f_E x) dp - \min_{p \in C} \int_\Omega u(f_E x) dp \to 0$$

for $u(x) \to \infty$, i.e. the lemma holds.

□

For the necessity of Approximate CR-S, recall Lemma A.1 implies that if $f \sim_E x$ then under ML updating

$$\min_{p \in C^*(E)} \int_\Omega u(f_E x^*) dp = \min_{p \in C} \int_\Omega u(x_E x^*) dp$$

Thus for any $\epsilon > 0$ there exists $\bar{x}_{E,f,\epsilon}$ such that

$$\min_{p \in C} \int_\Omega u(x_E x^*) dp - \min_{p \in C} \int_\Omega u(f_E x^*) dp < \epsilon$$

for all $x^* \succsim \bar{x}_{E,f,\epsilon}$. Then for each $z \succ w$, it suffices to let $\epsilon = u(z) - u(w)$ and then Approximate CR-S axiom holds.

For the sufficiency, fix any strict $\succsim$-nonnull E and consider the contra positive statement: not ML updating implies not Approximate CR-S axiom. Not ML means that $C_E \neq \{p/p(E) : p \in C^*(E)\}$. In other words, either there exists $\tilde{p} \in C^*(E)$ such that $\tilde{p}/\tilde{p}(E) \notin C_E$ or there exists $q \in C_E$ such that $q \notin \{p/p(E) : p \in C^*(E)\}$ or both.

Consider the first case, by the same argument implied by the Hahn-Banach Theorem in the proof of Theorem 2.4, there exists $f \in \mathcal{F}$ and $x \in X$ such that $f \sim_E x$ and

$$\min_{p \in C^*(E)} \int_E u(f) \frac{dp}{p(E)} < u(x)$$

Then for all $x^* \succsim x$,

$$\min_{p \in C} \int_\Omega u(x_E x^*) dp - \min_{p \in C^*(E)} \int_\Omega u(f_E x^*) dp = u(x) p^*(E) - \min_{p \in C^*(E)} \int_E u(f) dp > 0$$

That is, there exists $\delta_{E,f} > 0$ such that

$$\min_{p \in C} \int_\Omega u(x_E x^*) dp - \min_{p \in C^*(E)} \int_\Omega u(f_E x^*) dp > \delta_{E,f} > 0$$

Now for all $x^* \succsim x$,

$$\begin{aligned}
&\min_{p \in C} \int_\Omega u(x_E x^*) dp - \min_{p \in C} \int_\Omega u(f_E x^*) dp \\
&= \min_{p \in C} \int_\Omega u(x_E x^*) dp - \min_{p \in C^*(E)} \int_\Omega u(f_E x^*) dp + \min_{p \in C^*(E)} \int_\Omega u(f_E x^*) dp - \min_{p \in C} \int_\Omega u(f_E x^*) dp \\
&> \delta_{E,f} > 0
\end{aligned}$$

the first inequality comes from the fact that $\min\limits_{p \in C^*(E)} \int_\Omega u(f_E x^*) dp - \min\limits_{p \in C} \int_\Omega u(f_E x^*) dp \geq 0$. Then it suffices to find $z \succ w$ such that $u(z) - u(w) < \delta_{E,f}$ and it will imply that for all $x^* \succsim x$:

$$\frac{1}{2} f_E x^* + \frac{1}{2} z \prec \frac{1}{2} x_E x^* + \frac{1}{2} w$$

i.e. the Approximate CR-S axiom fails.

Now consider the second case, there exists $q \in C_E$ such that $q \notin \{p/p(E) : p \in C^*(E)\}$. By Hahn-Banach theorem, there exists $f \sim_E x$ such that

$$u(x) = \min_{p \in C_E} \int_\Omega u(f) dp \leq \int_\Omega u(f) dq < \min_{p \in C^*(E)} \int_E u(f) \frac{dp}{p(E)}$$

Then it further implies for all $x^* \succsim x$,

$$\min_{p \in C^*(E)} \int_\Omega u(f_E x^*) dp - \min_{p \in C} \int_\Omega u(x_E x^*) dp = \min_{p \in C^*(E)} \int_E u(f) dp - u(x) p^*(E) > 0$$

i.e. there exists $\delta_{E,f}$ such that

$$\min_{p \in C^*(E)} \int_\Omega u(f_E x^*) dp - \min_{p \in C} \int_\Omega u(x_E x^*) dp > \delta_{E,f} > 0$$

On the other hand, by Lemma B.2, for any $\epsilon > 0$ there exists $\bar{x}_{E,f,\epsilon}$such that

$$\min_{p \in C^*(E)} \int_\Omega u(f_E x^*) dp - \min_{p \in C} \int_\Omega u(f_E x^*) dp < \epsilon$$

for all $x^* \succsim \bar{x}_{E,f,\epsilon}$. Now for all $x^* \succsim \max\{x, \bar{x}_{E,f,\epsilon}\}$,

$$\begin{aligned}
&\min_{p\in C}\int_\Omega u(f_E x^*)dp - \min_{p\in C}\int_\Omega u(x_E x^*)dp \\
&= \min_{p\in C}\int_\Omega u(f_E x^*)dp - \min_{p\in C^*(E)}\int_\Omega u(f_E x^*)dp + \min_{p\in C^*(E)}\int_\Omega u(f_E x^*)dp - \min_{p\in C}\int_\Omega u(x_E x^*)dp \\
&> -\epsilon + \delta_{E,f}
\end{aligned}$$

Now find any $z \succ w$ such that $u(z) - u(w) = \eta < \delta_{E,f}$ and let $\epsilon = \delta_{E,f} - \eta > 0$. Then for all $x^* \succsim \max\{x, \bar{x}_{E,f,\epsilon}\}$ the previous result implies that

$$\min_{p\in C}\int_\Omega u(f_E x^*)dp - \min_{p\in C}\int_\Omega u(x_E x^*)dp > -\epsilon + \delta_{E,f} = \eta > 0$$

Then

$$\frac{1}{2}\min_{p\in C}\int_\Omega u(f_E x^*)dp + \frac{1}{2}u(w) - \min_{p\in C}\int_\Omega u(x_E x^*)dp - \frac{1}{2}u(z) > \frac{1}{2}[\eta - \eta] = 0$$

i.e. $\frac{1}{2}f_E x^* + \frac{1}{2}w \succ \frac{1}{2}x_E x^* + \frac{1}{2}z$ for all $x^* \succsim \bar{x}_{E,f,\epsilon}$. Thus the Approximate CR-S axiom fails in this case as well. Combine both cases shows that, not ML implies not Approximate CR-S.

□

Remark. The characterization results for Contingent RML and RML can be extended similarly to this more general case. For example, CR-UO can be approximated by requiring the difference between $f_E x^*$ and some $x'_E x^*$ with $x'_E x^* \succsim x_E x^*$ to be arbitrarily small as x^* goes to infinity. The statement of the axioms and also the additional steps in the proof will be analogously the same, thus omitted in this paper.

Appendix C Experimental results in Liang (2019)

I include the table 4.2 in Liang (2019) mentioned in the main text for reference. Because of the difference in terminologies, I first summarize the one-to-one mapping from his terms to the parameters I used in example 2:

Prior	β
(Midpoint) Information accuracy	$(\lambda_1 + \lambda_2)/2$
Good news	s_1
Bad news	s_2
Type of information (simple)	Probabilistic signals
Type of information (ambiguous)	Ambiguous signals

Table 4.2 in Liang (2019) is shown in the following. The Mean conditional CE represents the evaluation of a bet on the state θ_1. Thus, for example in the first row of the following table, the Mean conditional CE under the ambiguous information is 10.35, *lower* than under the simple information 10.80. This is consistent with the prediction summarized by table 1 (the row with $\beta < 1/2$ and s_1).

Prior	(Midpoint) Information accuracy	Good/Bad news	Type of information	Mean conditional CE	Standard error	N
30%	70%	good	simple	10.80	0.645	54
			compound	9.48	0.603	44
			ambiguous	10.35	0.739	47
40%	60%	good	simple	10.19	0.533	91
			compound	8.96	0.491	85
			ambiguous	10.10	0.490	60
50%	70%	good	simple	12.51	0.362	164
			compound	11.99	0.349	164
			ambiguous	10.88	0.346	165
60%	60%	good	simple	12.45	0.391	73
			compound	12.10	0.463	80
			ambiguous	9.61	0.452	105
70%	70%	good	simple	14.74	0.369	111
			compound	13.45	0.397	121
			ambiguous	13.74	0.381	118
30%	70%	bad	simple	5.70	0.368	111
			compound	5.38	0.355	121
			ambiguous	5.48	0.345	118
40%	60%	bad	simple	6.89	0.400	73
			compound	7.89	0.490	80
			ambiguous	5.95	0.390	105
50%	70%	bad	simple	6.47	0.345	165
			compound	6.99	0.306	163
			ambiguous	6.93	0.314	165
60%	60%	bad	simple	7.46	0.496	91
			compound	7.48	0.431	85
			ambiguous	9.58	0.474	60
70%	70%	bad	simple	7.20	0.672	54
			compound	9.70	0.651	44
			ambiguous	9.34	0.720	47
30%	50%		simple	7.10	0.361	163
			compound	7.50	0.374	164
			ambiguous	7.29	0.334	164
70%	50%		simple	10.88	0.336	163
			compound	10.34	0.349	164
			ambiguous	10.21	0.369	163

Table C.1: Table 4.2 in Liang (2019)

References

Beauchêne, D., Li, J., and Li, M. (2019). Ambiguous persuasion. *Journal of Economic Theory*, 179:312–365.

Bose, S. and Renou, L. (2014). Mechanism design with ambiguous communication devices. *Econometrica*, 82(5):1853–1872.

Brezis, H. (2010). *Functional analysis, Sobolev spaces and partial differential equations*. Springer Science & Business Media.

De Filippis, R., Guarinno, A., Jehiel, P., and Kitagawa, T. (2018). Non-bayesian updating in a social learning experiment. *Working paper*.

Dominiak, A., Duersch, P., and Lefort, J.-P. (2012). A dynamic ellsberg urn experiment. *Games and Economic Behavior*, 75(2):625–638.

Ellsberg, D. (1961). Risk, ambiguity, and the savage axioms. *The Quarterly Journal of Economics*, pages 643–669.

Epstein, L. G. and Schneider, M. (2003). Recursive multiple-priors. *Journal of Economic Theory*, 113(1):1–31.

Epstein, L. G. and Schneider, M. (2007). Learning under ambiguity. *The Review of Economic Studies*, 74(4):1275–1303.

Ghirardato, P. (2002). Revisiting savage in a conditional world. *Economic theory*, 20(1):83–92.

Gilboa, I. and Schmeidler, D. (1989). Maxmin expected utility with non-unique prior. *Journal of Mathematical Economics*, 18(2):141–153.

Gilboa, I. and Schmeidler, D. (1993). Updating ambiguous beliefs. *Journal of Economic Theory*, 59(1):33–49.

Gul, F. and Pesendorfer, W. (2017). Evaluating ambiguous random variables and updating by proxy. *Working paper*.

Hanany, E. and Klibanoff, P. (2007). Updating preferences with multiple priors. *Theoretical Economics*, 2(3):261–298.

Hanany, E. and Klibanoff, P. (2009). Updating ambiguity averse preferences. *The BE Journal of Theoretical Economics*, 9(1).

Hill, B. (2019). Updating confidence in beliefs. *Working paper*.

Kamenica, E. and Gentzkow, M. (2011). Bayesian persuasion. *American Economic Review*, 101(6):2590–2615.

Kellner, C. and Le Quement, M. T. (2018). Endogenous ambiguity in cheap talk. *Journal of Economic Theory*, 173:1–17.

Kellner, C., Le Quement, M. T., and Riener, G. (2019). Reacting to ambiguous messages: An experimental analysis. *Working paper*.

Kovach, M. (2015). Ambiguity and partial bayesian updating. *Working paper*.

Li, J. (2015). Preferences for partial information and ambiguity. *Theoretical Economics*.

Liang, Y. (2019). Learning from unknown information sources. *Working paper*.

Maccheroni, F., Marinacci, M., and Rustichini, A. (2006). Ambiguity aversion, robustness, and the variational representation of preferences. *Econometrica*, 74(6):1447–1498.

Machina, M. J. and Schmeidler, D. (1992). A more robust definition of subjective probability. *Econometrica*, (4):745–780.

Ortoleva, P. (2012). Modeling the change of paradigm: Non-bayesian reactions to unexpected news. *American Economic Review*, 102(6):2410–36.

Ortoleva, P. (2014). Hypothesis testing and ambiguity aversion. *Rivista di politica economica*, (7):45–66.

Pires, C. P. (2002). A rule for updating ambiguous beliefs. *Theory and Decision*, 53(2):137–152.

Shishkin, D. and Ortoleva, P. (2019). Ambiguous information and dilation: An experiment. *Working paper*.

Siniscalchi, M. (2006). A behavioral characterization of plausible priors. *Journal of Economic Theory*, 128(1):1–17.

Siniscalchi, M. (2009). Two out of three ain't bad: A comment on the ambiguity aversion literature: A critical assessment. *Economics & Philosophy*, 25(3):335–356.

Siniscalchi, M. (2011). Dynamic choice under ambiguity. *Theoretical Economics*, 6(3):379–421.

Suleymanov, E. (2018). Robust Maximum Likelihood Updating. *Working paper*.

Zhao, C. (2017). Pseudo-Bayesian Updating. *Working paper*.