A Theory of Contraction Updating*

Rui Tang †

July 21, 2022

Abstract

We introduce the contraction rule for updating ambiguous beliefs. With the rule, a realized event renders an individual's belief unambiguous when and only when the event has small ambiguity. The contraction rule is continuous and insensitive to priors that are less likely given the information. The contraction posterior set is independent of the order in which multiple pieces of information arrive. We axiomatize the rule and show that it accommodates recent experimental findings on updating ambiguous information and has robust predictions on individual learning.

KEYWORDS: Ambiguous beliefs; Contraction rule; Dilation of beliefs; Under-reaction to ambiguous information

JEL CODES: D01, D81, D93

PRESENTATION VIDEO: https://youtu.be/R811DT_y9Ig

* I am deeply indebted to Faruk Gul, Pietro Ortoleva and Wolfgang Pesendorfer for their invaluable advice, guidance and encouragement. I thank Roland Bénabou, Sylvain Chassang, Xiaoyu Cheng, Simon Grant, Gaoji Hu, Shaowei Ke, Alessandro Lizzeri, Xiaosheng Mu, John K.-H. Quah, Satoru Takahashi, Jingni Yang, Leeat Yariv, Mu Zhang, and Chen Zhao for helpful comments and discussions. All errors are my own.

† Department of Economics, HKUST, ruitang@ust.hk

1 Introduction

In decision theory, the term ambiguity refers to the situation in which states of the world have no objective probability distribution. Since the seminal work of Knight (1921), Keynes (1921), and Ellsberg (1961), a variety of models have been proposed to rationalize decision makers' (henceforth DM) choices over bets on ambiguous states.[1] In contrast, relatively few papers focus on updating ambiguous beliefs. The growing literature on incorporating ambiguous information into interactive decisions has highlighted the importance of the latter topic.[2]

As in most applications of ambiguity, the DM is a max-min expected utility maximizer (Gilboa and Schmeidler, 1989): she has a set of priors over states of the world and evaluates each prospect according to its minimal expected utility over all her priors. The set of priors, uniquely identified through the DM's choices, is considered as the DM's belief set. In this paper, we offer a new belief updating rule that revises the DM's prior set to her posterior set when new information arrives.

Two updating rules, Full-Bayesian rule (also known as the prior-by-prior rule, henceforth FB) and Maximum Likelihood (also known as the Dempster-Shafer rule, henceforth ML), are most frequently used in updating ambiguous beliefs. A common consequence of FB and ML is that information may dilate the DM's payoff-relevant set of beliefs and thus *increase* her payoff-relevant ambiguity.[3] The dilation of belief sets may occur even when there is no ex ante payoff-relevant ambiguity (Wasserman and Kadane, 1990). Shishkin and Ortoleva (2021) (henceforth SO21) experimentally test this prediction and find that ambiguity averse subjects typically do not dilate their payoff-relevant belief sets after receiving ambiguous information. This finding is inconsistent with both FB and ML. Our new updating rule, the contraction rule, does not create dilation of belief sets when the DM has no ex ante payoff-relevant ambiguity. In Section 4.1, we provide generic conditions under

[1] See, for instance, Gilboa and Schmeidler (1989), Schmeidler (1989), Maccheroni, Marinacci, and Rustichini (2006), Chew and Sagi (2008), Gul and Pesendorfer (2014), etc.

[2] See, for instance, Bose and Renou (2014), Beauchêne, Li, and Li (2019), and Chen (2021), etc.

[3] Shishkin and Ortoleva (2021) consider the case in which there is an urn containing 50 red balls and 50 blue balls, and a ball is randomly drawn from the urn. The color of the drawn ball determines the DM's final payoff. They show that if a DM, who initially does not observe the color of the drawn ball (so she initially has a flat prior over the color), is provided with a signal that indicates the color of the drawn ball but can be true or false with unknown probabilities, then the DM's posterior set over the color of the drawn ball can strictly contain the flat prior under both FB and ML.

which the contraction rule does not create belief dilation (Proposition 6).

To see how our updating rule works, consider the following example. There is an investor who is deciding whether to buy a stock or not. The investor believes that there is an equal chance for the stock price to go up (U) or down (D), where U and D denote the two states respectively. The investor consults an expert who knows perfectly about the true state. Nevertheless, the investor is *uncertain* about whether the expert tells her the truth or not: she believes that the maximal probability for the expert to lie is equal to the maximal probability for the expert to tell the truth.

Suppose that the expert tells the investor that the stock price would go up. Intuitively, this suggestion can be interpreted symmetrically in two contradicting ways, depending on whether the expert is telling the truth or not. Note also that the investor has an unambiguous prior over the stock price and thus there is no ex ante ambiguity to be resolved by the information. As a result, the investor should ignore the suggestion of the expert and maintain her initial belief.

We note that both FB and ML have inconsistent predictions with the intuition above. By FB or ML, the investor becomes more confused given the suggestion and her belief is rendered ambiguous. In contrast, the contraction rule has consistent predictions with the intuition above. Given the expert's suggestion, an investor who follows the contraction rule would evaluate the maximal likelihoods of U and D separately: U is mostly likely when the expert is telling the truth and D is mostly likely when the expert is lying. The maximal joint probability of state U with the expert's suggestion is equal to that of state D with the suggestion. The contraction rule maintains the ratio of the two maximal probabilities and leads to a flat posterior in this example.

With the contraction rule, whether the information resolves the DM's ambiguity or not depends on the *degree of ambiguity* on the realized event. Specifically, let S be the state space and let a set of priors P over S be the DM's prior set. A piece of information, also called an event, is a non-empty subset E of S. The contraction measure, denoted by $\mu_P|E$, is the one that assigns each state in E its maximal ex ante probability, i.e., $\mu_P|E(s) = \max_{p\in P} p(s)$ for each $s \in E$. We use the sum $\sum_{s\in E} \mu_P|E(s)$ as an indicator of whether the degree of ambiguity on E is large or not. When the degree of ambiguity on E is small ($\sum_{s\in E} \mu_P|E(s) \leq 1$), the realization of E resolves the DM's ambiguity; the DM's posterior maintains the ratio of maximal ex ante likelihoods for each pair of states in E. When the degree of ambiguity on E is large ($\sum_{s\in E}\mu_P|E(s) > 1$), each prior in P is updated *towards* the contraction measure; the DM's ambiguity is not resolved.

We compare the contraction rule with the two benchmark rules, FB and ML, in Section 2.3. With a stylized example, we show that the contraction rule moderates FB and ML in the case where the realized event has large ambiguity: the FB posterior set can be very sensitive to priors that assign near-zero probabilities to the realized event while the contraction posterior set is almost not affected by those priors; the ML completely disregards priors that do not maximize the probability of the realized event while the contraction rule still updates those priors by assigning less weights to them. We then show that the contraction rule satisfies continuity while neither FB nor ML does. In Section 2.4, we further show that the contraction posterior set is unaffected by the arrival order of multiple pieces of information.

In Section 3.1, we introduce evaluation functions. A function that maps states of the world to payoffs is called an act. An evaluation function, V, maps each act to its certainty equivalence and admits a max-min representation. Therefore, a DM's preference over acts can be described by an evaluation function, and her beliefs can be uniquely identified from it.

In Section 3.2, we characterize the contraction rule through evaluation functions. An updating rule Γ (defined over evaluation functions) specifies for each ex ante evaluation function V and each event E an ex post evaluation function V_E. Γ is the contraction rule if V_E identifies the contraction posterior set. We show that the contraction rule can be fully characterized by axioms ambiguity-driven decreased sensitivity after updating, independence of irrelevant states, mixture independence without ambiguity, mixture ambiguity betweenness, and nonincreasing ambiguity by information. While the last axiom, which says that information does not render an unambiguous prior over payoff-relevant states ambiguous, is violated by FB, all other axioms are satisfied by FB.

In Section 4, we provide several applications of the contraction rule. We first study the empirical relevance of the rule with SO21 and Liang (2021) (henceforth L21). We show that a large proportion of their experimental and empirical findings can be addressed by the rule. Then we investigate the learning behavior of a DM in a Wald-type experiment with ambiguous signals. We show that a DM who updates according to the contraction rule learns the true state of the world with probability close to one when provided with a sufficiently large number of independent, informative, and ambiguous signals.

Related Literature. One stream of literature studies how DMs update ambiguous beliefs and the associated economic implications (e.g., Epstein and Schneider (2007, 2008)). Proposed by Jaffray (1988), FB is analyzed by Wasserman and Kadane (1990) and Jaffray (1992) and axiomatized by Pires (2002). Introduced by Dempster (1967) and Shafer (1976), ML is axiomatized by Gilboa and Schmeidler (1993) and Cheng (2022). This paper contributes to the literature by providing the axiomatic foundation for a new updating rule that does not create belief dilation under generic conditions. Similar to the new updating rules proposed by Kovach (2021) and Cheng (2022), both of which nest FB and ML, the contraction rule moderates FB and ML when ambiguity is not fully resolved. While FB updates each prior equally and ML only updates the most likely priors, the contraction rule updates each prior but assigns less weight to those that are less likely. Our approach ensures the continuity of the contraction rule, which is violated by both FB and ML.

Two specific streams of theoretical literature on updating ambiguous beliefs are worth noting. The first one is about dynamic consistency.[4] Epstein and Schneider (2003) show that dynamic consistency is maintained when the DM has "rectangular" sets of priors and updates according to FB. Hanany and Klibanoff (2007, 2009) introduce the dynamic consistency updating rule. Following this rule, the DM figures out an optimal contingent plan based on information she might obtain and updates her beliefs in such a way that her ex post self finds it optimal to follow the contingent plan. The dynamic consistency updating rule violates consequentialism as the ex post beliefs of the DM may depend on unrealized parts of the choice problem. Since the contraction rule updates each prior set to some posterior set supported in the realized event, it satisfies consequentialism and violates dynamic consistency.

Another stream of literature concerns the martingale property of updating ambiguous beliefs. One distinctive paper in this line is by Gul and Pesendorfer (2021), in which they introduce the proxy rule. The core axiom that features the proxy rule is "not all news is bad news": given a prospect and a set of potentially realized signals, there exits one signal of which the realization does not decrease the DM's evaluation of the prospect. Similarly, under generic conditions, the contraction rule does not lead to belief dilation, which implies that "a piece of news cannot be bad for all prospects". That is, the realization of a signal has to weakly increase the DM's evaluation towards at least one uncertain prospect.

[4]See Gilboa and Marinacci (2013) for a comprehensive survey.

In addition, both our rule and the proxy rule predict that information does not render the DM's unambiguous payoff-relevant belief ambiguous. The contraction rule differs from the proxy rule in the preference domain in which the two rules are applied: the proxy rule is used to update max-min preferences that allow for totally monotone capacities, and the contraction rule can be used for updating general max-min preferences.[5]

A variety of experimental studies directly test how subjects react to ambiguous information, including Cohen, Gilboa, Jaffray, and Schmeidler (2000), Dominiak, Duersch, and Lefort (2012), Ert and Trautmann (2014), Moreno and Rosokha (2015), Kellner, Le Quement, and Riener (2019), Epstein and Halevy (2021), etc. A key contribution of our model is to provide consistent predictions with the experimental and empirical findings by SO21 and L21. In particular, we show that the contraction rule does not render the DM's unambiguous payoff-relevant belief ambiguous (SO21) and predicts under-reaction to ambiguous information (L21). There are other findings by SO21 and L21 that can also be accommodated by our model, which are discussed in Sections 4.1 and 4.2.

A recently emerging stream of literature applies ambiguous information to interactive decisions among multiple players such as Blume and Board (2014), Bose and Renou (2014), Kellner and Le Quement (2017), Kellner and Le Quement (2018), Beauchêne, Li, and Li (2019), Chen (2021), etc. Since the two most commonly adopted rules (FB and ML) create dilation of belief sets, the contraction rule can be considered as an alternative approach for those applications to disentangle the implications by belief dilation from those by other features of ambiguity.

The rest of the paper is organized as follows. We present the contraction rule in Section 2 and characterize it in Section 3. Section 4 contains all applications of our model. All omitted proofs and examples can be found in Appendix A, and a characterization of FB is provided in Appendix B.

[5]For instance, in the Online Appendix A of SO21, they show that the set of priors satisfying their assumptions cannot induce a totally monotone capacity. Thus the proxy rule cannot be directly applied to accommodate their findings.

2 Model

2.1 Preliminary

Let S be the state space that contains infinite states. Let $\mathcal{M}(S)$ be the set of all measures over S that have finite supports, i.e., a measure π is an element of $\mathcal{M}(S)$ if and only if there exists finite $\hat{S} \subseteq S$ such that $\pi(S \backslash \hat{S}) = 0$. Fix $\mathcal{M}$ to be an arbitrary non-empty subset of $\mathcal{M}(S)$. $\mathcal{M}$ is finitely supported if there exists finite $\hat{S} \subseteq S$ such that $\pi(S \backslash \hat{S}) = 0$ for all $\pi \in \mathcal{M}$. If $\mathcal{M}$ is finitely supported and $\sup_{\pi \in \mathcal{M}} \pi(\{s\}) < +\infty$ for all $s \in S$, then the measure $\mu_{\mathcal{M}} \in \mathcal{M}(S)$ is defined such that $\mu_{\mathcal{M}}(\{s\}) = \sup_{\pi \in \mathcal{M}} \pi(\{s\})$ for each $s \in S$. $\mu_{\mathcal{M}}$ is the lowest upper bound of all measures in $\mathcal{M}$. For any $\hat{S} \subseteq S$ and any $\pi \in \mathcal{M}(S)$, define $\pi|\hat{S} \in \mathcal{M}(S)$ such that $(\pi|\hat{S})(E) = \pi(E \cap \hat{S})$ for all $E \subseteq S$; let $\mathcal{M}|\hat{S} = \{\pi|\hat{S} : \pi \in \mathcal{M}\}$. For any finite partition $\Pi = \{S_i\}_{i=1}^n$ of S and any $\pi \in \mathcal{M}(S)$, π_Π is defined as the measure induced by π over the algebra generated by Π such that $\pi_\Pi(S_i) = \pi(S_i)$ for each $i \in \{1, ..., n\}$; let $\mathcal{M}_\Pi = \{\pi_\Pi : \pi \in \mathcal{M}\}$.

For any $\pi, \pi' \in \mathcal{M}(S)$ and $\alpha, \beta \in \mathbb{R}$, if $\alpha\pi(\{s\}) + \beta\pi'(\{s\}) \geq 0$ for all $s \in S$, then define $\alpha\pi + \beta\pi' \in \mathcal{M}(S)$ such that for each $E \subseteq S$, $(\alpha\pi + \beta\pi')(E) = \alpha\pi(E) + \beta\pi'(E)$. This notion can be extended to two sets of measures such that $\alpha\mathcal{M} + \beta\mathcal{M}' = \{\alpha\pi + \beta\pi' : \pi \in \mathcal{M}, \pi' \in \mathcal{M}'\}$ if each $\alpha\pi + \beta\pi'$ is a well-defined element in $\mathcal{M}(S)$. For any non-empty $\mathcal{M} \subseteq \mathcal{M}(S)$, let $co(\mathcal{M})$ be the convex hull of $\mathcal{M}$, i.e., $\pi \in co(\mathcal{M})$ if and only if there exists a non-empty and finite set $\{\pi_i\}_{i=1}^n \subseteq \mathcal{M}$ and a set of non-negative numbers $\{\alpha_i\}_{i=1}^n$ with $\sum_{i=1}^n \alpha_i = 1$ such that $\pi = \sum_{i=1}^n \alpha_i \pi_i$.

Let $\Delta(S) \subseteq \mathcal{M}(S)$ be the set of all finitely supported probability measures over S. We assume that all priors and posteriors of the DM are elements of $\Delta(S)$. For any $\pi \in \mathcal{M}(S)$ with $\pi(S) > 0$, define $\overline{\pi} \in \Delta(S)$ as the normalized probability measure of π such that $\overline{\pi}(E) = \frac{\pi(E)}{\pi(S)}$ for all $E \subseteq S$. For any $\pi, \pi' \in \mathcal{M}(S)$ with $\pi(S) \leq 1$ and $\pi'(S) > 1$, define $\Phi(\pi, \pi')$ as the unique probability measure in $co(\{\pi, \pi'\})$, i.e.,

$$\Phi(\pi, \pi') = \frac{\pi'(S) - 1}{\pi'(S) - \pi(S)}\pi + \frac{1 - \pi(S)}{\pi'(S) - \pi(S)}\pi'.$$

A set of probability measures P is said to be convex if $\alpha p + (1-\alpha)p' \in P$ for all $p, p' \in P$ and $\alpha \in [0, 1]$. P is said to be closed if it is a closed subset of $\mathbb{R}^S$, where each $p \in P$ is viewed as a vector in $\mathbb{R}^S$. Let $\mathscr{P}$ be the collection of non-empty, finitely supported, convex, and closed sets of probability measures over S.

An event is a non-empty and finite subset of S. Let $\mathcal{S}$ be the collection of all events. For any $P \in \mathscr{P}$ and any event E, E is P-non-null if there exists $p \in P$ such that $p(E) > 0$. E is P-null if it is not P-non-null.

Let $\mathbb{K}$ be the payoff space, which is an interval of $\mathbb{R}$ with non-empty interior. An act is a function $f : S \to \mathbb{K}$, which maps each state to some payoff. We use $x \in \mathbb{K}$ to denote the constant act that equals x everywhere. Let $\mathcal{F}$ denote the set of all acts. For any act f and any $p \in \Delta(S)$, let $\mathbb{E}_p(f) = \sum_{s \in S} p(s) f(s)$; let $\mathbb{E}_P(f) = \inf_{p \in P} \mathbb{E}_p(f)$ for any non-empty $P \subseteq \Delta(S)$. For any two acts f and g and any $E \subseteq S$, we write $f \geq g$ if $f(s) \geq g(s)$ for all $s \in S$, and write $f =^E g$ if $f(s) = g(s)$ for all $s \in E$. If for some $s \in S$, $f(s) > g(s)$ and $f =^{S \setminus \{s\}} g$, then we write $f \rhd^s g$. Let fEg denote the act that equals f on E and equals g on $S \setminus E$. For any $\alpha \in [0, 1]$, $\alpha f + (1 - \alpha) g$ denotes the act satisfying that $(\alpha f + (1 - \alpha) g)(s) = \alpha f(s) + (1 - \alpha) g(s)$ for all $s \in S$. Throughout the paper, we write s for $\{s\}$ whenever there is no confusion.

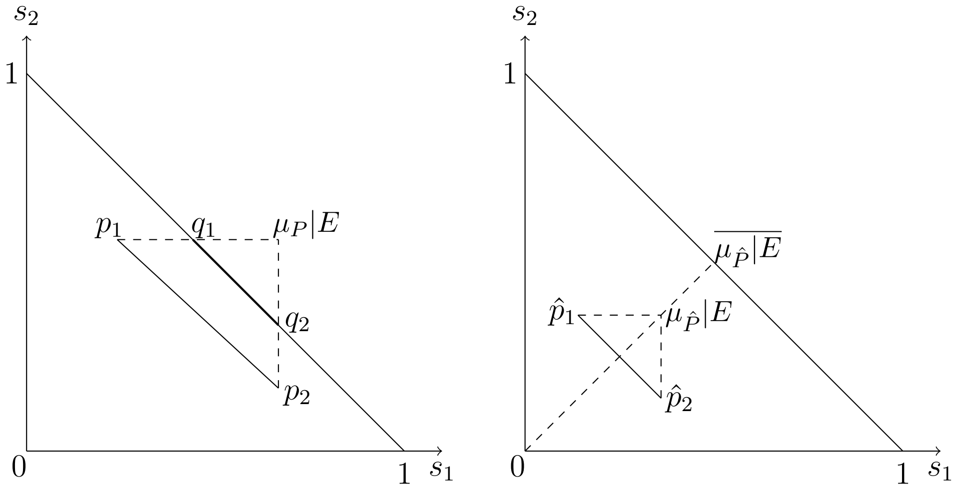
2.2 Contraction Rule

We present the contraction rule in this section. Consider some $P \in \mathscr{P}$ and a P-non-null event E. The contraction posterior set $Q^c(P, E)$ is defined as follows.

$$Q^c(P,E) = \begin{cases} \{\overline{\mu_P|E}\}, & \text{if } \mu_P(E) \leq 1, \\ \{\Phi(p|E, \mu_P|E) : p \in P\}, & \text{if } \mu_P(E) > 1. \end{cases} \tag{1}$$

$\mu_P|E$ is called the *contraction measure*. For each $s \in E$, we have $\mu_P|E(s) = \max_{p \in P} p(s)$, i.e., the contraction measure consists of the maximal likelihood of each state in E. We interpret the contraction rule as that the DM updates each one of her priors *towards* the contraction measure: when $\mu_P(E) > 1$, $\Phi(p|E, \mu_P|E)$ is the unique probability measure between $p|E$ and $\mu_P|E$, and the posterior set is formed by projecting $P|E$ onto the set of probability measures over E along the direction towards $\mu_P|E$; when $\mu_P(E) \leq 1$, each measure $p|E$ is first updated to the contraction measure, and the posterior is given by the normalization of the contraction measure. In Figure 1, we depict the two cases.

We interpret the value of $\mu_P(E)$ as an indicator of the degree of ambiguity on



(a) $\mu_P(E) > 1$ (b) $\mu_{\hat{P}}(E) \leq 1$

Figure 1: For both (a) and (b), all priors are supported in $\{s_1, s_2, s_3\} \subseteq S$. Thus each prior can be depicted in the two-dimensional space, where the horizontal axis denotes the probability of s_1 and the vertical axis denotes the probability of s_2. The realized event E is $\{s_1, s_2\}$ for both (a) and (b). In (a), the prior set P is the line segment between p_1 and p_2, and the contraction posterior set is the line segment between $q_1 = \Phi(p_1|E, \mu_P|E)$ and $q_2 = \Phi(p_2|E, \mu_P|E)$; In (b), the prior set $\hat{P}$ is the line segment between $\hat{p}_1$ and $\hat{p}_2$, and the contraction posterior set is $\{\overline{\mu_{\hat{P}}|E}\}$.

E.[6] When $\mu_P(E) \leq 1$, E has small or no ambiguity. The realization of E renders the DM's belief unambiguous. When $\mu_P(E) > 1$, E has large ambiguity, and thus its realization does not resolve the DM's ambiguity. As shown by the following proposition, when the ambiguity is not resolved, the contraction rule preserves the contraction measure.

Proposition 1. *For any $P \in \mathscr{P}$ and any P-non-null event E, $|Q^c(P, E)| = 1$ if and only if $\mu_P(E) \leq 1$; in the case where $\mu_P(E) > 1$, $\mu_P|E = \mu_{Q^c(P,E)}$.*

We note that requiring the ambiguity to be resolved when $\mu_P(E) \leq 1$ is not *ad hoc*. In fact, it is implied by two fundamental postulates of belief updating: (1) the DM's posterior set is function of $P|E$; (2) if the DM has an unambiguous belief over payoff-relevant states, then new information does not render her belief ambiguous.

Specifically, postulate (1) says that the DM only relies on prior distributions over states within the realized event E to update her beliefs. The prior distributions of states in $S \backslash E$ are irrelevant.[7]

[6]For the measurement of ambiguity of *alternatives*, see, for instance, Izhakian (2020).

[7]One can easily check the both FB and ML satisfy this postulate. Formal definitions of FB and ML can be found in Section 2.3.

To state the second postulate, assume that the state space takes the product structure $S = D \times \Theta$, where D contains all payoff-relevant states and Θ contains all signals. Postulate (2) then says that if the DM's prior set P induces a unique marginal distribution on D, then the DM's ex post belief over D should also be unambiguous no matter what signal is observed. The following example demonstrates why the two postulates imply that $\mu_P(E) = 1$ should be the cutoff for ambiguity resolving.

Example 1. Consider two prior sets $P_1 = co(\{p_1, \hat{p}_1\})$ and $P_2 = co(\{p_2, \hat{p}_2\})$, where all priors in the two prior sets are supported in $\hat{S} = \{d_1, d_2\} \times \{\theta_1, \theta_2\} \subseteq S$. The distributions of $p_1, \hat{p}_1, p_2$, and $\hat{p}_2$ are presented in Table 1. The realized event is $E = \{(d_1, \theta_1), (d_2, \theta_1)\}$, i.e., signal θ_1. □

	(d_1, θ_1)	(d_1, θ_2)	(d_2, θ_1)	(d_2, θ_2)
p_1	0.1	0.2	0.6	0.1
$\hat{p}_1$	0.4	0.1	0.3	0.2
p_2	0.1	0.3	0.6	0
$\hat{p}_2$	0.4	0	0.3	0.3

Table 1: Distributions of $p_1, \hat{p}_1, p_2$, and $\hat{p}_2$ in Example 1

Note that in Example 1, $P_1|E = P_2|E$ (since $p_1|E = p_2|E$ and $\hat{p}_1|E = \hat{p}_2|E$). By postulate (1), P_1 and P_2 should be updated to the same posterior set when E is realized. Also note that for P_2, both p_2 and $\hat{p}_2$ assign probability 0.4 to $\{d_1\} \times \Theta$ and probability 0.6 to $\{d_2\} \times \Theta$. Thus, P_2 is unambiguous on D. By postulate (2), the realization of signal θ_1 does not render the posterior of P_2 over D ambiguous. Hence, both P_1 and P_2 are updated to some singleton posterior set over E. Here, the condition $\mu_{P_1}(E) \leq 1$ is necessary for the existence of such P_2, since otherwise $\mu_{P_2}(E) = \mu_{P_1}(E) > 1$ implies that P_2 cannot be unambiguous on D. It is also easy to see that $\mu_{P_1}(E) \leq 1$ is sufficient for us to construct such P_2.[8] Therefore, postulates (1) and (2) justify the ambiguity resolving part of the contraction rule.

Next, we discuss the case where the information does not resolve the ambiguity. In this case, the posterior set is $Q^c(P, E) = \{\Phi(p|E, \mu_P|E) : p \in P\}$, where

$$\Phi(p|E, \mu_P|E) = \frac{\mu_P(E) - 1}{\mu_P(E) - p(E)} p|E + \frac{1 - p(E)}{\mu_P(E) - p(E)} \mu_P|E. \tag{2}$$

[8] We can always shift probabilities outside of the realized signal to ensure that each payoff-relevant state has a constant ex ante probability. For instance, in Example 1, we can move p_1's probability on (d_2, θ_2) to (d_1, θ_2) to obtain p_2 and $\hat{p}_1$'s probability on (d_1, θ_2) to (d_2, θ_2) to obtain $\hat{p}_2$. By this, we construct P_2 based on P_1.

Two observations based on formula (2) should be noted. First, the mixture weight $\frac{\mu_P(E)-1}{\mu_P(E)-p(E)}$ of $p|E$ is an increasing function of $p(E)$. This sharply captures how the DM evaluates the likelihoods of the priors: if a prior assigns a higher probability to E, then it is considered to be more likely given the information E and is thus weighed more during the updating process. In particular, if $p(E) = 1$, then $\Phi(p|E, \mu_P|E) = p|E$. Hence, a prior is completely preserved when it is fully consistent with the information. Second, the mixture weight of $p|E$ is an increasing function of $\mu_P(E)$. As $\mu_P(E)$ increases, the information becomes less informative and thus the DM relies more on her priors.

Finally, we note that the contraction posterior set is well-behaved: (1) the posterior set $Q^c(P, E)$ is always non-empty, convex, and closed; (2) if the new information is uninformative, then the posterior set equals the prior set.

Proposition 2. *For any $P \in \mathscr{P}$ and any P-non-null event E, $Q^c(P, E) \in \mathscr{P}$; if in addition $p(E) = 1$ for all $p \in P$, then $Q^c(P, E) = P$.*

2.3 Comparison with FB and ML

In this section, we compare the contraction rule with FB and ML using one stylized example. To start with, we formally introduce FB and ML.

For any $P \in \mathscr{P}$ and P-non-null event E, the FB posterior set is defined as

$$Q^{fb}(P, E) = cl\Big(\{\overline{p|E} : p \in P, p(E) > 0\}\Big),$$

where $cl(\cdot)$ denotes the closure of what is inside of the bracket.[9] With FB, the DM updates each prior to its posterior following the Bayes' rule.

For any $P \in \mathscr{P}$ and P-non-null event E, the ML posterior set is defined as

$$Q^{ml}(P, E) = \{\overline{p|E} : p \in P \text{ such that } p(E) \geq \hat{p}(E), \forall \hat{p} \in P\}.$$

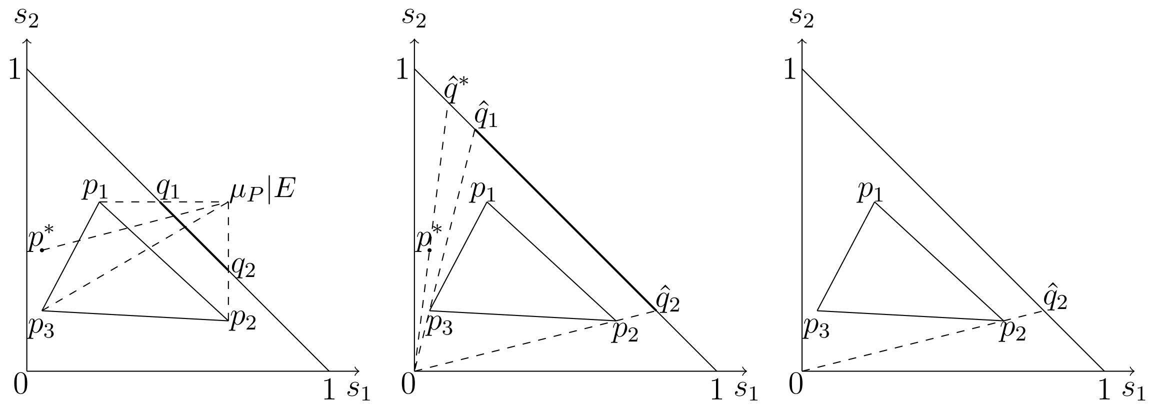
With ML, the DM only updates the priors that maximize the probability of E following the Bayes' rule. Our next example illustrates the differences among FB, ML, and the contraction rule.

[9] The set $\{\overline{p|E} : p \in P, p(E) > 0\}$ might not be closed. For instance, consider $P = \{p \in \Delta(S) : p(\{s_1, s_2, s_3\}) = 1, p^2(s_1) + (1 - p(s_2))^2 \leq 1\}$. One can verify that $P \in \mathscr{P}$ and if $E = \{s_1, s_2\}$, then $\{\overline{p|E} : p \in P, p(E) > 0\} = \{q \in \Delta(S) : q(\{s_1, s_2\}) = 1, q(s_2) > 0\}$, which is not closed.

Example 2. Let $P = co(\{p_1, p_2, p_3\})$ with support $\{s_1, s_2, s_3\} \subseteq S$. Distributions of p_1, p_2, and p_3 are given by

$$p_1(s_1) = 6/25, p_1(s_2) = 14/25, p_1(s_3) = 1/5,$$
$$p_2(s_1) = 2/3, p_2(s_2) = 1/6, p_2(s_3) = 1/6,$$
$$p_3(s_1) = 1/20, p_3(s_2) = 1/5, p_3(s_3) = 3/4.$$

Assume that the realized event is $E = \{s_1, s_2\}$. The contraction posterior set, the FB posterior set, and the ML posterior set, as depicted in Figure 2, are $Q^c(P, E) = co(\{q_1, q_2\})$, $Q^{fb}(P, E) = co(\{\hat{q}_1, \hat{q}_2\})$, and $Q^{ml}(P, E) = \{\hat{q}_2\}$ respectively, where $q_1(s_1) = 1 - q_1(s_2) = 11/25, q_2(s_1) = 1 - q_2(s_2) = 2/3$, $\hat{q}_1(s_1) = 1 - \hat{q}_1(s_2) = 1/5, \hat{q}_2(s_1) = 1 - \hat{q}_2(s_2) = 4/5$. □



(a) contraction posterior (b) FB posterior (c) ML posterior

Figure 2: (a) The triangle $p_1p_2p_3$ constitutes the prior set P and the line segment between q_1 and q_2 is the contraction posterior set; (b) The triangle $p_1p_2p_3$ constitutes the prior set P and the line segment between $\hat{q}_1$ and $\hat{q}_2$ is the FB posterior set; (c) The triangle $p_1p_2p_3$ constitutes the prior set P and $\{\hat{q}_2\}$ is the ML posterior set.

First note that with FB, any prior, regardless of the probability it assigns to the realized event, is updated with the Bayes' rule. Hence, the FB posterior set can be very sensitive to priors that are less likely. To see this, note that in Figure 2 (b), if p_3 shifts to p^*, i.e., the prior set changes from $co(\{p_1, p_2, p_3\})$ to $co(\{p_1, p_2, p^*\})$, then the FB posterior set is enlarged to $co(\{\hat{q}^*, \hat{q}_2\})$. Since priors that assign less probabilities to the realized event are amplified more by the Bayes' rule, changes among those priors can result in significant changes of the FB posterior set.

By contrast, the ML posterior set is a singleton as shown in Figure 2 (c). Different from FB, ML only updates priors that maximize the probability of the realized event, i.e., the priors that are most consistent with the information. As a consequence, ML might rule out many reasonable priors. For instance, in Example 2, prior p_1 only assigns slightly less probability to E than p_2 does. However, it is disregarded by ML all of a sudden when E is realized.

The contraction rule moderates FB and ML in the sense that it updates all priors (when the realized event has large ambiguity) but puts less weights to those that are less consistent with the information. As shown in Figure 2 (a), the shift of p_3 to p^* does not affect the contraction posterior set. Since p^* assigns a small probability to E, formula (2) implies that only a small weight is put on $p^*|E$ when it is updated to its posterior. Thus the contraction rule is not sensitive to less likely priors.

One observation from the above comparison is that neither FB nor ML is continuous: the FB posterior set is very sensitive to priors that assign small probabilities to the realized event; the ML posterior set is unstable when there are multiple priors simultaneously maximizing the probability of the realized event. We show in Appendix A (Examples 5 and 6) that FB is not upper hemicontinuous and ML is not lower hemicontinuous. In contrast, the following proposition states that the contraction rule is continuous.

Proposition 3. *For any sequence of prior sets $\{P_n\}_{n=1}^{+\infty} \subseteq \mathscr{P}$ that have the same finite support $\hat{S} \subseteq S$, if $\{P_n\}_{n=1}^{+\infty}$ converges to $P \in \mathscr{P}$ in Hausdorff metric,*[10] *then for any P-non-null event E, there exists m such that E is P_n-non-null for all $n \geq m$, and $\{Q^c(P_n, E)\}_{n=m}^{+\infty}$ converges to $Q^c(P, E)$ in Hausdorff metric.*

2.4 Divisibility

When a DM has multiple pieces of information, the order in which information arrives may affect the DM's final beliefs. To see this, consider a DM who receives two pieces of information: E and F. If the DM observes E at the first stage and F at the second stage, she might update her beliefs first to some posteriors on E then to some posteriors on $E \cap F$. If the DM observes F at the first stage and E at the second stage, then she might have a different updating path. There is no guarantee that the DM's posterior sets over $E \cap F$ are the same in the two cases. An updating rule is divisible, or path-independent, if the posterior set of

[10] Since $\{P_n\}_{n=1}^{+\infty}$ and P have finite support $\hat{S}$, we view them as subsets of $\mathbb{R}^{\hat{S}}$. Let d be the Euclidean metric in $\mathbb{R}^{\hat{S}}$. The Hausdorff metric d_h for any two compact sets A and B is given by $d_h(A, B) = \max\{\max_{x \in A}(\min_{y \in B} d(x, y)), \max_{z \in B}(\min_{w \in A} d(z, w))\}$.

the DM is unaffected by the order in which several pieces of information arrive. A divisible rule delivers unique predictions on the DM's final beliefs regardless of her information acquisition order. Divisible rules are characterized by Cripps (2019) when there is no ambiguity. In the remaining part of this section, we show that the contraction rule is divisible.

Example 3. Let $P = co(\{p, \hat{p}\})$ with support $\{s_1, s_2, s_3, s_4\} \subseteq S$. The distributions of p and $\hat{p}$ are given by $p(s_1) = 1/3, p(s_2) = 1/3, p(s_3) = 1/6, p(s_4) = 1/6$; $\hat{p}(s_1) = 2/3, \hat{p}(s_2) = 0, \hat{p}(s_3) = 1/3$, and $\hat{p}(s_4) = 0$. Suppose that the DM is first informed of event $E = \{s_1, s_2, s_3\}$. Since $\mu_P(E) = 2/3 + 1/3 + 1/3 = 4/3 > 1$, we have $Q^c(P, E) = co(\{q, \hat{p}\})$, where $q(s_1) = 4/9$, $q(s_2) = 1/3$, and $q(s_3) = 2/9$. Suppose that the DM is then informed of event $F = \{s_1, s_2\}$. Since $\mu_{Q^c(P,E)}(F) = 2/3 + 1/3 = 1$, the contraction posterior set is given by $Q^c(Q^c(P, E), F) = \{\mu_{Q^c(P,E)}|F\} = \{q^*\}$, where $q^*(s_1) = 2/3$ and $q^*(s_2) = 1/3$.

By contrast, suppose that the DM is directly informed of F at the beginning. Since $\mu_P(F) = 2/3 + 1/3 = 1$, her contraction posterior set is given by $Q^c(P, F) = \{\mu_P|F\} = \{q^*\}$. It follows that $Q^c(Q^c(P, E), F) = Q^c(P, F)$. □

In Example 3, either the DM is first informed of event E or directly informed of F does not affect her final belief set. The key observation from this example is that $\mu_{Q^c(P,E)}|F = \mu_P|F$ (which is implied by Proposition 1). Since the DM updates her priors towards the contraction measure, the DM's posterior set on F does not change as long as the contraction measure on F remains unchanged. The next proposition states a more general result: if we know that the DM observes two (or multiple) events, then we can assume without loss of generality that the DM observes the two (or multiple) events simultaneously and only updates her beliefs once. Thus the contraction rule is divisible.

Proposition 4. *For any $P \in \mathscr{P}$ and P-non-null events E and F with $F \subseteq E$, $Q^c(Q^c(P, E), F) = Q^c(P, F)$.*

Note that FB is also a divisible rule since it updates each prior with the Bayes' rule. The divisibility of FB is driven by the same geometric feature as the contraction rule: the origin—the measure that assigns each state a measure of zero—serves as a contraction point, and each prior is updated away from the contraction point. Clearly, the contraction point of FB remains unchanged after updating.

3 Axiomatic Foundation

In this section, we first introduce evaluation functions. The DM's beliefs can be uniquely identified by her evaluation function. We then provide an axiomatic foundation for the contraction rule based on DMs' evaluation functions.

3.1 Evaluation Function

The DM's set of beliefs over S can be identified through her choices over acts. We characterize her choices through an evaluation function, which is a map $V : \mathcal{F} \to \mathbb{K}$ such that for any $x \in \mathbb{K}$, any $f, g, h, l \in \mathcal{F}$ and any $\alpha \in (0, 1)$:

(1) (Identity): $V(x) = x$.

(2) (Continuity): $\{t \in [0, 1] : V(tf + (1 - t)g) \geq V(th + (1 - t)l)\}$ is closed.

(3) (Certainty Independence): $V(f) > V(g)$ if and only if $V(\alpha f + (1 - \alpha)x) > V(\alpha g + (1 - \alpha)x)$.

(4) (Monotonicity): $f \geq g$ implies $V(f) \geq V(g)$.

(5) (Uncertainty Aversion): $V(f) = V(g)$ implies $V(\frac{1}{2}f + \frac{1}{2}g) \geq V(f)$.

(6) (Finite Support): There exists a non-empty and finite subset $\hat{S} \subseteq S$ such that $V(f) = V(g)$ for all f and g with $f =^{\hat{S}} g$.

Conditions (1)-(5) are slightly modified from the axioms by Gilboa and Schmeidler (1989). Together with condition (6), the conditions are sufficient and necessary for V to have a max-min representation with finite support. That is, there exists $P \in \mathscr{P}$ such that $V(f) = \mathbb{E}_P(f)$ for each act f. Such a set of priors P is said to represent V. Let $\mathcal{V}$ be the set of all evaluation functions. Each element of $\mathcal{V}$ uniquely corresponds to an element in $\mathscr{P}$, and vice versa. Thus the set of evaluation functions characterizes all possible belief sets of the DM.

When some event E is realized, the DM revises her beliefs. She updates her set of priors to a set of posteriors. Her set of posteriors can be identified through her ex post evaluation function.

Now, we are ready to define updating rules over evaluation functions. We only consider events that happen with non-zero probabilities.[11] For this purpose, we define non-null events for a given evaluation function. A set $E \subseteq S$ is said to be

[11] We do not model how DMs react to unexpected information in this paper. For theories of updating events with zero probability, see, for example, Ortoleva (2012).

V-null if $V(f) = V(g)$ for all $f, g \in \mathcal{F}$ satisfying $f =^{S \setminus E} g$. E is V-non-null if it is not V-null. It is easy to see that if an evaluation function V is represented by P, then E is V-null if and only if E is P-null. Let $\mathcal{S}_V$ be the set of all V-non-null events.

Define $\mathcal{G} \subseteq \mathcal{V} \times \mathcal{S}$ such that $(V, E) \in \mathcal{G}$ if and only if $E \in \mathcal{S}_V$. By this definition, a tuple (V, E) is contained in $\mathcal{G}$ if and only if E is V-non-null.

Definition 1. *An updating rule over evaluation functions is a function* $\Gamma : \mathcal{G} \to \mathcal{V}$ *such that* (i) $S \setminus E$ *is* $\Gamma(V, E)$*-null, and* (ii) *if* $S \setminus E$ *is* V*-null, then* $\Gamma(V, E) = V$.

By Definition 1, an updating rule maps each evaluation function V and one of its non-null events E to a new evaluation function $\Gamma(V, E)$, where V characterizes the DM's ex ante beliefs and $\Gamma(V, E)$ captures the DM's ex post beliefs after E is realized. Condition (i) requires that states outside of the realized event should not affect the DM's ex post evaluation. Condition (ii) says that when there is essentially no new information, the DM does not revise her beliefs.

Definition 2. *An updating rule* Γ *over evaluation functions is the contraction rule (respectively FB and ML) if for any* $(V, E) \in \mathcal{G}$ *with* V *being represented by* P*,* $\Gamma(V, E)$ *is represented by* $Q^c(P, E)$ *(respectively* $Q^{fb}(P, E)$ *and* $Q^{ml}(P, E)$*).*

Discussion of the updating framework. First, we restrict our attention to max-min DMs. This is aligned with the main objective of this paper, i.e., investigating how new information shapes DMs' multiple beliefs. As will be clear later, the axioms that we use to characterize our updating rule do not necessarily require the DM's preference to admit a max-min representation. We leave the study of our axioms in other models of ambiguity for future work.

Second, updating rules are defined over the set of all possible evaluation functions. We provide two interpretations for the rich choice domain we adopt. The first interpretation is that we can observe choices made by the DM in various choice scenarios and aggregate those choices to the universal state space S. For instance, suppose that we observe the DM's choices as well as how she reacts to new information in a specific choice scenario with a finite state space $\hat{S}$. We can then define an *arbitrary* injection $\phi : \hat{S} \to S$ to map the DM's preferences over acts on $\hat{S}$ (before and after information) to her preferences over acts on S with support $\phi(\hat{S})$. An implicit assumption under this interpretation is that for any two choice scenarios with state space S_1 and S_2 respectively, if there is an isomorphism $\psi : S_2 \to S_1$ such that the DM's evaluation functions V_1 and V_2 under the two scenarios are equivalent in the sense that $V_1(f) = V_2(\psi(f))$ for all acts $f : S_1 \to \mathbb{K}$[12], then how she reacts to information should be identical in the two scenarios, i.e., the ex post evaluation function given event $E \subseteq S_1$ in the first scenario and that given event $\psi^{-1}(E) \subseteq S_2$ in the second scenario are equivalent in the same sense. Following this interpretation, the axioms in the next section can be viewed as consistency conditions obeyed by the DM when processing information in different choice scenarios.

Another interpretation is that we observe the choice behavior of the whole population. In this regard, our axioms can be viewed as consistent conditions imposed on the whole population whose preferences allow for max-min representations. When illustrating the axioms in the next section, we adopt the first interpretation.

3.2 Characterization

We provide an axiomatic foundation for the contraction rule in this section. Whenever Γ is clear, we write V_E for $\Gamma(V, E)$.

Increased Sensitivity after Updating (ISU). For any $(V, E) \in \mathcal{G}$ and $f, g \in \mathcal{F}$ with $g \rhd^s f$ for some $s \in E$, $V(f) = V_E(f)$ implies $V(g) \leq V_E(g)$.

Axiom ISU says that the DM becomes more sensitive to payoff differences on state s if the new information (E) does not rule out s. Since $V(f) = V_E(f)$, the condition $V(g) \leq V_E(g)$ implies that $V(g) - V(f) \leq V_E(g) - V_E(f)$, which means that the DM reacts more to the payoff-increase on s after the information.

Axiom ISU is closely related to dynamic consistency. To see this, note that axiom ISU can be restated as: for any $(V, E) \in \mathcal{G}$ and $f, g \in \mathcal{F}$ with $g \rhd^s f$ for some $s \in E$ and $V(f) = V_E(f)$, if $V(\frac{1}{2}f + \frac{1}{2}x) = V(\frac{1}{2}g + \frac{1}{2}y)$ for some $x, y \in \mathbb{K}$, then $V_E(\frac{1}{2}f + \frac{1}{2}x) \leq V_E(\frac{1}{2}g + \frac{1}{2}y)$.[13] The new condition can be interpreted as a version of dynamic consistency with *monotonicity*. Since $V(f) \leq V(g)$, we know $x \geq y$, and thus for each $\hat{s} \in S \backslash E$, $1/2f(\hat{s}) + 1/2x \geq 1/2g(\hat{s}) + 1/2y$. Since $1/2f + 1/2x$

[12] $\psi(f)$ is an act over S_2 such that $\psi(f)(s_2) = f(\psi(s_2))$ for all $s_2 \in S_2$

[13] To see why the statement is equivalent to axiom ISU, assume first that axiom ISU holds. Since $V(f) = V_E(f)$, we have $V(g) \leq V_E(g)$ and thus $V(g) - V(f) \leq V_E(g) - V_E(f)$. If $V(\frac{1}{2}f + \frac{1}{2}x) = V(\frac{1}{2}g + \frac{1}{2}y)$ for some $x, y \in \mathbb{K}$, we have $V(g) - V(f) = x - y$. It follows that $x - y \leq V_E(g) - V_E(f)$. Then we have $V_E(\frac{1}{2}f + \frac{1}{2}x) \leq V_E(\frac{1}{2}g + \frac{1}{2}y)$. The inverse direction can be shown similarly.

	s_1	s_2	s_3	s_4
p	0.1	0.6	0.2	0.1
$\hat{p}$	0.7	0.1	0.1	0.1
f	1	0	10000	x
$\hat{f}$	2	0	10000	x
g	0	1	-10000	y
$\hat{g}$	0	2	-10000	y

Table 2: Distributions of the priors and payoffs of the acts in Example 4

is as good as $1/2g + 1/2y$ at the ex ante stage and the realization of E rules out the states on which $1/2f + 1/2x$ is better than $1/2g + 1/2y$, $1/2f + 1/2x$ becomes weakly worse than $1/2g + 1/2y$ after the realization of E. Nevertheless, as shown by Hanany and Klibanoff (2007), it is impossible to maintain dynamic consistency in a general belief updating framework under ambiguity. The following example shows that axiom ISU also fails when the ambiguity on the realized event is large.

Example 4. Let V be an evaluation function represented by $P = co(\{p, \hat{p}\})$ with support $\{s_1, s_2, s_3, s_4\} \subseteq S$. The distributions of p and $\hat{p}$ are presented in Table 2. Consider four acts f, $\hat{f}$, g, and $\hat{g}$. The payoffs of the acts on $\hat{S}$ are also presented in Table 2. The realized event is $E = \{s_1, s_2\}$. □

We argue that in Example 4, axiom ISU is violated for any updating rule. First, note that for both f and $\hat{f}$, the prior in P that minimizes their expected payoffs is $\hat{p}$, since $\hat{p}$ assigns a smaller probability to state s_3 than p does. $\hat{f}$ differs from f on state s_1 where $\hat{f}$ yields a higher payoff. Thus $V(\hat{f}) - V(f) = \hat{p}(s_1) = 0.7$.

Next, for g and $\hat{g}$, by a similar observation, the prior in P that minimizes their expected payoffs is p. g and $\hat{g}$ differ on state s_2 with $\hat{g}$ yielding a higher payoff. Thus we have $V(\hat{g}) - V(g) = p(s_2) = 0.6$.

Now, consider an arbitrary updating rule and assume that V is updated to V_E when E is realized. We argue that axiom ISU cannot hold. Since $\{s_4\}$ is V_E-null, there exist x and y such that $V(f) = V_E(f)$ and $V(g) = V_E(g)$. Since $\hat{f} \rhd^{s_1} f$ and $\hat{g} \rhd^{s_2} g$, axiom ISU requires that $V_E(\hat{f}) - V_E(f) \geq 0.7$ and $V_E(\hat{g}) - V_E(g) \geq 0.6$. If V_E is represented by some $Q \in \mathscr{P}$, then the above two inequalities imply that $\min_{q \in Q} q(s_1) \geq 0.7$ and $\min_{q \in Q} q(s_2) \geq 0.6$, which are impossible. Hence, axiom ISU is violated by any updating rule.

By Example 4, if the ambiguity of the realized event E is too large, then axiom ISU may not hold. Indeed, when E has large ambiguity, the DM is likely to exhibit excessive sensitivity to each state in E. It then becomes harder for the DM to increase her sensitivity to each state in E simultaneously after the realization of E. In what follows, we introduce a weakening version of axiom ISU. Prior to presenting the new axiom, we need the following definition.

Definition 3. *For any partition $\Pi = \{S_i\}_{i=1}^n$ of S, an evaluation function V is unambiguous with respect to Π if $V(\frac{1}{2}f + \frac{1}{2}g) = V(f)$ for any two acts f and g that are measurable with respect to Π and satisfy $V(f) = V(g)$; otherwise, V is ambiguous with respect to Π. V is unambiguous if $V(\frac{1}{2}f + \frac{1}{2}g) = V(f)$ for any two acts f and g that satisfy $V(f) = V(g)$; otherwise, V is ambiguous.*

By the definition, an evaluation function is unambiguous if and only if the DM does not benefit from hedging.

Ambiguity-driven Decreased Sensitivity after Updating (ADSU). For any $(V, E) \in \mathcal{G}$ and $f, g \in \mathcal{F}$ with $g \rhd^s f$ for some $s \in E$, if $V(f) = V_E(f)$ and $V(g) > V_E(g)$, then V_E is ambiguous.

As we have argued, the violation of axiom ISU is due to the large ambiguity of the realized event. By axiom ADSU, such ambiguity cannot be resolved after updating, and thus the ex post evaluation function V_E is ambiguous.

To proceed, we define mixture over evaluation functions. For any $V, \hat{V} \in \mathcal{V}$ and $\alpha \in [0,1]$, define $\alpha V + (1-\alpha)\hat{V}$ such that $(\alpha V + (1-\alpha)\hat{V})(f) = \alpha V(f) + (1-\alpha)\hat{V}(f)$ for all $f \in \mathcal{F}$.[14]

Mixture Independence (MI). For any $V, \hat{V} \in \mathcal{V}$, $E \in \mathcal{S}_V \backslash \mathcal{S}_{\hat{V}}$, and $\alpha \in (0,1)$, $V_E = (\alpha V + (1-\alpha)\hat{V})_E$.

Since E is $\hat{V}$-null, V and $\alpha V + (1-\alpha)\hat{V}$ share the same structure of ambiguity on E. Axiom MI says that the DM's ex post choice behavior is determined by the ambiguity structure on the realized event. However, since V and $\alpha V + (1-\alpha)\hat{V}$ might have different degrees of ambiguity on E, axiom MI cannot accommodate the situation in which the DM's ex post choices are affected by the degree of ambiguity on the realized event. Thus we consider the following weaker version of axiom MI.

Mixture Independence without Ambiguity (MIA). For any $V, \hat{V} \in \mathcal{V}$, $E \in \mathcal{S}_V \backslash \mathcal{S}_{\hat{V}}$, and $\alpha \in (0,1)$, if V_E is unambiguous, then $V_E = (\alpha V + (1-\alpha)\hat{V})_E$.

[14]Lemma 5 shows that the mixture of two evaluation functions is indeed an evaluation function.

By axiom MIA, the degree of ambiguity on the realized event does not affect the DM's ex post choices if its realization already resolves the DM's ambiguity.

Betweenness (B). For any $V, \hat{V} \in \mathcal{V}$ and $E \in \mathcal{S}_V \cap \mathcal{S}_{\hat{V}}$, if $\max\{V(fEx), V_E(f)\} \geq \hat{V}(fEx) \geq \min\{V(fEx), V_E(f)\}$ for all $f \in \mathcal{F}$ and $x \in \mathbb{K}$, then $V_E = \hat{V}_E$.

To understand axiom B, consider three choice scenarios. The DM's ex ante evaluation function is V, $\hat{V}$, and V_E in scenario 1,2, and 3 respectively. The primitive conditions of axiom B indicate that the DM's ex ante choices on E in scenario 3 are more similar to those in scenario 2 than to those in scenario 1. It follows that when event E is realized, the DM's ex post choices in scenario 3 are more similar to those in scenario 2 than to those in scenario 1. Since the DM's ex post choices in scenarios 1 and 3 are identical (both evaluation functions are updated to V_E), her ex post choices in scenario 2 must also be the same as those in scenario 3. That is, she revises $\hat{V}$ to V_E when E is realized.

By axiom B, the similarity between the DM's ex ante choices in different choice scenarios is preserved after information. However, this axiom can be violated if the DM's belief is rendered unambiguous when an event with small ambiguity is realized. In such case, the DM must ignore some part of the information contained in the realized event, say E, as she disregards the ex ante ambiguity on E. As a result, even if the DM makes similar ex ante choices on E in two different choice scenarios, her ex post choices in the two scenarios might become dissimilar when E is realized, since she might ignore different parts of the information contained on E in the two scenarios. Hence, axiom B can be violated in such case. To exclude situations like this, we modify axiom B to the next axiom .

Before stating the axiom, we need some definitions. For a given updating rule Γ, we say that E does not resolve ambiguity of V if V_E is ambiguous, and that E does not *strongly* resolve ambiguity of V if there exists an evaluation function $\hat{V}$ such that $(\alpha V + (1-\alpha)\hat{V})_E$ is ambiguous for all $\alpha \in (0,1)$. We note that if E does not resolve ambiguity of V, then E does not strongly resolve ambiguity of V.[15] The latter one can be considered as the limiting case of the former one.

Mixture Ambiguity Betweenness (MAB). For any $V, \hat{V}, \tilde{V} \in \mathcal{V}$, $\alpha \in (0,1]$, and $E \in (\mathcal{S}_V \cap \mathcal{S}_{\hat{V}}) \backslash \mathcal{S}_{\tilde{V}}$ that does not strongly resolve ambiguity of $\alpha V + (1-\alpha)\tilde{V}$,

[15] To see this, assume that E does not resolve ambiguity of V. It follows that $(\alpha V + (1-\alpha)V)_E$ is ambiguous for all $\alpha \in (0,1)$. Thus, E does not strongly resolve ambiguity of V.

let $W = \alpha V + (1-\alpha)\tilde{V}$ and $\hat{W} = \alpha\hat{V} + (1-\alpha)\tilde{V}$. If $\max\{W(fEx), W_E(f)\} \geq \hat{W}(fEx) \geq \min\{W(fEx), W_E(f)\}$ for all $f \in \mathcal{F}$ and $x \in \mathbb{K}$, then $V_E = \hat{V}_E$.

Axiom MAB can be implied by axioms MI and B. We interpret axiom MAB as follows. The requirement that the ambiguity of W on E is not resolved indicates that the DM processes all information contained in E when doing the update. Thus, following the interpretation of axiom B, the condition imposed on W and $\hat{W}$ indicates that the DM updates them to the same ex post evaluation function. Since E is $\tilde{V}$-null, V $(\hat{V})$ and W $(\hat{W})$ share the same ambiguity structure on E. The only difference between V $(\hat{V})$ and W $(\hat{W})$ is that V $(\hat{V})$ has $1/\alpha$ times larger degree of ambiguity on E than W $(\hat{W})$ does. Since the ambiguity on E is enlarged to the same extent for both cases, the DM updates V and $\hat{V}$ to the same ex post evaluation function when E is realized.

Independence of Irrelevant States (IIS). For any $V, \hat{V} \in \mathcal{V}$ and $E \in \mathcal{S}_V \cap \mathcal{S}_{\hat{V}}$, if $V(fEx) = \hat{V}(fEx)$ for all $f \in \mathcal{F}$ and $x \in \mathbb{K}$, then $V_E = \hat{V}_E$.

Clearly, axiom B implies axiom IIS. The primitive conditions of axiom IIS ensure that V and $\hat{V}$ completely agree on E. The axiom then says that unrealized states do not affect the DM's belief updating.

In Appendix B, we show that FB satisfies axioms ADSU, MI, and B (and thus FB satisfies all the axioms introduced so far except for axiom ISU) and can be characterized by the latter two axioms. The next axiom is the only one that is satisfied by the contraction rule but not by FB.

Nonincreasing Ambiguity by Information (NAI). For any $V \in \mathcal{V}$ that is unambiguous with respect to the partition $\{S_i\}_{i=1}^n$ of S, and any $E \in \mathcal{S}_V$, if $|E \cap S_i| \leq 1$ for each i, then V_E is unambiguous.

To interpret axiom NAI, assume that each block S_i is a payoff-relevant state. Then the realized event E can be viewed as a signal. By axiom NAI, if the DM is unambiguous with respect to the payoff-relevant states at the ex ante stage, then new information does not render her evaluation function ambiguous.

Theorem 1. *An updating rule over evaluation functions is the contraction rule if and only if it satisfies axioms ADSU, MIA, IIS, MAB, and NAI.*

Discussion of the axioms. The key motivation of our axiomatic exercise is that the ambiguity on the realized event can be resolved when the its degree is small.

Axioms ADSU, MIA, and MAB are all related to this motivation. Axiom IIS can be considered as a probabilistic version of consequentialism and is satisfied by the contraction rule as well as the two benchmark rules FB and ML. Axiom NAI is motivated by recent experimental evidence that information does not increase ambiguity. It distinguishes our model from the two benchmark ones.

An important question is whether the axioms provided in this section can be tested in the lab. Clearly, axioms ADSU and NAI can be directly tested through the DM's choices in a single choice scenario. Axioms MIA, IIS, and NAI are also testable. For instance, consider the testing of axiom IIS. Assume that there are two independent choice scenarios. In scenario 1, there is an urn (urn 1) that contains 100 balls, of which 50 are green and 50 are either red or blue with unknown composition. In scenario 2, there is an urn (urn 2) that contains 100 balls, of which 50 are either green or yellow and 50 are either blue or red with unknown composition. From each urn, a ball is drawn. Consider an act f ($\hat{f}$) in scenario 1 (2) such that f ($\hat{f}$) yields payoff x when the ball drawn from urn 1 (2) is red, payoff y when the drawn ball is blue, and z otherwise. If the DM always has the same evaluation over f and $\hat{f}$ in the two choice scenarios, then the primitive condition of axiom IIS is satisfied, i.e., the two evaluation functions in the two choice scenarios agree on the "event" that the ball drawn from the urn is either red or blue. We can then inform the DM in the two scenarios that the drawn ball is either red or blue and collect the DM's ex post evaluations over acts. By comparing the DM's ex post evaluations towards the "same" act in the two scenarios, we are able to falsify whether the DM's choice behavior satisfies axiom IIS or not.

4 Application

Throughout this section, we assume that each prior of the DM assigns probability one to a subset $\hat{S} \subseteq S$ where $\hat{S} = D \times \Theta$. D is a finite payoff-relevant state space and Θ is a finite signal space. A piece of information takes the form of some signal $\theta \in \Theta$, i.e., event $D \times \{\theta\}$. We fix $\hat{\Pi}$ to be the partition $\{\{d\} \times \Theta : d \in D\}$ of $\hat{S}$.[16]

We consider three applications of the contraction rule. First, we connect the contraction rule with the experimental findings by SO21 and show that the rule does not create belief dilation under mild conditions. Second, we study how DMs

[16]We ignore all states outside of $\hat{S}$ since they play no role in the analysis.

update their beliefs with information of unknown accuracy by associating the contraction rule with the findings by L21. Finally, we investigate whether the DM learns with a sequence of independent and ambiguous signals. We find that the contraction rule leads to learning of the true state.

4.1 Dilation

SO21 study how ambiguous information shapes the DM's beliefs over payoff-relevant states. Through lab experiments, they empirically test the hypothesis that ambiguous information increases payoff-relevant ambiguity and reject it. In what follows, we show that the contraction rule provides consistent predictions with their findings.

Consider a DM with prior set P over $D \times \Theta$. A signal $\theta \in \Theta$ dilates the DM's payoff-relevant belief set under updating rule Q if $P_{\hat{\Pi}} \subsetneq \left(Q(P, D \times \{\theta\})\right)_{\hat{\Pi}}$, i.e., after observing signal θ, the DM's posterior set over D strictly contains her prior one.[17] The following proposition establishes a non-dilation result for the contraction rule when D contains only two states.

Proposition 5. *If $|D| = 2$, then for any signal θ such that $D \times \{\theta\}$ is P-non-null, θ does not dilate the DM's payoff-relevant belief set under the contraction rule.*

We discuss the special case of Proposition 5 where there are two symmetric signals. This is exactly the case studied by SO21. Let $D = \{d_1, d_2\}$ and $\Theta = \{\theta_1, \theta_2\}$. Assume that $p \in P$ if and only if $p(d_1, \theta_1) = \alpha\beta$, $p(d_1, \theta_2) = \alpha(1 - \beta)$, $p(d_2, \theta_1) = (1 - \alpha)(1 - \beta)$, and $p(d_2, \theta_2) = (1 - \alpha)\beta$ for some $\alpha \in [1/2 - a, 1/2 + a]$ and $\beta = [1/2 - b, 1/2 + b]$, where $a \in [0, 1/2]$ and $b \in (0, 1/2]$ are constants. That is, the DM believes the probability of d_1 to be at least $1/2 - a$ and at most $1/2 + a$ and the conditional probability of θ_1 (respectively θ_2) on d_1 (respectively d_2) to be at least $1/2 - b$ and at most $1/2 + b$.

When $a = 0$, there is no ex ante payoff-relevant ambiguity: the DM assigns probability half to both d_1 and d_2. As shown by SO21, the realization of any signal dilates the DM's belief set over D if she follows FB or ML. By contrast, if the DM updates with the contraction rule, her ex post belief over D would be the same as her ex ante one. Thus, for a given prospect (call it prospect K) that yields a

[17] Our definition of belief dilation is a weak version. Wasserman and Kadane (1990) define dilation as the case in which each signal enlarges the payoff-relevant set of priors.

high payoff on d_1 and a low payoff on d_2, only the contraction rule predicts that information does not change the DM's evaluation over prospect K. This prediction is tested to be true for a large proportion of ambiguity averse subjects by SO21.

When $a > 0$, there is ex ante payoff-relevant ambiguity: the DM believes that the probability of d_1 is at least $1/2 - a$ and at most $1/2 + a$. For any given signal, there are two cases to be considered for the contraction rule. If $(1/2+a)(1/2+b) \leq 1/2$, then the signal resolves the DM's payoff-relevant ambiguity. In this case, the DM believes d_1 and d_2 to be equally likely after observing the signal. If $(1/2+a)(1/2+b) > 1/2$, then the DM's beliefs are revised to that the probability of d_1 ranges from $1-(1/2+a)(1/2+b)$ to $(1/2+a)(1/2+b)$. In both cases, the information decreases payoff-relevant ambiguity. Thus, an ambiguity averse DM increases her evaluation over prospect K after the information. This is indeed the case for a non-negligible proportion of ambiguity averse subjects in SO21's experiments.

Our next proposition provides generic conditions under which belief dilation does not occur with the contraction rule: when there is no ex ante payoff-relevant ambiguity or when the prior set is contained in the interior of $\Delta(D \times \Theta)$.

Proposition 6. *If either $P_{\hat{\Pi}}$ is a singleton or $p(d,\theta) > 0$ for all $p \in P$, $d \in D$, and $\theta \in \Theta$, then no signal dilates the DM's belief set over D under the contraction rule.*

4.2 Information with Ambiguous Accuracy

In this section, we study how DMs react to information with unknown accuracy following the framework of L21. Let $D = \{d_1, d_2\}$ and $\Theta = \{\hat{d}_1, \hat{d}_2\}$. The DM's priors over D are characterized by an interval $[\underline{r}, \overline{r}] \subseteq (0,1)$, i.e., she believes that the probability of d_1 ranges from $\underline{r}$ to $\overline{r}$. Following L21, we consider two scenarios.

In scenario 1, the DM can seek information from an expert with unknown accuracy. The expert informs the DM of her prediction of the true state by sending either signal $\hat{d}_1$ or $\hat{d}_2$: $\hat{d}_1$ (respectively $\hat{d}_2$) refers to prediction d_1 (respectively d_2). The DM believes that there are two possible accuracy levels of the expert's predictions: H and L. That is, the DM considers two conditional probabilities of the signals: $c^H(\hat{d}_1|d_1) = c^H(\hat{d}_2|d_2) = H$ and $c^L(\hat{d}_1|d_1) = c^L(\hat{d}_2|d_2) = L$. We require that $1 > H > L > 0$ and $H + L > 1$.[18] We allow L to be smaller than

[18]If $H + L < 1$, we can exchange the labels of the two signals.

1/2 in order to accommodate the possibility that the expert intentionally lies to the DM. We assume that the accuracy of the expert's predictions is independent of the DM's priors over D. Thus, the DM's prior set P over $D \times \Theta$ is given by $co(\{p_{\underline{r},L}, p_{\underline{r},H}, p_{\overline{r},L}, p_{\overline{r},H}\})$, where for each $r \in \{\underline{r}, \overline{r}\}$ and $J \in \{L, H\}$, we have

$$p_{r,J}(d_1, \hat{d}_1) = rJ, p_{r,J}(d_1, \hat{d}_2) = r(1-J),$$
$$p_{r,J}(d_2, \hat{d}_1) = (1-r)(1-J), p_{r,J}(d_2, \hat{d}_2) = (1-r)J.$$

In scenario 2, there is an expert with accuracy $\frac{H+L}{2}$. Now, conditional on the true state being $d_i \in D$, the probability for the expert predicting correctly is $\frac{H+L}{2}$. The prior set of the DM is thus given by $co(\{p_{\overline{r}}, p_{\underline{r}}\})$ where for each $r \in \{\overline{r}, \underline{r}\}$,

$$p_r(d_1, \hat{d}_1) = r\frac{H+L}{2}, p_r(d_1, \hat{d}_2) = r(1-\frac{H+L}{2}),$$
$$p_r(d_2, \hat{d}_1) = (1-r)(1-\frac{H+L}{2}), p_r(d_2, \hat{d}_2) = (1-r)\frac{H+L}{2}.$$

We compare the DM's ex post evaluations over certain prospects in the two scenarios. For this purpose, we consider two acts f and g where $f(d_1, \hat{d}_1) = f(d_1, \hat{d}_2) = g(d_2, \hat{d}_1) = g(d_2, \hat{d}_2) = 1$ and $f(d_2, \hat{d}_1) = f(d_2, \hat{d}_2) = g(d_1, \hat{d}_1) = g(d_1, \hat{d}_2) = 0$. Thus f yields payoff 1 on d_1 and 0 on d_2; g yields 0 on d_1 and 1 on d_2. Let v_f^a (respectively v_f^u) be the DM's ex post evaluation of f after observing $\hat{d}_1$ in scenario 1 (respectively scenario 2); let v_g^a (respectively v_g^u) be the DM's ex post evaluation of g after observing $\hat{d}_1$ in scenario 1 (respectively scenario 2).

Proposition 7. *If the DM updates her beliefs with the contraction rule, then $v_f^u > v_f^a$, and the comparison between v_g^u and v_g^a depends on $\frac{1-\overline{r}}{1-\underline{r}}$:*

(1) *if $\frac{1-\overline{r}}{1-\underline{r}} < \frac{1-\underline{r}}{1-\underline{r}H}\frac{2-H-L}{H+L}$, then $v_g^a < v_g^u$,*

(2) *if $\frac{1-\overline{r}}{1-\underline{r}} = \frac{1-\underline{r}}{1-\underline{r}H}\frac{2-H-L}{H+L}$, then $v_g^a = v_g^u$,*

(3) *if $\frac{1-\underline{r}}{1-\underline{r}H}\frac{2-H-L}{H+L} < \frac{1-\overline{r}}{1-\underline{r}}$, then $v_g^u < v_g^a$.*

We interpret Proposition 7 as follows. The condition $H + L > 1$ ensures that the information is asymmetrically informative in both scenarios. Thus $\hat{d}_1$ is good news for f and bad news for g. The first inequality $v_f^u > v_f^a$ indicates that with the contraction rule, the DM under-reacts to good news if it is ambiguous.

However, contraction rule does not always predict under-reaction to ambiguously bad news. To understand our result, note that bad news not only pushes the DM's priors towards the bad state (d_1) of g but also partially resolves payoff-relevant ambiguity. Thus a DM who exhibits ambiguity aversion may increase her evaluation of g even receiving bad news. Note that in the proposition, the term $\frac{1-\bar{r}}{1-\underline{r}}$ captures the degree of the DM's ex ante ambiguity on D. A larger value of $\frac{1-\bar{r}}{1-\underline{r}}$ corresponds to less ex ante ambiguity. When $\frac{1-\bar{r}}{1-\underline{r}}$ is small (case (1)), the DM's ex ante ambiguity is large and she benefits more from resolving of the ambiguity. Thus, the DM benefits more from unambiguous information than the ambiguous one. As the ex ante ambiguity decreases, the effect of ambiguity resolving is dominated by the effect of under-reaction to ambiguous information, which leads to $v_g^u < v_g^a$ (case (3)).

The theoretical predictions by the contraction rule are consistent with the experimental and empirical evidence provided by L21, who finds that subjects exhibit under-reaction to ambiguous information and pessimism to ambiguously bad news through both lab experiments and stock price reactions to analyst earnings forecasts. However, as shown by Proposition 7, our DM exhibits pessimism for ambiguously bad news only when the ex ante ambiguity is large while a certain proportion of subjects in L21's experiments exhibit pessimism for such news when there is no ex ante ambiguity. Nevertheless, our results are relevant in the empirical analysis of the stock price reactions by L21, where the ex ante ambiguity of stock prices is typically large.

4.3 Learning with Ambiguous Information

In this section, we extend our analysis in Section 4.2 to a Wald-type learning scenario. Consider a binary payoff-relevant state space $D = \{d_1, d_2\}$, for which the DM's prior set is given by $[\underline{r}, \bar{r}] \subseteq (0, 1)$. Instead of observing one signal, the DM observes n independent signals. The signal space is thus given by $\Theta = \hat{\Theta}^n$, where $\hat{\Theta} = \{\hat{d}_1, \hat{d}_2\}$. To motivate, one can imagine that there are n independent experts. Each expert has prediction accuracy H or L, where $1 > H > L > 0$ and $H + L > 1$.

When a sequence of signals $(\hat{d}^1, ..., \hat{d}^n) \in \hat{\Theta}^n$ are realized, the DM updates her prior set over D to some posterior set. We are interested in the DM's ex post belief set when n is large. In particular, we investigate whether the DM's beliefs converge to the true state of the world if provided with a long sequence of independent and ambiguous signals. For this purpose, we assume that the true prediction accuracy of each expert is $\frac{H+L}{2} > 1$. Let $I_{d_i}^n \subseteq [0, 1]$ be the set of the DM's posterior beliefs over d_1 (which is a random variable) when the true state is $d_i \in D$ and the number of signals is n. The following proposition asserts that with the contraction rule, the DM would finally learn the true state when there are enough signals.

Proposition 8. *If the DM updates with the contraction rule, then for any* $\epsilon > 0$*, the probabilities of* $\min_{r \in I^n_{d_1}} \geq 1 - \epsilon$ *and* $\max_{r \in I^n_{d_2}} \leq \epsilon$ *both converge to one.*

We close this section by comparing Proposition 8 with the predictions of FB and ML on learning. If the DM updates with FB, she cannot learn the true state if $L < 1/2$. In this case, any given signal can be interpreted as supporting state d_1 as well as against d_1. Consequently, the DM dilates her belief set over D no matter what signal is observed at each round, and her posterior set converges to $[0, 1]$. That is, after many rounds of learning, the DM becomes completely confused about the true state and believes that the probability of d_1 ranges from 0 to 1.

If the DM follows ML, then whether the DM learns the true state depends on *how she processes the information.* If the DM updates her beliefs after receiving all signals, then she can finally learn the true state.[19] However, if the DM updates her prior set over D signal by signal, then she cannot learn the true state with probability one. To see why, note that ML is not a divisible rule. If the DM observes a lot of signals $\hat{d}_1$ at the first several rounds, she would update her beliefs towards state d_1. If she assigns a high enough probability to d_1, then no matter what new signal is realized, she will always interpret the new signal as supporting d_1. In this case, whether the DM can learn the true state or not depends on the initial sequence of signals. By contrast, since the contraction rule is divisible, it robustly predicts that the DM would learn the true state with probability close to one regardless of how she processes the information.

5 Conclusion

In this paper, we axiomatize a new rule, the contraction rule, for updating ambiguous beliefs. The rule moderates the two benchmark models of belief updating, FB and ML, and departs from them by requiring that information does not render an unambiguous belief over payoff-relevant states ambiguous.

[19] For instance, suppose that the true state is d_1. If the total number of signals is large enough, then with probability closed to one, there is a larger proportion of signal $\hat{d}_1$ than that of signal $\hat{d}_2$. The maximum likelihood is achieved by associating signal $\hat{d}_1$ with accuracy H and associating signal $\hat{d}_2$ with accuracy L. By this, the DM believes the probability of d_1 to be close to 1.

Our new rule shuts down the channel of belief dilation under mild conditions and provides new predictions for applications involving with belief revising under ambiguity. We leave the work of applying our theory to interactive decision-making for future research.[20]

6 Appendix

6.1 Appendix A: Omitted Proofs and Examples

Proof of Proposition 1. Consider $P \in \mathscr{P}$ and P-non-null event E. If $\mu_P(E) \leq 1$, then $Q^c(P, E) = \{\overline{\mu_P|E}\}$. Clearly, $|Q^c(P, E)| = 1$. If $\mu_P(E) > 1$, then $Q^c(P, E) = \{\Phi(p|E, \mu_P|E)\}_{p \in P}$. Fix some $s \in E$ and pick $\hat{p} \in P$ such that $\hat{p}(s) = \mu_P(s)$. Since $\mu_P(E) > 1$, there exists $s^* \in E \backslash s$ such that $\hat{p}(s^*) < \mu_P(s^*)$. For such s^*, pick $p^* \in P$ such that $p^*(s^*) = \mu_P(s^*)$. Since $\Phi(p^*|E, \mu_P|E)(s^*) = \mu_P(s^*) > \Phi(\hat{p}|E, \mu_P|E)(s^*)$, we know $\Phi(p^*|E, \mu_P|E) \neq \Phi(\hat{p}|E, \mu_P|E)$. Thus, $|Q^c(P, E)| \neq 1$.

For the second statement of the proposition, assume $\mu_P(E) > 1$. Since $p(s) \leq \mu_P(s)$ for all $s \in E$ and $p \in P$, we have $\Phi(p|E, \mu_P|E)(s) \leq \mu_P(s)$ for all $s \in E$ and $p \in P$. For each $s \in E$, there exists $\tilde{p} \in P$ such that $\tilde{p}(s) = \mu_P(s)$. Clearly, $\Phi(\tilde{p}|E, \mu_P|E)(s) = \mu_P(s)$. Hence, $\max_{q \in Q^c(P,E)} q(s) = \max_{p \in P} p(s)$ for all $s \in E$. That is, $\mu_P|E = \mu_{Q^c(P,E)}$. □

Proof of Proposition 2. Consider any $P \in \mathscr{P}$ and any P-non-null event E. If $\mu_P(E) \leq 1$, then $Q^c(P, E) = \{\overline{\mu_P|E}\}$, which is non-empty, convex, and closed. If $\mu_P(E) > 1$, then $Q^c(P, E) = \{\Phi(p|E, \mu_P|E)\}_{p \in P}$, which is non-empty. To see that $Q^c(P, E)$ is convex, first note that $G = \{\alpha p|E + (1 - \alpha)\mu_P|E : \alpha \in [0, 1], p \in P\}$ is a convex set of measures. Since $Q^c(P, E)$ is the intersection of G with the set of probability measures over E, $Q^c(P, E)$ is convex. To proceed, we show that $Q^c(P, E)$ is closed. Consider a sequence of probability distributions $\{q_n\}_{n=1}^{+\infty} \subseteq Q^c(P, E)$ that converges to some q. For each q_n, there exists some $p_n \in P$ such that $q_n = \Phi(p_n|E, \mu_P|E)$. Since P is closed (and thus compact since it is bounded), it is without loss of generality to assume that $\{p_n\}_{n=1}^{+\infty}$ converges to some $p \in P$. It

[20] In an earlier version of the current paper, Tang (2020) applies the contraction rule to study information design problems à la Kamenica and Gentzkow (2011) and shows that if the agent updates with the contraction rule, then the principal can exact almost all the revenue through suitable ambiguous information structures.

follows that

$$q_n = \Phi(p_n|E, \mu_P|E) \text{ converges to } \Phi(p|E, \mu_P|E).$$

That is, $q = \Phi(p|E, \mu_P|E) \in Q^c(P, E)$ and thus $Q^c(P, E)$ is closed.

For the second statement of the proposition, assume that $p(E) = 1$ for all $p \in P$. We want to show that $Q^c(P, E) = P$. If $\mu_P(E) \leq 1$, then P is a singleton, since otherwise we can find different $p, \hat{p} \in P$ such that $1 < \mu_{\{p,\hat{p}\}}(E) \leq \mu_P(E)$, which is a contradiction. Hence, $P = \{p^*\}$ for some p^*. It follows that $\mu_P = p^*$ and $Q^c(P, E) = \{\overline{\mu_P|E}\} = \{\overline{p^*|E}\} = \{p^*\} = P$. If $\mu_P(E) > 1$, then by formula (2), we know $\Phi(p|E, \mu_P|E) = p|E = p$ for each $p \in P$. Again, we have $Q^c(P, E) = P$. □

Example 5. For each $n \geq 1$, let $P_n = co(\{p, \hat{p}, \tilde{p}_n\})$ with support $\{s_1, s_2, s_3\} \subseteq S$. Distributions of p, $\hat{p}$, and $\tilde{p}_n$ are: $p(s_1) = 1/2$, $p(s_2) = 1/2$, $p(s_3) = 0$, $\hat{p}(s_1) = 0$, $\hat{p}(s_2) = 0$, $\hat{p}(s_3) = 1$, $\tilde{p}_n(s_1) = 1/(2n)$, $\tilde{p}_n(s_2) = 0$, and $\tilde{p}_n(s_3) = 1-1/(2n)$. Let $E = \{s_1, s_2\}$ and $\tilde{p} = \tilde{p}_1$. It follows that $Q^{fb}(P_n, E) = co(\{\overline{p|E}, \overline{\tilde{p}_n|E}\}) = co(\{\overline{p|E}, \overline{\tilde{p}|E}\})$. Note that $\{P_n\}_{n=1}^{+\infty}$ converges to $P = \{p, \hat{p}\}$. However, $\{Q^{fb}(P_n, E)\}_{n=1}^{+\infty}$ does not converge to $Q^{fb}(P, E)$ since $Q^{fb}(P, E) = \{\overline{p|E}\}$. □

Example 6. Let $P = co(\{p, \hat{p}\})$ with support $\{s_1, s_2, s_3\} \subseteq S$. Distributions of p and $\hat{p}$ are: $p(s_1) = 1/5, p(s_2) = 2/5, p(s_3) = 2/5$, $\hat{p}(s_1) = 2/5$, $\hat{p}(s_2) = 1/5$, and $\hat{p}(s_3) = 2/5$. Let $E = \{s_1, s_2\}$. It follows that $Q^{ml}(P, E) = co(\{\overline{p|E}, \overline{\hat{p}|E}\})$ since $p(\{s_1, s_2\}) = \hat{p}(\{s_1, s_2\}) = 3/5$. Next, for each $n \geq 1$, consider $P_n = co(\{p_n, \hat{p}\})$ where $p_n(s_1) = 1/5 - 1/(n+5)$, $p_n(s_2) = 2/5$, and $p_n(s_3) = 2/5 + 1/(n+5)$. It follows that $Q^{ml}(P_n, E) = \{\overline{\hat{p}|E}\}$ for each n. Note that $\{P_n\}_{n=1}^{+\infty}$ converges to P, but $\{Q^{ml}(P_n, E)\}_{n=1}^{+\infty}$ does not converge to $Q^{ml}(P, E)$. □

Proof of Proposition 3. Since $\{P_n\}_{n=1}^{+\infty}$ converges to P, we know that for any event E, $\{\max_{p \in P_n} p(E)\}_{n=1}^{+\infty}$ converges to $\max_{p \in P} p(E)$. Since E is P-non-null, there exists m such that for all $n \geq m$, $\max_{p \in P_n} p(E) > 0$, i.e., for all $n \geq m$, E is P_n-non-null.

To proceed, we show that $\{Q^c(P_n, E)\}_{n=m}^{+\infty}$ converges to $Q^c(P, E)$. Without loss of generality, assume $m = 1$. We consider three cases: (a) $\mu_P(E) < 1$, (b) $\mu_P(E) > 1$, and (c) $\mu_P(E) = 1$. In case (a), there exists m^* such that for all $n \geq m^*$, $\mu_{P_n}(E) < 1$. Hence, for each $n \geq m^*$, $Q^c(P_n, E) = \{\overline{\mu_{P_n}|E}\}$. Note that for each $s \in E$, $\{\mu_{P_n}(s)\}_{n=1}^{+\infty}$ converges to $\mu_P(s)$. It follows that $\{\overline{\mu_{P_n}|E}\}_{n=m^*}^{+\infty}$ converges to $\overline{\mu_P|E}$. That is, $\{Q^c(P_n, E)\}_{n=1}^{+\infty}$ converges to $Q^c(P, E)$.

In case (b), suppose to the contrary that $\{Q^c(P_n, E)\}_{n=1}^{+\infty}$ does not converge to $Q^c(P, E)$. Then there exists $\epsilon > 0$ and a subsequence $\{Q^c(P_{n_t}, E)\}_{t=1}^{+\infty}$ such that $d^h(Q^c(P_{n_t}, E), Q^c(P, E)) \geq \epsilon$ for all t. Without loss of generality, assume that $\{Q^c(P_{n_t}, E)\}_{t=1}^{+\infty} = \{Q^c(P_n, E)\}_{n=1}^{+\infty}$. We only need to discuss two cases: (ib) there exists a sequence $\{q_n\}_{n=1}^{+\infty}$ such that $q_n \in Q^c(P_n, E)$ and $d^h(\{q_n\}, Q^c(P, E)) \geq \epsilon$ for each n, or (iib) there exists a sequence $\{q_n\}_{n=1}^{+\infty} \subseteq Q^c(P, E)$ such that $d^h(\{q_n\}, Q^c(P_n, E)) \geq \epsilon$ for each n.

In case (ib), for each q_n, there exists $p_n \in P_n$ such that $q_n = \Phi(p_n|E, \mu_{P_n}|E)$. Consider a convergent subsequence $\{p_{n_t}\}_{t=1}^{+\infty}$ of $\{p_n\}_{n=1}^{+\infty}$ that converges to some p. Since $\{P_n\}_{n=1}^{+\infty}$ converges to P, we know that $p \in P$. By the definition of the contraction rule, we know

$$q_{n_t} = \frac{\mu_{P_{n_t}}(E) - 1}{\mu_{P_{n_t}}(E) - p_{n_t}(E)} p_{n_t}|E + \frac{1 - p_{n_t}(E)}{\mu_{P_{n_t}}(E) - p_{n_t}(E)} \mu_{P_{n_t}}|E.$$

It follows that $\{q_{n_t}\}_{t=1}^{+\infty}$ converges to

$$\frac{\mu_P(E) - 1}{\mu_P(E) - p(E)} p|E + \frac{1 - p(E)}{\mu_P(E) - p(E)} \mu_P|E = \Phi(p|E, \mu_P|E),$$

which is contained in $Q^c(P, E)$. This is a contradiction since $d^h(\{q_n\}, Q^c(P, E)) \geq \epsilon$ for each n. Therefore, case (ib) is impossible.

In case (iib), there exists a convergent subsequence $\{q_{n_t}\}_{t=1}^{+\infty}$ that converges to some $q \in Q^c(P, E)$. There exists t^* such that for all $t \geq t^*$, $d^h(\{q\}, Q^c(P_{n_t}, E)) \geq \frac{\epsilon}{2}$. Note that there exists some $p \in P$ such that $q = \Phi(p|E, \mu_P|E)$. Since $\{P_{n_t}\}_{t=1}^{+\infty}$ converges to P, there exists a sequence $\{p_{n_t}\}_{t=1}^{+\infty}$ converging to p where $p_{n_t} \in P_{n_t}$ for each t. By a similar argument, we know that $\{\Phi(p_{n_t}|E, \mu_{P_{n_t}}|E)\}_{t=1}^{+\infty}$ converges to $\Phi(p|E, \mu_P|E) = q$. This is a contradiction since $d^h(\{q\}, Q^c(P_{n_t}, E)) \geq \frac{\epsilon}{2}$ for all $t \geq t^*$. Therefore, case (iib) is also impossible. Hence, $\{Q^c(P_n, E)\}_{n=1}^{+\infty}$ converges to $Q^c(P, E)$ in case (b).

In case (c), we know that $Q^c(P, E) = \{\overline{\mu_P|E}\} = \{\mu_P|E\}$. Divide $\{P_n\}_{n=1}^{+\infty}$ to two subsequences $\{P_{n_t}\}_{t=1}^{\infty}$ and $\{P_{m_t}\}_{t=1}^{\infty}$ such that $\mu_{P_{n_t}}(E) > 1$ and $\mu_{P_{m_t}}(E) \leq 1$ for all t (in some cases, one of the two sequences is finite, then we only consider the other sequence). For $\{P_{m_t}\}_{t=1}^{\infty}$, it is easy to show that $\{\overline{\mu_{P_{m_t}}|E}\}_{t=1}^{+\infty}$ converges to $\mu_P|E$. For $\{P_{n_t}\}_{t=1}^{\infty}$, suppose to the contrary that $\{Q^c(P_{n_t}, E)\}_{t=1}^{+\infty}$ does not converge to $\{\mu_P|E\}$. By a similar argument, (without loss of generality) we can find $p_{n_t} \in P_{n_t}$ for each t such that the sequence $\{p_{n_t}\}_{t=1}^{+\infty}$ converges to some $p \in P$, but $\{\Phi(p_{n_t}|E, \mu_{P_{n_t}}|E)\}_{t=1}^{+\infty}$ does not converge to $\mu_P|E$. If $p(E) < 1$, then

$\{\Phi(p_{n_t}|E, \mu_{P_{n_t}}|E)\}_{t=1}^{+\infty}$ converges to $\mu_P|E$, which is a contradiction. If $p(E) = 1$, then for each $\hat{p} \in P$, either $\hat{p} = p$ or $\hat{p}(s) \leq p(s)$ for all $s \in E$, since otherwise $\mu_P(E) > 1$. This implies that $\mu_P|E = p$. It follows that both $\{p_{n_t}|E\}_{t=1}^{+\infty}$ and $\{\mu_{P_{n_t}}|E\}_{t=1}^{+\infty}$ converge to $\mu_P|E$. Thus $\{\Phi(p_{n_t}|E, \mu_{P_{n_t}}|E)\}_{t=1}^{+\infty}$ converges to $\mu_P|E$, which is again a contradiction. Hence, $\{Q^c(P_{n_t}, E)\}_{t=1}^{+\infty}$ converges to $\{\mu_P|E\}$. $\square$

Proof of Proposition 4. We consider three cases. In case (i), $\mu_P(E) \leq 1$. We have $Q^c(P, E) = \{\overline{\mu_P|E}\}$. Then $Q^c(Q^c(P, E), F) = Q^c(\{\overline{\mu_P|E}\}, F) = \{\overline{\overline{\mu_P|E}|F}\} = \{\overline{\mu_P|F}\} = Q^c(P, F)$. In case (ii), $\mu_P(E) > 1$ and $\mu_P(F) \leq 1$. By Proposition 1, we have $\mu_P|E = \mu_{Q^c(P,E)}$. Clearly, it implies $\mu_P|F = \mu_{Q^c(P,E)}|F$. By a similar argument as case (i), we know that $Q^c(Q^c(P, E), F) = Q^c(P, F)$. In case (iii), $\mu_P(E) > 1$ and $\mu_P(F) > 1$. We need the following lemma, of which the proof is simple algebra and thus omitted.

Lemma 1. *For any $\pi, \pi' \in \mathcal{M}(S)$ and any $\alpha \in (0, 1]$ such that $\pi(S) \leq 1$, $\pi'(S) > 1$, and $(\alpha\pi + (1-\alpha)\pi')(S) \leq 1$, we have $\Phi(\pi, \pi') = \Phi(\alpha\pi + (1-\alpha)\pi', \pi')$.*

Back to the proof for case (iii), since $\mu_P|F = \mu_{Q^c(P,E)}|F$, we have $Q^c(Q^c(P, E), F) = \{\Phi(q|F, \mu_P|F)\}_{q \in Q^c(P,E)} = \{\Phi(\Phi(p|E, \mu_P|E)|F, \mu_P|F)\}_{p \in P}$. Note that for each $p \in P$, $\Phi(p|E, \mu_P|E) = \alpha p|E + (1-\alpha)\mu_P|E$ for some $\alpha \in (0, 1]$. Hence, $\Phi(p|E, \mu_P|E)|F = \alpha p|F + (1-\alpha)\mu_P|F$. By Lemma 1, we have

$$\{\Phi(\Phi(p|E, \mu_P|E)|F, \mu_P|F)\}_{p \in P} = \{\Phi(p|F, \mu_P|F)\}_{p \in P} = Q^c(P, F).$$

$\square$

Proof of Theorem 1. We start with several lemmas. Then we prove the necessity and sufficiency parts of the theorem. Throughout the proof, we fix the partition $\Pi = \{\{s\} : s \in E\} \cup \{S \backslash E\}$.

Lemma 2. *For any $P \in \mathscr{P}$ and event E, if $\mu_P(E) > 1$, then*

$$\Big(co\,(P \cup Q^c(P, E))\Big)_\Pi = \bigcup_{p \in P} \Big(co(\{p, \Phi(p|E, \mu_P|E)\})\Big)_\Pi. \tag{3}$$

Proof. Let $B = \bigcup_{p \in P} \Big(co(\{p, \Phi(p|E, \mu_P|E)\})\Big)_\Pi$. Clearly, $\Big(P \cup Q^c(P, E)\Big)_\Pi \subseteq B \subseteq \Big(co\,(P \cup Q^c(P, E))\Big)_\Pi$. It suffices to show that B is convex. Define the set of measures $G = \{\alpha p|E + (1-\alpha)\mu_P|E : \alpha \in [0, 1], p \in P\}$. G is a convex set since $P|E$ is convex. Consider the subset $G' \subseteq G$ such that $G' = \{\pi \in G : \pi(E) \leq 1\}$.

It follows that G' is convex. Since $B|E = G'$ and $\Pi = \{\{s\} : s \in E\} \cup \{S \backslash E\}$, B is also convex. We are done. $\square$

Lemma 3. *For any $P, \hat{P}, \tilde{P} \in \mathscr{P}$, $E \in \left(\mathcal{S}_P \cap \mathcal{S}_{\hat{P}}\right) \backslash \mathcal{S}_{\tilde{P}}$, and $\alpha \in (0,1]$ with $\alpha\mu_P(E) > 1$, let $P^1 = \alpha P + (1-\alpha)\tilde{P}$ and $P^2 = \alpha\hat{P} + (1-\alpha)\tilde{P}$. If* (i) *$P^2_\Pi \subseteq \left(co(P^1 \cup Q^c(P^1, E))\right)_\Pi$ and* (ii) *for each $p \in P^1$, there exists $\hat{p} \in P^2$ such that $\hat{p}_\Pi \in \left(co(\{p, \Phi(p|E, \mu_{P^1}|E)\})\right)_\Pi$, then $Q^c(P, E) = Q^c(\hat{P}, E)$.*

Proof. Condition (i) and Lemma 2 imply that for each $\hat{p} \in P^2$, there exists $p \in P^1$ such that $\hat{p}_\Pi \in \left(co(\{p, \Phi(p|E, \mu_{P^1}|E)\})\right)_\Pi$. Thus, for each $\hat{p} \in P^2$, there exists $p \in P^1$ such that $\hat{p}|E = \beta p|E + (1-\beta)\Phi(p|E, \mu_{P^1}|E) = \eta p|E + (1-\eta)\mu_{P^1}|E$ for some $\beta \in [0,1]$ and $\eta \in (0,1]$. Since E is $\tilde{P}$-null, we know $\alpha P|E = P^1|E$, $\alpha\hat{P}|E = P^2|E$, $\mu_{P^1}|E = \alpha\mu_P|E$, and $\mu_{P^2}|E = \alpha\mu_{\hat{P}}|E$. Thus, for each $\hat{p}^* \in \hat{P}$, there exists $p^* \in P$ and $\eta \in (0,1]$ such that $\alpha\hat{p}^*|E = \eta\alpha p^*|E + (1-\eta)\alpha\mu_P|E$, i.e.,

$$\hat{p}^*|E = \eta p^*|E + (1-\eta)\mu_P|E. \tag{4}$$

By a similar argument, condition (ii) implies that for each $p^* \in P$, there exists $\hat{p}^* \in \hat{P}$ such that condition (4) holds for some $\eta \in (0,1]$. Clearly, it implies $\mu_P|E = \mu_{\hat{P}}|E$. Note that if $\hat{p}^*$ and p^* satisfy condition (4) for some $\eta \in (0,1]$, then $\Phi(\hat{p}^*|E, \mu_{\hat{P}}|E) = \Phi(\hat{p}^*|E, \mu_P|E) = \Phi(p^*|E, \mu_P|E)$, where the second equality follows from Lemma 1. We conclude that for each $\hat{p}^* \in \hat{P}$, there exists $p^* \in P$ such that $\Phi(\hat{p}^*|E, \mu_{\hat{P}}|E) = \Phi(p^*|E, \mu_P|E)$, and vice versa. Thus, $Q^c(P, E) = Q^c(\hat{P}, E)$. $\square$

The next lemma can be proved similarly as Lemma 3, and thus we omit its proof.

Lemma 4. *For any $P, \hat{P}, \tilde{P} \in \mathscr{P}$, $E \in \left(\mathcal{S}_P \cap \mathcal{S}_{\hat{P}}\right) \backslash \mathcal{S}_{\tilde{P}}$, and $\alpha \in (0,1]$ with $\alpha\mu_P(E) = 1$, let $P^1 = \alpha P + (1-\alpha)\tilde{P}$ and $P^2 = \alpha\hat{P} + (1-\alpha)\tilde{P}$. If* (i) *$P^2_\Pi \subseteq \left(co(P^1 \cup \{\mu_{P^1}|E\})\right)_\Pi$ and* (ii) *for each $p \in P^1$, there exists $\hat{p} \in P^2$ such that $\hat{p}_\Pi \in \left(co(\{p, \mu_{P^1}|E\})\right)_\Pi$, then $Q^c(P, E) = Q^c(\hat{P}, E)$.*

Lemma 5. *For any $V, \hat{V} \in \mathcal{V}$ and $\alpha \in [0,1]$, if P represents V and $\hat{P}$ represents $\hat{V}$, then $\alpha V + (1-\alpha)\hat{V}$ is an evaluation function that is represented by $\alpha P + (1-\alpha)\hat{P}$.*

Proof. Consider an arbitrary $f \in \mathcal{F}$. We have $\mathbb{E}_{\alpha P+(1-\alpha)\hat{P}}(f) = \min_{p \in P, \hat{p} \in \hat{P}} \mathbb{E}_{\alpha p+(1-\alpha)\hat{p}}(f) = \min_{p \in P} \min_{\hat{p} \in \hat{P}} \mathbb{E}_{\alpha p+(1-\alpha)\hat{p}}(f) = \min_{p \in P} \min_{\hat{p} \in \hat{P}} (\alpha\mathbb{E}_p(f) + (1-\alpha)\mathbb{E}_{\hat{p}}(f)) = \alpha \min_{p \in P} \mathbb{E}_p(f) + (1-$

$\alpha)\min_{\hat{p}\in\hat{P}}\mathbb{E}_{\hat{p}}(f)=\alpha V(f)+(1-\alpha)\hat{V}(f)=(\alpha V+(1-\alpha)\hat{V})(f)$. Thus, $\alpha V+(1-\alpha)\hat{V}$ is an evaluation function and is represented by $\alpha P+(1-\alpha)\hat{P}$. □

The next two lemmas are trivial and thus their proofs are omitted. We write them here only for the purpose of reference.

Lemma 6. *For any $V \in \mathcal{V}$ and partition $\Pi^* = \{S_i\}_{i=1}^n$ of S, V is unambiguous with respect to Π^* if and only if P_{Π^*} is a singleton, where $P \in \mathscr{P}$ represents V.*

Lemma 7. *For any $V \in \mathcal{V}$ and event E, suppose that V is represented by some $P \in \mathscr{P}$. Then E is P-null if and only if E is V-null*

Lemma 8. *With the contraction rule, for any $V \in \mathcal{V}$ and V-non-null event E, suppose that V is represented by some $P \in \mathscr{P}$. Then E does not strongly resolve ambiguity of V if and only if $\mu_P(E) \geq 1$.*

Proof. If E does not strongly resolve ambiguity of V, then there exists evaluation function $\hat{V}$ such that $(\alpha V + (1-\alpha)\hat{V})_E$ is ambiguous for all $\alpha \in (0,1)$. Let $\hat{V}$ be represented by $\hat{P}$. By Lemma 5, $\alpha V + (1-\alpha)\hat{V}$ is represented by $\alpha P + (1-\alpha)\hat{P}$. It follows from Proposition 1 and Lemma 6 that for all $\alpha \in (0,1)$, $\mu_{\alpha P+(1-\alpha)\hat{P}}(E) > 1$. Thus, $\mu_P(E) \geq 1$. Inversely, if $\mu_P(E) \geq 1$, then consider some evaluation function $\hat{V}$ such that $\hat{P}$ represents $\hat{V}$ and $\mu_{\hat{P}}(E) > 1$. We have for any $\alpha \in (0,1)$, $\mu_{\alpha P+(1-\alpha)\hat{P}}(E) > 1$. By Proposition 1 and Lemma 6, $(\alpha V+(1-\alpha)\hat{V})_E$ is ambiguous for all $\alpha \in (0,1)$. Hence, E does not strongly resolve ambiguity of V. □

Lemma 9. *For any $V, \hat{V} \in \mathcal{V}$ and $E \in \mathcal{S}_V \cap \mathcal{S}_{\hat{V}}$, if P represents V, $\hat{P}$ represents $\hat{V}$, and $V(fEx) = \hat{V}(fEx)$ for all $f \in \mathcal{F}$ and $x \in \mathbb{K}$, then $P|E = \hat{P}|E$.*

Proof. It is clear that V and $\hat{V}$ agree on evaluations for acts measurable with respect to the partition Π. It follows from the uniqueness of the max-min representation that $P_\Pi = \hat{P}_\Pi$. This implies $P|E = \hat{P}|E$. □

(Necessity) We show that the contraction rule satisfies all the axioms.

Claim 1. *The contraction rule satisfies axiom ADSU.*

Proof. Consider any $V \in \mathcal{V}$ and V-non-null event E. Let V be represented by P, and thus V_E is represented by $Q^c(P,E)$. It suffices to show that if V_E is unambiguous, then for any f and g with $g \rhd^s f$ for some $s \in E$, we have $V(g) - V(f) \leq V_E(g) - V_E(f)$. Since V_E is unambiguous, by Lemma 6 and

Proposition 1, we know $\mu_P(E) \leq 1$. Thus, $Q^c(P,E) = \{\overline{\mu_P|E}\}$. Since $\overline{\mu_P|E}(s) \geq p(s)$ for each $p \in P$, it follows that

$$V(g) - V(f) = V(g) - \mathbb{E}_{p^*}(f) \leq \mathbb{E}_{p^*}(g) - \mathbb{E}_{p^*}(f)$$

$$= p^*(s)[g(s) - f(s)] \leq \overline{\mu_P|E}(s)[g(s) - f(s)] = V_E(g) - V_E(f),$$

where $p^* \in P$ minimizes the expected payoff of f, and the first inequality follows from the max-min representation of V. □

Claim 2. *The contraction rule satisfies axiom MIA.*

Proof. Consider any $V, \hat{V} \in \mathcal{V}$, event E, and $\alpha \in (0,1]$ that satisfy the conditions stated in the axiom. Let V be represented by P and $\hat{V}$ represented by $\hat{P}$. By Lemma 5, $\alpha V + (1-\alpha)\hat{V}$ is represented by $\alpha P + (1-\alpha)\hat{P}$. Since E is $\hat{V}$-null, by Lemma 7, E is $\hat{P}$-null. Since V_E is unambiguous, we know that V_E is represented by $\{\overline{\mu_P|E}\}$. Then we have $\mu_{\alpha P+(1-\alpha)\hat{P}}(E) = \alpha\mu_P(E) \leq 1$ and $\overline{\mu_{\alpha P+(1-\alpha)\hat{P}}|E} = \overline{\mu_P|E}$. Thus, we conclude that $V_E = (\alpha V + (1-\alpha)\hat{V})_E$. □

Claim 3. *The contraction rule satisfies axiom IIS.*

Proof. Consider two evaluation functions V and $\hat{V}$. Let V be represented by P and $\hat{V}$ represented by $\hat{P}$. Consider an event $E \in \mathcal{S}_V \cap \mathcal{S}_{\hat{V}}$. Suppose that for any $f \in \mathcal{F}$ and any $x \in \mathbb{K}$, $V(fEx) = \hat{V}(fEx)$. By Lemma 9, we know $P|E = \hat{P}|E$. Clearly, it implies that $Q^c(P,E) = Q^c(\hat{P},E)$, and thus $V_E = \hat{V}_E$. □

Claim 4. *The contraction rule satisfies axiom MAB.*

Proof. Consider evaluation functions $V, \hat{V}$, and $\tilde{V}$, event E, and $\alpha \in (0,1]$ that satisfy the conditions stated in the axiom. Let V be represented by P, $\hat{V}$ represented by $\hat{P}$, and $\tilde{V}$ represented by $\tilde{P}$. Let $W = \alpha V + (1-\alpha)\tilde{V}$, $\hat{W} = \alpha\hat{V} + (1-\alpha)\tilde{V}$, $P^1 = \alpha P + (1-\alpha)\tilde{P}$, and $P^2 = \alpha\hat{P} + (1-\alpha)\tilde{P}$. Since E does not strongly resolve ambiguity of W, we know by Lemma 8 that $\mu_{P^1}(E) \geq 1$. We only prove the axiom for the case where $\mu_{P^1}(E) > 1$. The case where $\mu_{P^1}(E) = 1$ can be shown similarly.

First, we show that $P^2_\Pi \subseteq \left(co(P^1 \cup Q^c(P^1,E))\right)_\Pi$. Suppose to the contrary that $P^2_\Pi \not\subseteq \left(co(P^1 \cup Q^c(P^1,E))\right)_\Pi$. Then there exists $\hat{p} \in P^2$ such that $\hat{p}_\Pi \notin \left(co(P^1 \cup Q^c(P^1,E))\right)_\Pi$. By the separating hyperplane theorem, there exists $f \in \mathcal{F}$ and $x \in \mathbb{K}$ such that $\sum_{s\in E}\hat{p}(s)f(s) + \hat{p}(E)x < \sum_{s\in E} p(s)f(s) + p(E)x$ for all

$p \in co(P^1 \cup Q^c(P^1, E))$. It follows that $\hat{W}(fEx) < \min\{W(fEx), W_E(f)\}$, a contradiction.

Next, we show that for each $p \in P^1$, there exists $\hat{p} \in P^2$ such that $\hat{p}_\Pi \in \left(co(\{p, \Phi(p|E, \mu_{P^1}|E)\})\right)_\Pi$. Suppose to the contrary that for some $p \in P^1$, $\left(co(\{p, \Phi(p|E, \mu_{P^1}|E)\})\right)_\Pi$ intersects P^2_Π at an empty set. By the separating hyperplane theorem, we can find some $f \in \mathcal{F}$ and some $x \in \mathbb{K}$ such that

$$\hat{W}(fEx) > \sum_{s \in E} p(s)f(s) + p(E)x \geq W(fEx),$$

$$\hat{W}(fEx) > \sum_{s \in E} \Phi(p|E, \mu_{P^1}|E)(s)f(s) \geq W_E(f),$$

which contradicts to that $\hat{W}(fEx) \leq \max\{W(fEx), W_E(f)\}$.

Since (i) $P^2_\Pi \subseteq \left(co(P^1 \cup Q^c(P^1, E))\right)_\Pi$, and (ii) for each $p \in P^1$, there exists $\hat{p} \in P^2$ such that $\hat{p}_\Pi \in \left(co(\{p, \Phi(p|E, \mu_{P^1}|E)\})\right)_\Pi$, it follows from Lemma 3 that $Q^c(P, E) = Q^c(\hat{P}, E)$. Thus we conclude that $V_E = \hat{V}_E$. □

Claim 5. *The contraction rule satisfies axiom NAI.*

Proof. Consider V, partition $\Pi^* = \{S_i\}_{i=1}^n$, and event E that satisfy the conditions stated in the axiom. Let V be represented by P. Since V is unambiguous with respect to $\{S_i\}_{i=1}^n$, we know P_{Π^*} is a singleton. It follows that $\mu_P(E) = \sum_{s \in E} (\max_{p \in P} p(s)) \leq \sum_{i=1}^n (\max_{p \in P} p(S_i)) = 1$, and thus V_E is unambiguous. □

(Sufficiency) In what follows, we show that the axioms are sufficient for an updating rule Γ to be the contraction rule. We assume throughout the proof that Γ satisfies all the axioms. We use V_E to denote the ex post evaluation function under the updating rule Γ.

Lemma 10. *For any $V \in \mathcal{V}$ and event E, if P represents V and $\mu_P(E) > 1$, then V_E is ambiguous.*

Proof. Consider V, P, and E that satisfy the conditions of the lemma. Since E is finite and S is infinite, there exists finite $\hat{E} \subseteq S$ such that $|E| = |\hat{E}|$ and $E \cap \hat{E} = \emptyset$. Consider an arbitrary isomorphism $\tau : E \to \hat{E}$ and let $\hat{s} = \tau(s)$ for each $s \in E$. For each $p \in P$, define $\hat{p}$ such that

$$\hat{p}(s) = p(s), \forall s \in E,$$

$$\hat{p}(\hat{s}) = \Phi(p|E, \mu_P|E)(s) - p(s), \forall \hat{s} \in \hat{E}.$$

Each $\hat{p}$ is well-defined since $\Phi(p|E, \mu_P|E)(s) \geq p(s)$ for all $s \in E$. It follows that (i) $\hat{p}(E \cup \hat{E}) = 1$ for each $\hat{p} \in \hat{P}$, (ii) $P_\Pi = \hat{P}_\Pi$, and (iii) $\hat{P}_{\Pi^*} = \Big(Q^c(P, E)\Big)_{\Pi^*}$, where $\Pi^* = \{\{s, \hat{s}\} : s \in E\} \cup \{S \backslash (E \cup \hat{E})\}$. Also, note that $\hat{P}$ is closed, and thus $co(\hat{P}) \in \mathscr{P}$. Let $\hat{V}$ be the evaluation function that is represented by $co(\hat{P})$. Since $\hat{P}_\Pi = P_\Pi$, we know $V(fEx) = \hat{V}(fEx)$ for all $f \in \mathcal{F}$ and $x \in \mathbb{K}$. By axiom IIS, to show that V_E is ambiguous, it suffices to show that $\hat{V}_E$ is ambiguous.

Suppose to the contrary that $\hat{V}_E$ is unambiguous and represented by $\{q\}$. Fix some $s \in E$. Consider act f such that $f(s) = 0$, $f(\hat{s}) = 1$, and $f(\tilde{s}) = 0$ for all $\tilde{s} \in S \backslash \{s, \hat{s}\}$. It is clear that $\hat{V}_E(f) = 0$ since f equals 0 on E. We argue that $\hat{V}(f) = 0$. Note that

$$\hat{V}(f) = \min_{\tilde{p} \in co(\hat{P})} \mathbb{E}_{\tilde{p}}(f) = \min_{\tilde{p} \in co(\hat{P})} \tilde{p}(\hat{s}) = 0,$$

where the second equality holds since for any $p \in P$ satisfying $p(s) = \mu_P(s)$, we have $\hat{p}(\hat{s}) = \Phi(p|E, \mu_P|E)(s) - p(s) = 0$. Define $\hat{P}^* \subseteq \hat{P}$ such that $\hat{p} \in \hat{P}^*$ if and only if $p(s) = \mu_P(s)$. It follows that

$$\arg \min_{\tilde{p} \in co(\hat{P})} \tilde{p}(\hat{s}) = co(\hat{P}^*).$$

To proceed, for each $\epsilon \in (0, 1)$, define act f^ϵ such that $f^\epsilon(s) = \epsilon$, $f^\epsilon(\hat{s}) = 1$, and $f^\epsilon(\tilde{s}) = 0$ for all $\tilde{s} \in S \backslash \{s, \hat{s}\}$. Since $\hat{V}$ is concave and $co(\hat{P}^*)$ contains all the supergradients of $\hat{V}$ at f, we know that when ϵ is sufficiently small,

$$\hat{V}(f^\epsilon) - \hat{V}(f) = \min_{\tilde{p} \in co(\hat{P}^*)} \mathbb{E}_{\tilde{p}}(f^\epsilon - f) + o(\epsilon),$$

where $o(\epsilon)$ denotes an infinitesimal term of ϵ. Since for each $\tilde{p} \in co(\hat{P}^*)$, $\tilde{p}(s) = \mu_P(s)$, the above equality can be written as

$$\hat{V}(f^\epsilon) - \hat{V}(f) = \mu_P(s)\epsilon + o(\epsilon).$$

Hence, for any $k < \mu_P(s)$, we can find small enough ϵ such that $\hat{V}(f^\epsilon) - \hat{V}(f) > k\epsilon$. Since $\hat{V}_E$ is represented by $\{q\}$, we know $\hat{V}_E(f^\epsilon) - \hat{V}_E(f) = q(s)\epsilon$. It follows from axiom ADSU that $q(s) \geq \mu_P(s)$. By a similar argument, we can show that for each $\tilde{s} \in E$, $q(\tilde{s}) \geq \mu_P(\tilde{s})$, which is impossible since $\mu_P(E) > 1$. Therefore, we conclude that $\hat{V}_E$ is ambiguous, and so is V_E. □

Lemma 11. *For any $V \in \mathcal{V}$ and V-non-null event E, if P represents V and $\mu_P(E) \leq 1$, then V_E is unambiguous, and if $\mu_P(E) = 1$, then V_E is represented by $\{\mu_P|E\}$.*

Proof. Consider V, E, and P that satisfy the conditions of the first statement of the lemma. Since E is V-non-null, $\overline{\mu_P|E}$ is well-defined. In addition, since $\mu_P(E) \leq 1$, we have $\overline{\mu_P|E}(s) \geq \mu_P|E(s) \geq p(s)$ for each $p \in P$ and each $s \in E$. Since S is infinite and E is finite, there exists finite $\hat{E} \subseteq S$ such that $|E| = |\hat{E}|$ and $E \cap \hat{E} = \emptyset$. Consider an arbitrary isomorphism $\tau : E \to \hat{E}$ and let $\hat{s} = \tau(s)$ for each $s \in E$. Define $\hat{P}$ as follows. For each $p \in P$, define some $\hat{p} \in \hat{P}$ such that $\hat{p}(s) = p(s)$ for each $s \in E$ and $\hat{p}(\hat{s}) = \overline{\mu_P|E}(s) - p(s)$ for each $\hat{s} \in \hat{E}$. It is easy to verify that $\hat{P} \in \mathscr{P}$ and $\hat{P}|E = P|E$. Let $\hat{V}$ be the evaluation function that is represented by $\hat{P}$. It follows that $V(fEx) = \hat{V}(fEx)$ for all $f \in \mathcal{F}$ and $x \in \mathbb{K}$. By axiom IIS, $V_E = \hat{V}_E$. Note that for each $\hat{p} \in \hat{P}$ and each $s \in E$, $\hat{p}(\{s, \hat{s}\}) = \overline{\mu_P|E}(s)$. It follows that $\hat{V}$ is unambiguous with respect to the partition $\{\{s, \hat{s}\} : s \in E\} \cup \{S \backslash (E \cup \hat{E})\}$. By axiom NAI, $\hat{V}_E$ is unambiguous, and so is V_E.

If $\mu_P(E) = 1$, then $\mu_{\hat{P}}(E) = 1$. It suffices to show that $\hat{V}_E$ is represented by $\{\mu_P|E\}$. Suppose that $\hat{V}_E$ is represented by $\{q\}$. Fix some $s \in E$. Consider acts f and g such that $f(s) = 0$, $f(\hat{s}) = 1$, $f(\tilde{s}) = 0$ for all $\tilde{s} \in S\backslash\{s, \hat{s}\}$, $g(s) = 1/2$, $g(\hat{s}) = 1$, and $g(\tilde{s}) = 0$ for all $\tilde{s} \in S\backslash\{s, \hat{s}\}$. Note that

$$\begin{aligned}\hat{V}(g) = \min_{\hat{p}\in\hat{P}} \mathbb{E}_{\hat{p}}(g) = \min_{\hat{p}\in\hat{P}} \left(\frac{1}{2}\hat{p}(s) + \hat{p}(\hat{s})\right) &= \min_{\hat{p}\in\hat{P}} \left(\frac{1}{2}\hat{p}(s) + \mu_P(s) - \hat{p}(s)\right) \\ &= \mu_P(s) - \frac{1}{2}\max_{\hat{p}\in\hat{P}} \hat{p}(s) = \mu_P(s) - \frac{1}{2}\mu_P(s) = \frac{1}{2}\mu_P(s).\end{aligned}$$

Similarly, we have $\hat{V}(f) = 0$. Since $\hat{V}_E$ is represented by $\{q\}$, we know $\hat{V}_E(f) = 0$ and $\hat{V}_E(g) = \frac{1}{2}q(s)$. Since $\hat{V}(f) = \hat{V}_E(f)$ and $\hat{V}_E$ is unambiguous, axiom ADSU implies that $\hat{V}(g) \leq \hat{V}_E(g)$. That is, $\mu_P(s) \leq q(s)$. By a similar argument, for each $\tilde{s} \in E$, we should have $\mu_P(\tilde{s}) \leq q(\tilde{s})$. Hence, we conclude that $\mu_P|E = q$. □

The next two lemmas establish the sufficiency part of the proof.

Lemma 12. *For any $V \in \mathcal{V}$ and V-non-null event E, if P represents V and $\mu_P(E) < 1$, then V_E is represented by $\{\overline{\mu_P|E}\}$.*

Proof. Consider V, E, and P that satisfy the conditions of the lemma. Fix some $s^* \in S \backslash E$. Define $\hat{P}$ such that (i) $\hat{P}|E = P|E$ and (ii) $\hat{p}(E \cup s^*) = 1$ for all $\hat{p} \in \hat{P}$. Note that $\hat{P}$ can be constructed by defining $\hat{p} \in \hat{P}$ for each $p \in P$ such that $\hat{p}(s) = p(s)$ for each $s \in E$ and $\hat{p}(s^*) = 1 - p(E)$. Since $P \in \mathscr{P}$, we know that $\hat{P} \in \mathscr{P}$. Let $\hat{V}$ be the evaluation function that is represented by $\hat{P}$. $P|E = \hat{P}|E$ implies that $V(fEx) = \hat{V}(fEx)$ for all $f \in \mathcal{F}$ and $x \in \mathbb{K}$. By axiom IIS, we have $V_E = \hat{V}_E$.

To proceed, we construct another set of priors $\tilde{P}$ as follows. Let $\tilde{p} \in \tilde{P}$ if and only if there exists $\hat{p} \in \hat{P}$ such that $\tilde{p}(s) = \hat{p}(s)/\mu_{\hat{P}}(E)$ for each $s \in E$ and $\tilde{p}(s^*) = 1 - \tilde{p}(E)$. It is easy to verify that $\tilde{P} \in \mathscr{P}$. Note that $\hat{P} = \alpha\tilde{P} + (1-\alpha)\{q\}$, where $\alpha = \mu_{\hat{P}}(E)$ and $q(s^*) = 1$. Let $\tilde{V}$ be the evaluation function that is represented by $\tilde{P}$. By Lemma 5, we know $\hat{V} = \alpha\tilde{V} + (1-\alpha)V^*$, where V^* is represented by $\{q\}$. Since E is $\{q\}$-null and $\mu_{\tilde{P}}(E) = 1$ (which means that $\tilde{V}_E$ is unambiguous by Lemma 11), axiom MIA implies that $\hat{V}_E = \tilde{V}_E$. By Lemma 11, $\tilde{V}_E$ is represented by $\{\overline{\mu_P|E}\}$, and so are $\hat{V}_E$ and V_E. □

Lemma 13. *For any $V \in \mathcal{V}$ and V-non-null event E, if P represents V and $\mu_P(E) > 1$, then V_E is represented by $Q^c(P, E) = \{\Phi(p|E, \mu_P|E) : p \in P\}$.*

Proof. Consider V, E, and P that satisfy the conditions of the lemma. Assume that for some $s^* \in S \backslash E$, $p(E \cup s^*) = 1$ for all $p \in P$. By axiom IIS, imposing this assumption will not affect V_E . By Proposition 2, $Q^c(P, E)$ is an element of $\mathscr{P}$. Hence, there exists an evaluation function $\hat{V}$ that is represented by $Q^c(P, E)$. Consider $q^* \in \Delta(S)$ such that $q^*(s^*) = 1$ and let $\tilde{V}$ be represented by $\{q^*\}$. Let $\alpha = 1/\mu_P(E) \in (0, 1)$ and define $W = \alpha V + (1-\alpha)\tilde{V}$ and $\hat{W} = \alpha\hat{V} + (1-\alpha)\tilde{V}$. By Lemma 5, W is represented by $\alpha P + (1-\alpha)\{q^*\}$ and $\hat{W}$ represented by $\alpha Q^c(P, E) + (1-\alpha)\{q^*\}$. Note that $\mu_{\alpha P+(1-\alpha)\{q^*\}}(E) = 1$. By Lemma 8, E does not strongly resolve ambiguity of W.

To proceed, we show that for all $f \in \mathcal{F}$ and $x \in \mathbb{K}$, $\min\{W(fEx), W_E(f)\} \leq \hat{W}(fEx) \leq \max\{W(fEx), W_E(f)\}$. Consider any $f \in \mathcal{F}$ and any $x \in \mathbb{K}$. Note that

(i) $\forall q \in Q^c(P, E)$, $\exists t \in [0, 1]$ and $\exists p \in P$ such that $tp|E + (1-t)\mu_P|E = q$, and

(ii) $\forall p \in P$, $\exists t \in [0, 1]$ and $\exists q \in Q^c(P, E)$ such that $tp|E + (1-t)\mu_P|E = q$.

It follows that

(iii) $\forall q \in \alpha Q^c(P,E) + (1-\alpha)\{q^*\}$, $\exists t \in [0,1]$ and $\exists p \in \alpha P + (1-\alpha)\{q^*\}$ such that $tp|E + (1-t)\mu_{\alpha P+(1-\alpha)\{q^*\}}|E = q|E$, and

(iv) $\forall p \in \alpha P + (1-\alpha)\{q^*\}$, $\exists t \in [0,1]$ and $\exists q \in \alpha Q^c(P,E) + (1-\alpha)\{q^*\}$ such that $tp|E + (1-t)\mu_{\alpha P+(1-\alpha)\{q^*\}}|E = q|E$.

Note that $\mu_{\alpha P+(1-\alpha)\{q^*\}}|E = \overline{\mu_P|E}$. Thus, condition (iii) implies that $\alpha Q^c(P,E) + (1-\alpha)\{q^*\} \subseteq co\Big((\alpha P + (1-\alpha)\{q^*\}) \cup \{\overline{\mu_P|E}\}\Big)$. By Lemma 11, W_E is represented by $\{\overline{\mu_P|E}\}$. It follows that $\hat{W}(fEx) \geq \min\{W(fEx), W_E(f)\}$.

Next, we argue that condition (iv) implies that $\hat{W}(fEx) \leq \max\{W(fEx), W_E(f)\}$. To see this, let $W(fEx) = \mathbb{E}_p(fEx)$ for some $p \in \alpha P + (1-\alpha)\{q^*\}$. By condition (iv), there exists $t \in [0,1]$ and $q \in \alpha Q^c(P,E) + (1-\alpha)\{q^*\}$ such that $tp|E + (1-t)\{\overline{\mu_P|E}\} = q|E$. That is, $tp + (1-t)\{\overline{\mu_P|E}\} = q$. It follows that $\hat{W}(fEx) \leq \mathbb{E}_q(fEx) = t\mathbb{E}_p(fEx) + (1-t)\mathbb{E}_{\overline{\mu_P|E}}(f) = tW(fEx) + (1-t)W_E(f) \leq \max\{W(fEx), W_E(f)\}$.

Since $\min\{W(fEx), W_E(f)\} \leq \hat{W}(fEx) \leq \max\{W(fEx), W_E(f)\}$ for all $f \in \mathcal{F}$ and $x \in \mathbb{K}$, by axiom MAB, we know that $V_E = \hat{V}_E$. Since $\hat{V}$ is represented by $Q^c(P,E)$, which means that $q(E) = 1$ for all $q \in Q^c(P,E)$, we know that $\hat{V}_E = \hat{V}$. Therefore, we conclude that V_E is represented by $Q^c(P,E)$. □

□

Proof of Proposition 5. Let $D = \{d_1, d_2\}$. The DM's prior beliefs over D can be fully captured by $[1 - \max_{p\in P} p(\{d_2\}\times\Theta), \max_{p\in P} p(\{d_1\}\times\Theta)]$: the probability of d_1 is at least $1 - \max_{p\in P} p(\{d_2\}\times\Theta)$ and at most $\max_{p\in P} p(\{d_1\}\times\Theta)$. After the DM observes some signal θ, there are two cases to be considered. First, the DM's contraction posterior set is a singleton. Clearly, there is no dilation in this case. Second, the DM's contraction posterior set is not a singleton. In this case, the DM's ex post beliefs over D are given by $[1 - \max_{p\in P} p(d_2,\theta), \max_{p\in P} p(d_1,\theta)]$: the probability of d_1 is at least $1 - \max_{p\in P} p(d_2,\theta)$ and at most $\max_{p\in P} p(d_1,\theta)$. Clearly, the ex post belief set does not strictly contain the ex ante one. □

Proof of Proposition 6. The first case is trivial since when the DM has no prior ambiguity over D, her contraction posterior set is a singleton. Consider the second case of the proposition. Note that the primitive conditions imply that $\max_{p\in P} p(\{d\}\times\Theta) > \max_{p\in P} p(d,\theta)$ for all $d \in D$ and $\theta \in \Theta$. Thus, $\max_{p\in P} p(\{d\}\times\Theta) > \mu_P(d,\theta)$. When θ is observed, if contraction posterior set is a singleton, then clearly there is no dilation. If not, then $\max_{p\in P} p(\{d\}\times\Theta) > \mu_P(d,\theta)$

implies the ex ante maximal probability of d is strictly higher than the ex post one of d. Therefore, there is no dilation of payoff-relevant belief sets. □

Proof of Proposition 7. The DM's evaluation of f is equal to the minimal probability of state d_1. The DM's evaluation of g is given by the minimal probability of state d_2.

First, consider scenario 1 where the information is ambiguous. Note that the maximal probability of $(d_1, \hat{d}_1)$ is $\bar{r}H$ and that of $(d_2, \hat{d}_1)$ is $(1-\underline{r})(1-L)$. The values of v_f^a and v_g^a depend on whether $\hat{d}_1$ resolves the DM's ambiguity or not, i.e., whether $\bar{r}H + (1-\underline{r})(1-L)$ is less than one or not. If $\bar{r}H + (1-\underline{r})(1-L) \leq 1$, then we have $v_f^a = \frac{\bar{r}H}{\bar{r}H+(1-\underline{r})(1-L)}$ and $v_g^a = \frac{(1-\underline{r})(1-L)}{\bar{r}H+(1-\underline{r})(1-L)}$. If $\bar{r}H + (1-\underline{r})(1-L) > 1$, then we have $v_f^a = 1-(1-\underline{r})(1-L)$ and $v_g^a = 1-\bar{r}H$.

Next, consider scenario 2. With signal $\hat{d}_1$, the maximal probability of $(d_1, \hat{d}_1)$ is $\bar{r}(H+L)/2$ and that of $(d_2, \hat{d}_1)$ is $(1-\underline{r})(2-H-L)/2$. Since $\bar{r}(H+L)/2 + (1-\underline{r})(2-H-L)/2 \leq 1$, the realization of $\hat{d}_1$ resolves the ambiguity. Thus we have $v_f^u = \frac{\bar{r}(H+L)}{\bar{r}(H+L)+(1-\underline{r})(2-H-L)}$ and $v_g^u = \frac{(1-\underline{r})(2-H-L)}{\bar{r}(H+L)+(1-\underline{r})(2-H-L)}$.

It can be easily verified that $\frac{H+L}{2-H-L} > \frac{H}{1-L}$ whenever $H+L > 1$. Thus when $\bar{r}H + (1-\underline{r})(1-L) \leq 1$, $v_f^a = \frac{\bar{r}H}{\bar{r}H+(1-\underline{r})(1-L)} < \frac{\bar{r}(H+L)}{\bar{r}(H+L)+(1-\underline{r})(2-H-L)} = v_f^u$. When $\bar{r}H + (1-\underline{r})(1-L) > 1$, we have $v_f^a = 1-(1-\underline{r})(1-L) \leq \frac{\bar{r}H}{\bar{r}H+(1-\underline{r})(1-L)} < \frac{\bar{r}(H+L)}{\bar{r}(H+L)+(1-\underline{r})(2-H-L)} = v_f^u$, where the first inequality holds since $\frac{v_f^a}{1-v_f^a} = \frac{1-(1-\underline{r})(1-L)}{(1-\underline{r})(1-L)} \leq \frac{\bar{r}H}{(1-\underline{r})(1-L)}$. The comparison between v_g^a and v_g^u follows from similar calculations. □

Proof of Proposition 8. We only prove that for any $\epsilon > 0$, the probability of $\min_{r \in I_{d_1}^n} \geq 1-\epsilon$ goes to one. When n is large enough, the contraction posterior set of the DM is a singleton. With signal sequence $(\hat{d}^1, ..., \hat{d}^n)$, the contraction posterior likelihood ratio between d_1 and d_2 is $\frac{\bar{r}}{1-\underline{r}} \prod_{t=1}^n (H/(1-L))^{\mathbf{1}[\hat{d}^t=\hat{d}_1]} \prod_{t=1}^n ((1-L)/H)^{\mathbf{1}[\hat{d}^t=\hat{d}_2]}$. Thus, $\min_{r \in I_{d_1}^n} \geq 1-\epsilon$ if and only if $\prod_{t=1}^n (H/(1-L))^{\mathbf{1}[\hat{d}^t=\hat{d}_1]} \prod_{t=1}^n ((1-L)/H)^{\mathbf{1}[\hat{d}^t=\hat{d}_2]} \geq \frac{1-\underline{r}}{\bar{r}}\frac{1-\epsilon}{\epsilon}$, i.e., $\sum_{t=1}^n (\ln((H/(1-L))^{\mathbf{1}[\hat{d}^t=d_1]}) + \ln(((1-L)/H)^{\mathbf{1}[\hat{d}^t=d_2]})) \geq \ln\left(\frac{(1-\epsilon)(1-\underline{r})}{\epsilon\bar{r}}\right)$. Since $\{\ln((H/(1-L))^{\mathbf{1}[\hat{d}^t=d_1]}) + \ln(((1-L)/H)^{\mathbf{1}[\hat{d}^t=d_2]})\}_{t=1}^n$ is a sequence of independent random variables with positive mean, by law of large number, we know that the probability of the above inequality goes to one as n goes to infinity. □

6.2 Appendix B: Full-Bayesian Rule

Let V_E^{fb} denote the FB ex post evaluation function of V when the realized event is E. For any $P \in \mathscr{P}$ and any P-non-null event E, the FB posterior set is $Q^{fb}(P,E) = cl\Big(\{\overline{p|E} : p \in P, p(E) > 0\}\Big) \in \mathscr{P}$. We say that E is strictly V-non-null if $V(f) < V(g)$ for all f, g satisfying that $g(s) \geq f(s)$ for all $s \in S$ and $g(s') > f(s')$ for all $s' \in E$. It can be shown that when V is represented by P, E is strictly V-non-null if and only if $\min_{p\in P} p(E) > 0$. When $\min_{p\in P} p(E) > 0$, the set $\{\overline{p|E} : p \in P\}$ is closed and thus $Q^{fb}(P,E) = \{\overline{p|E} : p \in P\}$. We fix partition $\Pi = \{\{s\} : s \in E\} \cup \{S \backslash E\}$.

Proposition 9. *FB satisfies axiom ADSU.*

Proof. Let $V \in \mathcal{V}$ be represented by P and event E be V-non-null. If V_E^{fb} is unambiguous, then $Q^{fb}(P,E)$ is a singleton. Let $Q^{fb}(P,E) = \{q\}$. Consider any $f, g \in \mathcal{F}$ such that $V(f) = V_E^{fb}(f)$ and $g \rhd^s f$ for some $s \in E$. Let $p^* \in P$ satisfy $p^*(s) \geq p(s)$ for all $p \in P$. It follows that $V(g) - V(f) \leq p^*(s)(g(s) - f(s))$. Since either $p^*(s) = 0$ or $\overline{p^*|E} = q$, we have $q(s) \geq p^*(s)$. Thus $V_E^{fb}(g) - V_E^{fb}(f) = q(s)(g(s) - f(s)) \geq V(g) - V(f)$. By $V_E^{fb}(f) = V(f)$, we know $V_E^{fb}(g) \geq V(g)$. □

Proposition 10. *FB satisfies axiom MI.*

Proof. Consider V, $\hat{V}$, α, and E that satisfy the conditions stated in axiom MI. Let V and $\hat{V}$ be represented by P and $\hat{P}$ respectively. Since E is $\hat{V}$-null, by Lemma 7, it is $\hat{P}$-null. It follows that $(\alpha P + (1-\alpha)\hat{P})|E = \alpha P|E$. Therefore, we have $\{\overline{p|E} : p \in P, p(E) > 0\} = \{\overline{\alpha p|E} : p \in P, p(E) > 0\} = \{\overline{\tilde{p}|E} : \tilde{p} \in \alpha P + (1-\alpha)\hat{P}, \tilde{p}(E) > 0\}$. Hence, we conclude that $V_E^{fb} = (\alpha V + (1-\alpha)\hat{V})_E^{fb}$. □

Proposition 11. *FB satisfies axiom B.*

Proof. Consider V, $\hat{V}$, and E that satisfy the conditions stated in axiom B. Let V and $\hat{V}$ be represented by P and $\hat{P}$ respectively. Then V_E^{fb} is represented by $Q^{fb}(P,E) = cl\Big(\{\overline{p|E} : p \in P, p(E) > 0\}\Big)$. Clearly, we have

$$\bigcup_{p \in P: p(E) > 0} co\Big(\{p, \overline{p|E}\}\Big)_\Pi \subseteq co\Big(P \cup Q^{fb}(P,E)\Big)_\Pi.$$

Since $\min\{V(fEx), V_E^{fb}(f)\} \leq \hat{V}(fEx)$ for all $f \in \mathcal{F}$ and $x \in \mathbb{K}$, we know $\hat{P}_\Pi \subseteq co\Big(P \cup Q^{fb}(P,E)\Big)_\Pi$. It follows that $Q^{fb}(\hat{P},E) \subseteq Q^{fb}(P,E)$. Next, since $\max\{V(fEx), V_E^{fb}(f)\} \geq \hat{V}(fEx)$ for all $f \in \mathcal{F}$ and $x \in \mathbb{K}$, by a similar proof as that of Claim 4, we know that for each $p \in P$ with $p(E) > 0$, there exists $\hat{p} \in \hat{P}$ such that $\hat{p}_\Pi \in co\Big(\{p, \overline{p|E}\}\Big)_\Pi$. It implies $\{\overline{p|E} : p \in P, p(E) > 0\} \subseteq \{\overline{\hat{p}|E} : \hat{p} \in \hat{P}, \hat{p}(E) > 0\}$, and thus $Q^{fb}(P, E) \subseteq Q^{fb}(\hat{P}, E)$. Hence we conclude that $V_E^{fb} = \hat{V}_E^{fb}$. □

Proposition 12. *If an updating rule* Γ *satisfies axioms MI and B, then* $\Gamma(V, E) = V_E^{fb}$ *for any strictly* V*-non-null event* E*.*

Proof. Consider $V \in \mathcal{V}$ and an event E that is strictly V-non-null. Let P represent V. If $p(E) = 1$ for each $p \in P$, then clearly we have $\Gamma(V, E) = V_E^{fb}$. If $p(E) < 1$ for some $p \in P$, since $\min_{p \in P} p(E) > 0$, we know $Q^{fb}(P, E) = \{\overline{p|E} : p \in P\}$. Let $\hat{V}$ be the evaluation function that is represented by $Q^{fb}(P, E)$. Consider $\tilde{V}$ such that $\tilde{V}$ is represented by $\{q^*\}$ where $q^*(E) = 0$. Let $0 < \alpha < \min_{p \in P} p(E)$ and define $W = \alpha\hat{V} + (1 - \alpha)\tilde{V}$. By the construction, W is represented by $\{\alpha\overline{p|E} + (1 - \alpha)q^* : p \in P\}$. By axiom MI, $\Gamma(W, E) = \Gamma(\hat{V}, E)$, and thus $\Gamma(W, E)$ is represented by $Q^{fb}(P, E)$, i.e., $\Gamma(W, E) = \hat{V}$. It remains to show that

$$\min\{W(fEx), \Gamma(W, E)(f)\} \leq V(fEx) \leq \max\{W(fEx), \Gamma(W, E)(f)\}$$

for all $f \in \mathcal{F}$ and $x \in \mathbb{K}$ (then by axiom B, $\Gamma(V, E) = \Gamma(W, E) = V_E^{fb}$, and we are done). Fix some f and x. Since $\alpha < \min_{p \in P} p(E)$, we have $P_\Pi \subseteq co\Big(Q^{fb}(P, E) \cup \{\alpha\overline{p|E} + (1 - \alpha)q^* : p \in P\}\Big)_\Pi$. It follows that $\min\{W(fEx), \Gamma(W, E)(f)\} \leq V(fEx)$. Next, let $p^* \in P$ satisfy that $\overline{p^*|E} \in \arg\min_{q \in Q^{fb}(P,E)} \mathbb{E}_q(f)$. It follows that $\mathbb{E}_{p^*}(fEx) = p^*(E)\Gamma(W, E)(f) + (1 - p^*(E))x$. Since $W(fEx) = \alpha\Gamma(W, E)(f) + (1 - \alpha)x$ and $\alpha < p^*(E)$, we know that $\mathbb{E}_{p^*}(fEx)$ is a convex combination of $W(fEx)$ and $\Gamma(W, E)(f)$. Hence, $\mathbb{E}_{p^*}(fEx) \leq \max\{W(fEx), \Gamma(W, E)(f)\}$, and this implies $V(fEx) \leq \max\{W(fEx), \Gamma(W, E)(f)\}$. □

References

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

Blume, A., and O. Board (2014): "Intentional Vagueness," *Erkenntnis*, 79(4), 855–899.

Bose, S., and L. Renou (2014): "Mechanism Design With Ambiguous Communication Devices," *Econometrica*, 82(5), 1853–1872.

Chen, J. Y. (2021): "Sequential Learning under Informational Ambiguity," *Available at SSRN 3480231.*

Cheng, X. (2022): "Relative Maximum Likelihood Updating of Ambiguous Beliefs," *Journal of Mathematical Economics*, 99, 102587.

Chew, S. H., and J. S. Sagi (2008): "Small Worlds: Modeling Attitudes Toward Sources of Uncertainty," *Journal of Economic Theory*, 139(1), 1–24.

Cohen, M., I. Gilboa, J.-Y. Jaffray, and D. Schmeidler (2000): "An Experimental Study of Updating Ambiguous Beliefs," *Risk, Decision and Policy*, 5(2), 123–133.

Cripps, M. W. (2019): "Divisible Updating," *Working Paper.*

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

Dominiak, A., P. Duersch, and J.-P. Lefort (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*, 70(4), 643–669.

Epstein, L. G., and Y. Halevy (2021): "Hard-to-Interpret Signals," *Working Paper.*

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

——— (2007): "Learning under Ambiguity," *The Review of Economic Studies*, 74(4), 1275–1303.

——— (2008): "Ambiguity, Information Quality, and Asset Pricing," *The Journal of Finance*, 63(1), 197–228.

Ert, E., and S. T. Trautmann (2014): "Sampling Experience Reverses Preferences for Ambiguity," *Journal of Risk and Uncertainty*, 49(1), 31–42.

Gilboa, I., and M. Marinacci (2013): *Ambiguity and the Bayesian Paradigm*vol. 1 of *Econometric Society Monographs*, pp. 179–242. Cambridge University Press.

Gilboa, I., and D. Schmeidler (1989): "Maxmin Expected Utility with Non-unique Prior," *Journal of Mathematical Economics*, 18(2), 141–153.

——— (1993): "Updating Ambiguous Beliefs," *Journal of Economic Theory*, 59(1), 33–49.

Gul, F., and W. Pesendorfer (2014): "Expected Uncertain Utility Theory," *Econometrica*, 82(1), 1–39.

——— (2021): "Evaluating Ambiguous Random Variables from Choquet to Maxmin Expected Utility," *Journal of Economic Theory*, 192, 105129.

Hanany, E., and P. Klibanoff (2007): "Updating Preferences with Multiple Priors," *Theoretical Economics*, 2(3), 261–298.

——— (2009): "Updating Ambiguity Averse Preferences," *The BE Journal of Theoretical Economics*, 9.

Izhakian, Y. (2020): "A Theoretical Foundation of Ambiguity Measurement," *Journal of Economic Theory*, 187, 105001.

Jaffray, J.-Y. (1988): "Application of Linear Utility Theory to Belief Functions," in *International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems*, pp. 1–8. Springer.

Jaffray, J.-Y. (1992): "Bayesian Updating and Belief Functions," *IEEE Transactions on Systems, Man, and Cybernetics*, 22(5), 1144–1152.

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

Kellner, C., and M. T. Le Quement (2017): "Modes of Ambiguous Communication," *Games and Economic Behavior*, 104, 271–292.

——— (2018): "Endogenous Ambiguity in Cheap Talk," *Journal of Economic Theory*, 173, 1–17.

Kellner, C., M. T. Le Quement, and G. Riener (2019): "Reacting to Ambiguous Messages: An Experimental Analysis," *Working Paper*.

Keynes, J. M. (1921): *A Treatise on Probability*. Macmillan and Company, limited.

Knight, F. H. (1921): *Risk, Uncertainty and Profit*, vol. 31. Houghton Mifflin.

Kovach, M. (2021): "Ambiguity and Partial Bayesian Updating," *Available at arXiv:2102.11429.*

Liang, Y. (2021): "Learning from Unknown Information Sources," *Available at SSRN 3314789.*

Maccheroni, F., M. Marinacci, and A. Rustichini (2006): "Ambiguity Aversion, Robustness, and the Variational Representation of Preferences," *Econometrica*, 74(6), 1447–1498.

Moreno, O. M., and Y. Rosokha (2015): "Learning under Compound Risk vs. Learning under Ambiguity-An Experiment," *Journal of Risk and Uncertainty*, 53, 137–162.

Ortoleva, P. (2012): "Modeling the Change of Paradigm: Non-Bayesian Reactions to Unexpected News," *American Economic Review*, 102(6), 2410–2436.

Pires, C. P. (2002): "A Rule for Updating Ambiguous Beliefs," *Theory and Decision*, 53(2), 137–152.

Schmeidler, D. (1989): "Subjective Probability and Expected Utility without Additivity," *Econometrica*, pp. 571–587.

Shafer, G. (1976): *A Mathematical Theory of Evidence*, vol. 42. Princeton university press.

Shishkin, D., and P. Ortoleva (2021): "Ambiguous Information and Dilation: An Experiment," *Working Paper.*

Tang, R. (2020): "A Theory of Updating Ambiguous Information," *memo.*

Wasserman, L. A., and J. B. Kadane (1990): "Bayes' Theorem for Choquet Capacities," *The Annals of Statistics*, pp. 1328–1339.