A Theory of Contraction Updating[*]

Rui Tang [†]

April 5, 2024

Abstract

We propose and axiomatize a new rule, the contraction rule, for belief updating in the presence of ambiguity. With the rule, a realized event renders an individual's belief unambiguous if and only if the event has small ambiguity. The rule is insensitive to unlikely priors, independent of the order in which multiple pieces of information arrive, and consistent with recent experimental findings on updating ambiguous information. We also show that the rule can be utilized by a dynamically inconsistent individual to maximally align her future choices with her current preference through suitable ambiguous information structures.

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

JEL CODES: D01, D80

[*]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, Joyee Deb, Simon Grant, Gaoji Hu, Shaowei Ke, Peter Klibanoff, Alessandro Lizzeri, Xiaosheng Mu, John K.-H. Quah, Satoru Takahashi, Jingni Yang, Leeat Yariv, Mu Zhang and Chen Zhao for helpful comments and discussions. I am grateful to the financial support by the Stephen Goldfeld Memorial Graduate Summer Fellowship of Princeton University and the start-up research fund of Hong Kong University of Science and Technology. All errors are my own.

[†]Department of Economics, The Hong Kong University of Science and Technology, Contact Email: ruitang.economics@gmail.com

1 Introduction

Ambiguity refers to the situation in which states of the world have no objective probability distribution. Since the work of Knight (1921), Keynes (1921) and Ellsberg (1961), many models have been proposed to rationalize decision makers' (henceforth DM) choices over bets on ambiguous states.[1] In a seminal paper, Gilboa and Schmeidler (1989) (henceforth GS) propose the maxmin expected utility (MEU) model according to which the DM behaves as if she forms a set of priors over the states of the world and evaluates each prospect according to its minimal expected utility over all her priors. In a wide variety of applications, the assumption that economic agents are MEU maximizers leads to novel economic implications.[2]

The prevalence of ambiguity in various economic scenarios and the critical role of information in economic interactions underscore the significance of comprehending how DMs react to new information in the presence of ambiguity. The relevance of this issue is further accentuated by the ubiquity of ambiguous information, as argued by Epstein and Halevy (2024). In this paper, we focus on MEU maximizers and explore how they update their beliefs when new information arrives.

Our study is inspired by recent experimental evidence on how individuals react to ambiguous information. A debatable issue related to ambiguous information is whether it can lead to the dilation of the DM's set of beliefs over payoff-relevant states, thereby increasing her payoff-relevant ambiguity (Wasserman and Kadane, 1990). The two benchmark updating rules in the MEU framework—the Full-Bayesian rule (henceforth FB) and Maximum Likelihood (henceforth ML)—allow for such dilation even when there is no ex-ante ambiguity on the payoff-relevant states.[3] By contrast, Shishkin and Ortoleva (2023) (henceforth SO23) find that ambiguity averse subjects typically do not dilate their payoff-relevant belief sets after receiving ambiguous signals.[4] The new updating rule that we propose in this paper—the contraction rule—accommodates this new finding and does not lead to belief dilation in general.

To better demonstrate the contraction rule and to highlight its normative appeal, consider the following motivating example. A judge is contemplating the most suitable punishment for a suspect based on the information (signal) provided by the investigator. The suspect can either be innocent (d_I) or guilty (d_G), and the investigator can send

[1]See, for instance, Schmeidler (1989), Klibanoff, Marinacci, and Mukerji (2005), Maccheroni, Marinacci, and Rustichini (2006), Chew and Sagi (2008), Gul and Pesendorfer (2014), etc.

[2]The mechanism design literature has extensively explored the optimal design of mechanisms for both maxmin agents and principals. For example, Wolitzky (2016), Di Tillio, Kos, and Messner (2017), and Tang and Zhang (2021), among others, investigate mechanism design with maxmin agents. Chung and Ely (2007), Carroll (2019) and Brooks and Du (2021), among others, study the robust design of mechanisms when the principal faces uncertainty and adopts the maxmin criterion.

[3]FB is also known as the prior-by-prior rule, and ML is also known as the Dempster-Shafer rule.

[4]See Section 4.1 for more details.

either signal θ_1 or θ_2. The judge has ambiguity over the interpretations of the signals, resulting in two prior beliefs over $S = \{d_I, d_G\} \times \{\theta_1, \theta_2\}$, as shown in Table 1, where $\epsilon \in [0, 1/10)$ is a constant number.

	(d_I, θ_1)	(d_I, θ_2)	(d_G, θ_1)	(d_G, θ_2)
Prior 1	$1/10 + \epsilon$	$2/5 - \epsilon$	$2/5$	$1/10$
Prior 2	$2/5 + \epsilon$	$1/10 - \epsilon$	$1/10$	$2/5$

Table 1: Judge's Prior Distributions

According to the priors, the judge has an unambiguous belief that the suspect has probability 1/2 of being guilty. Upon receiving the signal, the judge updates her beliefs regarding the suspect's innocence and suggests a decision to the jury. We assume that the judge is an MEU maximizer, and her utility function is given by

$$U(y, d) = -y^2 + 2yI_{d=d_G},$$

where $d \in \{d_I, d_G\}$ denotes whether the suspect is indeed innocent or guilty, I is the indicator function, and $y \in [0, 1]$ denotes the severity of the punishment. According to the utility function, punishment incurs a cost of y^2, and the gain ($2yI_{d=d_G}$) from punishment is positive only when the suspect is indeed guilty. When the judge's belief is that the suspect has a probability of either $\underline{p}$ or $\bar{p}$ of being guilty ($\underline{p} \leq \bar{p}$), her optimal punishment y^* solves

$$\arg\max_{y\in[0,1]} \min_{p\in\{\underline{p},\bar{p}\}} \left(-y^2 + 2yp\right),$$

which leads to $y^* = \underline{p}$. In words, the judge tends to be lenient by suggesting the punishment according to the lowest probability for the suspect to be guilty.

We first consider the case where $\epsilon = 0$, and analyze the judge's posteriors when she uses FB or ML for belief updating. With FB, the judge updates every prior according to the realized signal following the Bayes' rule: When either θ_1 or θ_2 is observed, the judge updates her belief to that the suspect has a probability that ranges from 1/5 to 4/5 of being guilty. In this case, the judge's posterior beliefs over $\{d_I, d_G\}$ form a non-singleton set that contains her prior belief, which we refer to as *belief dilation*. Notably, without information, the judge would have suggested a punishment with severity 1/2, but given the information, she would suggest a less severe punishment with severity 1/5 *no matter* which signal she receives.

If the judge uses ML for belief updating, she only updates the priors that *maximize* the probability of the realized signal following the Bayes' rule. Since $\epsilon = 0$, the two priors both assign probability 1/2 to each one of the two signals, indicating that the two priors are both updated regardless which signal is realized. Thus, ML leads to the same predictions on the judge's decisions as FB does.

Nevertheless, the judge may find it hard to justify her recommended punishment to the jury if she updates with either FB or ML. This is because both signals allow for two symmetric interpretations that point in opposite directions, but the suggested punishment would always be less severe than the ex-ante optimal punishment, regardless of the signal that is realized. Another consequence of adopting the two rules is that the defense attorney of the suspect can decrease the punishment by providing additional information that enables symmetric but opposite interpretations. In conclusion, the excessive leniency implied by FB and ML in this example may appear impractical, particularly given that the maxmin criterion already proposes a lenient punishment.

In contrast to FB and ML, the contraction rule suggests a different punishment for the suspect. Following the contraction rule, the judge would use the highest probability of each tuple in $\{d_I, d_G\} \times \{\theta_1, \theta_2\}$ for belief updating. When θ_1 is realized, the maximal probabilities of (d_I, θ_1) and (d_G, θ_1) are both $2/5$. The judge's posterior belief is then given by the normalization of the two probabilities, leading her to maintain the prior belief over $\{d_I, d_G\}$. Similarly, the judge would maintain her prior belief when θ_2 occurs. Consequently, the judge would always suggest a punishment of severity $1/2$ to the jury by arguing that the new information is ambiguously uninformative and should be ignored.

While the contraction rule leads to the ignorance of the information when $\epsilon = 0$, it does not always predict so. To see this, assume that ϵ is positive. Compared with the benchmark case where $\epsilon = 0$, the joint probabilities of θ_1 and d_I increase, and those of θ_2 and d_I decrease. In this case, the contraction rule predicts that signal θ_1 would increase the judge's belief over d_I from $1/2$ to $(2 + 5\epsilon)/(4 + 5\epsilon)$, and signal θ_2 would decrease the judge's belief over d_I from $1/2$ to $(2 - 5\epsilon)/(4 - 5\epsilon)$.

Observe that in the example above, the contraction rule leads to the resolution of ambiguity. This observation relates to a crucial feature of the contraction rule: The degree of ambiguity on the realized event determines whether the information resolves the DM's ambiguity or not. More specifically, let S be the state space and P the DM's set of priors over S. An event, denoted by E, is a nonempty subset of S. The contraction measure, denoted by $\mu_P|E$, assigns each state in E a measure that equals its maximal ex-ante probability and each state outside E a measure of zero. That is, for every $s \in E$ and $s' \in S \backslash E$, $\mu_P|E(s) = \max_{p \in P} p(s)$ and $\mu_P|E(s') = 0$. The degree of ambiguity on E is then determined by the value of the summation $\sum_{s \in E} \mu_P|E(s)$. If 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, and the DM's posterior is the normalization of the contraction measure $\mu_P|E$. If 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 to form a posterior, leading to a non-singleton set of posteriors, which means that the DM's ambiguity is not resolved. In our motivating example, since there is no prior ambiguity over the innocence of the suspect, the ambiguity, which is only contained in the signals, has a small degree and is fully resolved after updating.

In Section 2.3, we compare the contraction rule with the two benchmark rules—FB and ML—through a stylized example in which the realized event has large ambiguity. The example demonstrates that the contraction rule moderates both FB and ML: The FB posterior set is highly sensitive to extreme priors that assign near-zero probabilities to the realized event, but the contraction posterior set is almost unaffected by such priors; ML completely disregards priors that do not maximize the probability of the realized event even though some of these priors are highly likely, while the contraction rule still updates these priors by assigning slightly less weights to them. We then discuss scenarios in which the contraction rule is more suitable to be applied than the rest two rules.

In Section 2.4, we show the divisibility of the contraction rule. That is, the contraction posterior set remains unchanged regardless of the order in which multiple pieces of information arrive. This property is normatively appeal as it suggests that the DM's updating process is unaffected by certain factors that are irrelevant with the content of the information. The divisibility of the contraction rule also enables us to obtain unique predictions on the DM's choices in dynamic settings in which the arrival order of multiple pieces of information is unobserved or partially observed.

In Section 3, we present our axiomatic framework. We consider a DM who is an MEU maximizer. The MEU representation allows us to uniquely identify the DM's set of beliefs. We consider a rich data set that documents the DM's preferences in different choice scenarios. An updating rule is then defined as a function that maps each ex-ante preference of the DM and each piece of information to an ex-post preference.

We characterize the contraction rule with six axioms. The Alignment Consistency and Sensitivity Congruence axioms are consistency conditions imposed on the DM's updating behavior when ex-ante ambiguity remains unresolved after updating, and the Sensitivity Independence axiom is the consistency condition when updating leads to ambiguity resolution. The axiom of Non-Ambiguity Persistence is the key departure of the contraction rule from the existing rules, which posits that information does not render an unambiguous prior over the payoff-relevant states ambiguous.[5] The axiom of Increased Sensitivity after Updating is our key behavioral axiom. The underlying behavioral postulate of this axiom is that the DM should become more sensitive to payoff differences on a given state after new information than before, provided that the new information does not rule out the state. While this postulate cannot be satisfied universally, the axiom of Increased Sensitivity after Updating requires the DM's updating behavior to satisfy this postulate to the largest possible extent. Together with the Continuity axiom, the aforementioned axioms fully characterize the contraction rule.

In Section 4, we provide several applications of the contraction rule. We first study the empirical relevance of the contraction rule. We show that a large proportion of the

[5]According to our motivating example, FB and ML violate this axiom.

experimental and empirical findings in SO23 and Liang (forthcoming) (henceforth L24) can be addressed by the rule. We then investigate the information design problem when ambiguous information is available. We show that by adopting the contraction rule for belief updating, an individual can design suitable information structure to manipulate her future choices to the largest extent if her future preference is not fully aligned with her current preference. Therefore, the contraction rule, combined with suitable information, is a valid approach for exercising self-control.

Our paper is part of the body of literature that examines belief updating in non-Bayesian frameworks (e.g., Epstein and Schneider (2007, 2008), Sadowski and Sarver (2021), Zhao (2022) and Ke, Wu, and Zhao (2023)). 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). Motivated by recent empirical findings, this paper contributes to the literature by providing a new updating rule that accommodates these findings. Similar to the updating rules proposed by Kovach (2023) and Cheng (2022) which nest FB and ML, the contraction rule also moderates FB and ML but does not result in belief dilation under generic conditions.

Our study is also related to the literature on dynamically consistent updating behavior, as surveyed by Gilboa and Marinacci (2013). 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, which enables the DM to devise an optimal contingent plan based on the information she might obtain and update her beliefs in a way that makes it optimal for her ex-post self 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, whereas the contraction rule violates dynamic consistency but satisfies consequentialism, as it updates each set of priors to some set of posteriors supported in the realized event and is independent of the choice problem faced by the DM.

Another stream of literature concerns the martingale property of updating ambiguous beliefs, with a distinctive paper in this line being Gul and Pesendorfer (2021) in which the proxy rule is introduced. The core axiom of the proxy rule is "not all news is bad news," meaning that given a prospect and a set of potentially realized signals, there exists one signal whose realization does not decrease the DM's evaluation of the prospect. Similarly, since the contraction rule does not lead to belief dilation in general, it implies that "a piece of news cannot be bad for all prospects." That is, the realization of a given signal has to weakly increase the DM's evaluation towards at least one uncertain prospect. Another feature shared by the two rules is that they both predict that information does not render the DM's unambiguous payoff-relevant belief ambiguous. While the proxy rule is used for updating maxmin preferences that allow for totally monotone capacities, the contraction rule can be used for updating arbitrary maxmin preferences.[6]

Several 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 (2022), Epstein and Halevy (2024), among others. Our model provides consistent predictions with several findings that cannot be accommodated by FB and ML. Specifically, we show that the contraction rule does not render the DM's unambiguous payoff-relevant belief ambiguous (SO23) and predicts under-reaction to ambiguous information (L24). We also demonstrate in Sections 4.1 and 4.2 that our model can accommodate some other findings by SO23 and L24.

Our paper is also related to the literature that examines how ambiguous information impacts decision-making in interactive settings involving multiple players. This literature includes works 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), and Chen (2023), etc. Since the contraction rule does not lead to belief dilation in general, it can be considered as an alternative approach for the applications that involve updating ambiguity to disentangle the influence of belief dilation from the effects of other features of ambiguity. In Section 4.3, we provide one such application by exploring the optimal ambiguous information design with the contraction rule. This application demonstrates how the contraction rule can provide new insights to the literature on information design and, more broadly, the literature on interactive decision-making.

The rest of the paper is organized as follows. In Section 2.1, we set out the basic definitions used in the paper. We formally introduce the contraction rule in Section 2.2 and discuss its properties in Sections 2.3 and 2.4. We characterize the contraction rule in Section 3. Section 4 contains empirical evidence and applications of the contraction rule. All omitted proofs can be found in the Appendix.

2 Model

2.1 Preliminary

State space and measures. Let S be a state space that has *infinitely* many states. A measure π over S is said to have a finite support if there exists finite $\hat{S} \subseteq S$ such that $\pi(S\backslash\hat{S}) = 0$. Denote by $\mathcal{M}(S)$ the set of all measures over S with finite supports. Consider an arbitrary nonempty subset $\mathcal{M}$ of $\mathcal{M}(S)$. The set $\mathcal{M}$ is said to be finitely

[6] In the Online Appendix A of SO23, 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.

supported if there is finite $\hat{S} \subseteq S$ such that for every $\pi \in \mathcal{M}$, $\pi(S \backslash \hat{S}) = 0$. If $\mathcal{M}$ is finitely supported and $\max_{s \in S} (\sup_{\pi \in \mathcal{M}} \pi(\{s\})) < +\infty$, then define $\mu_{\mathcal{M}}$ as the measure in $\mathcal{M}(S)$ such that for every $s \in S$, $\mu_{\mathcal{M}}(\{s\}) = \sup_{\pi \in \mathcal{M}} \pi(\{s\})$. For all $\hat{S} \subseteq S$ and $\pi \in \mathcal{M}(S)$, let $\pi|\hat{S}$ be the measure in $\mathcal{M}(S)$ that satisfies $(\pi|\hat{S})(E) = \pi(E \cap \hat{S})$ for all $E \subseteq S$, and 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 $\pi \in \mathcal{M}(S)$, let π_Π be the measure induced by π over the algebra generated by Π such that $\pi_\Pi(S_i) = \pi(S_i)$ for all $i \in \{1, ..., n\}$. Correspondingly, define $\mathcal{M}_\Pi = \{\pi_\Pi : \pi \in \mathcal{M}\}$.

For all $\pi, \pi' \in \mathcal{M}(S)$ and $\alpha, \beta \in \mathbb{R}$, if $\alpha\pi(\{s\}) + \beta\pi'(\{s\}) \geq 0$ for every $s \in S$, then let $\alpha\pi + \beta\pi'$ be the measure in $\mathcal{M}(S)$ such that for every $E \subseteq S$, $(\alpha\pi + \beta\pi')(E) = \alpha\pi(E) + \beta\pi'(E)$. For any nonempty $\mathcal{M}, \mathcal{M}' \subseteq \mathcal{M}(S)$, let $\alpha\mathcal{M} + \beta\mathcal{M}' = \{\alpha\pi + \beta\pi' : \pi \in \mathcal{M}, \pi' \in \mathcal{M}'\}$ if each $\alpha\pi + \beta\pi'$ is a well-defined measure in $\mathcal{M}(S)$. For any nonempty $\mathcal{M} \subseteq \mathcal{M}(S)$, let $co(\mathcal{M})$ be the convex hull of $\mathcal{M}$, i.e., $\pi \in co(\mathcal{M})$ if and only if there exist $\{\pi_i\}_{i=1}^n \subseteq \mathcal{M}$ and $\{\alpha_i\}_{i=1}^n \subseteq \mathbb{R}_{++}$ with $n \in \mathbb{N}_+$ such that $\sum_{i=1}^n \alpha_i = 1$ and $\pi = \sum_{i=1}^n \alpha_i\pi_i$.

Probability measures and events. Denote by $\Delta(S) \subseteq \mathcal{M}(S)$ the set of all finitely supported probability measures over S. For every $\pi \in \mathcal{M}(S)$ with $\pi(S) > 0$, denote by $\overline{\pi}$ the normalized probability measure of π such that for every $E \subseteq S$, $\overline{\pi}(E) = \frac{\pi(E)}{\pi(S)}$. 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 for all $p, q \in P$ and $\alpha \in [0, 1]$, $\alpha p + (1 - \alpha)q \in P$, and closed if it is a closed subset of $\mathbb{R}^S$ when each $p \in P$ is viewed as a vector in $\mathbb{R}^S$. Let $\mathscr{P}$ be the collection of nonempty, finitely supported, convex and closed sets of probability measures over S.

An event is a nonempty and finite subset of S. Let $\mathcal{S}$ be the collection of all events. For all $P \in \mathscr{P}$ and $E \in \mathcal{S}$, E is P-non-null if there exists $p \in P$ such that $p(E) > 0$; otherwise, we say that E is P-null. Denote by $\mathcal{S}_P$ the set of all P-non-null events.

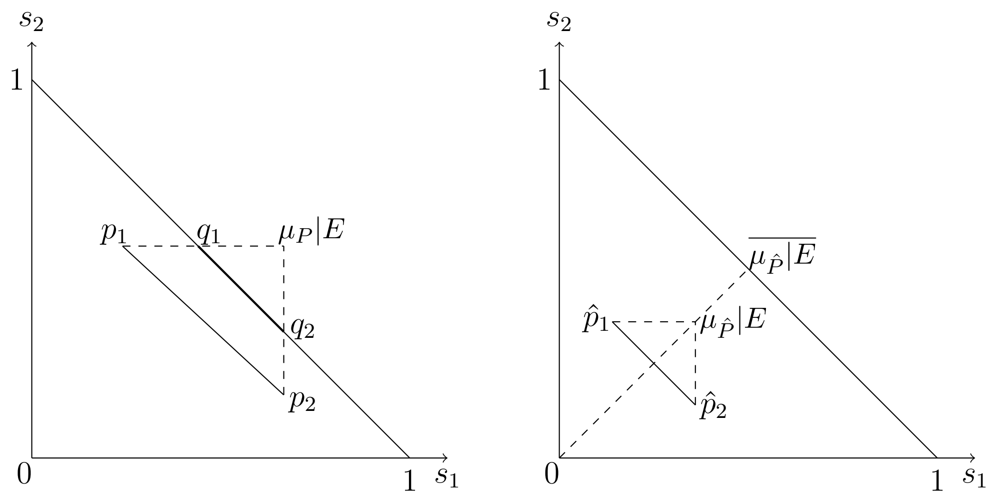
Acts and evaluations. Let X be a nonempty and convex set of consequences. An act is a function $f : S \to X$ that maps each state to some consequence. We also use each x in X to denote the constant act that maps each state to x. Let $\mathcal{F}$ denote the set of all acts. For all $f \in \mathcal{F}$, $p \in \Delta(S)$, and function $u : X \to \mathbb{R}$, define $u(f; p) = \sum_{s \in S} p(s)u(f(s))$, and for every nonempty $P \subseteq \Delta(S)$, define $u^{\downarrow}(f; P) = \inf_{p \in P} u(f; p)$. For any acts f and g and $E \subseteq S$, we write $f \overset{E}{=} g$ if f and g agree on E, and denote by fEg the act that equals f on E and equals g on $S \backslash E$. For any given $\alpha \in [0, 1]$, let $\alpha f + (1 - \alpha)g$ be the mixture act of f and g such that for all $s \in S$, $(\alpha f + (1 - \alpha)g)(s) = \alpha f(s) + (1 - \alpha)g(s)$. Throughout the paper, we write s for $\{s\}$ whenever there is no confusion.

2.2 Contraction Rule

In this section, we introduce the contraction rule. For every $P \in \mathscr{P}$ and P-non-null event E, the contraction posterior set $Q^c(P, E)$ is defined as follows:

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

Recall that $\mu_P|E$ is defined such that for all $s \in E$, $\mu_P|E(s) = \max_{p \in P} p(s)$, and for all $s \notin E$, $\mu_P|E(s) = 0$. This measure is referred to as the *contraction measure.* With the contraction rule, the DM updates every prior *towards* the contraction measure. When $\mu_P(E) > 1$, $\Phi(p|E, \mu_P|E)$ is the unique probability measure between $p|E$ and $\mu_P|E$. In this case, the posterior set is formed by projecting $P|E$ onto the set of probability measures over E in 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. Figure 1 illustrates these two cases.



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

Figure 1: For both (a) and (b), all priors have support $\{s_1, s_2, s_3\}$. We depict each prior in the 2-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 $\{\overline{\mu_{\hat{P}}|E}\}$.

We interpret the value of $\mu_P(E)$ as an indicator of the degree of ambiguity on E.[7] When $\mu_P(E) \leq 1$, E has small or no ambiguity, and the realization of E renders the DM's belief unambiguous. When $\mu_P(E) > 1$, E has large ambiguity, and the DM's ambiguity remains unresolved after E occurs. In the latter case, the contraction measure is maintained after updating, as stated by the following proposition.

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

Proposition 1. *For all $P \in \mathscr{P}$ and $E \in \mathcal{S}_P$, $|Q^c(P,E)| = 1$ 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 would like to emphasize that setting $\mu_P(E) = 1$ as the cutoff for ambiguity resolution is not *ad hoc*. In fact, it can be implied by two postulates of belief updating, which we explain in detail below.

Postulate 1 states that the DM should rely solely on prior distributions over the realized event E to update her beliefs. Consequently, the DM's set of posteriors is a function of $P|E$.[8]

To state Postulate 2, we 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 states that if the DM's prior belief set P induces a unique marginal distribution on D, then the DM's ex-post belief set over D should also be a singleton, regardless of which signal in Θ is realized. According to Postulate 2, if the DM has an unambiguous belief over payoff-relevant states, then new information should not render her belief ambiguous. The following example illustrates why the two postulates imply the cutoff for ambiguity resolution to be $\mu_P(E) = 1$.

Example 1. Consider two prior belief sets $P_1 = co(\{p_1, \hat{p}_1\})$ and $P_2 = co(\{p_2, \hat{p}_2\})$ with support $\{d_1, d_2\} \times \{\theta_1, \theta_2\}$. The distributions of the priors are presented in Table 2. The realized event is $E = \{(d_1, \theta_1), (d_2, \theta_1)\}$, i.e., signal θ_1. $\square$

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

In Example 1, since $p_1|E = p_2|E$ and $\hat{p}_1|E = \hat{p}_2|E$, we have $P_1|E = P_2|E$. Postulate 1 then implies that the two prior belief sets should be updated to the same set of posteriors when E occurs. As both p_2 and $\hat{p}_2$ assign a probability of 0.4 to $\{d_1\} \times \Theta$ and 0.6 to $\{d_2\} \times \Theta$, P_2 is unambiguous on $\{d_1, d_2\}$. Postulate 2 then implies that P_2 is updated to a singleton posterior set given θ_1, and so is P_1.

Based on the above analysis, for any given prior belief set P_1, if we can construct another belief set P_2 that is unambiguous on $\{d_1, d_2\}$ and satisfies $P_2|E = P_1|E$, then by the two postulates, P_1 should be updated to a singleton belief set after the realization of θ_1, i.e., the ambiguity is fully resolved. It turns out that the condition $\mu_{P_1}(E) \leq 1$ is

[8] It is worth noting that both FB and ML satisfy Postulate 1. Formal definitions of FB and ML can be found in Section 2.3.

both sufficient and necessary for us to construct such a belief set P_2.[9] Therefore, the two postulates jointly justify the cutoff for ambiguity resolution.

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

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

Two observations based on equation (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 captures how the DM incorporates the likelihoods of the priors into the updating process: If a prior assigns a higher probability to E, then it is considered more likely given the information and thus weighed more. In particular, if $p(E) = 1$, then $\Phi(p|E, \mu_P|E) = p|E$. Hence, a prior is entirely 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 more ambiguous and thus less informative. As a result, the DM relies more on her priors.

Our next proposition demonstrates that the contraction posterior sets are well-behaved: The set $Q^c(P,E)$ is always nonempty, convex and closed; if the new information is uninformative, then the set of posteriors is the same as the set of priors.

Proposition 2. *For all $P \in \mathscr{P}$ and $E \in \mathcal{S}_P$, $Q^c(P,E) \in \mathscr{P}$; if in addition $p(E) = 1$ for all $p \in P$, then $Q^c(P,E) = P$.*

2.3 Applicability of the Contraction Rule

In this section, we compare the contraction rule with FB and ML to discuss situations in which the contraction rule is more suitable to be applied. To start with, we formally introduce FB and ML.

For all $P \in \mathscr{P}$ 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. With FB, the DM updates each prior to its posterior following the Bayes' rule.

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

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

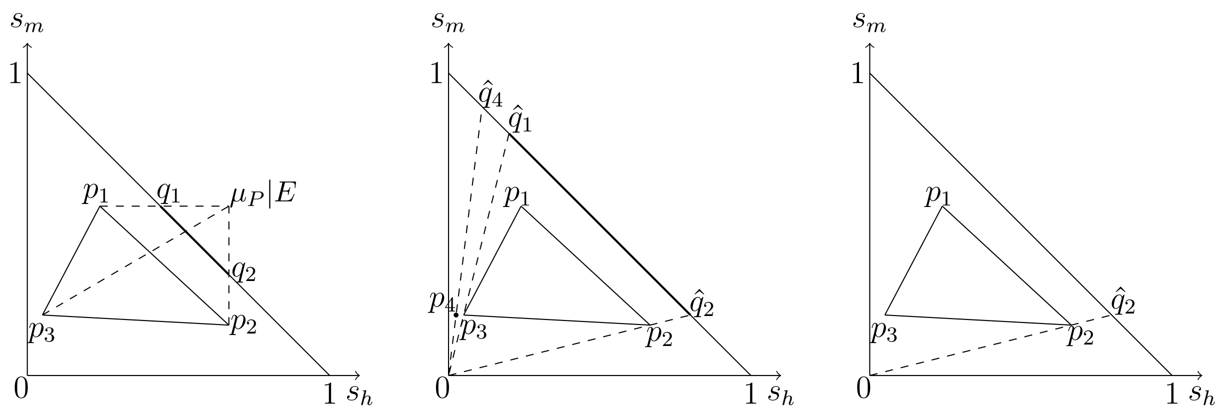
[9] To see the necessity, note that if $\mu_{P_1}(E) > 1$, then the existence of such a belief set P_2 implies that for all $p \in P_2$, $1 = p(\{d_1\} \times \{\theta_1, \theta_2\}) + p(\{d_2\} \times \{\theta_1, \theta_2\}) \geq \mu_{P_2}(E) = \mu_{P_1}(E) > 1$, which is a contradiction. For the sufficiency, note that we can shift probabilities outside of the realized event E 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 $\hat{p}_2$.

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

Example 2. An employer (she) is considering hiring a job seeker (he) whose true ability can be high (s_h), medium (s_m), or low (s_l). There are three possible types of job seekers, where the type is the job seeker's private information and correlates with his true ability. Specifically, for every $t \in \{1, 2, 3\}$, a type-t job seeker's ability is distributed according to $p_t \in \Delta(\{s_h, s_m, s_l\})$, where:

$$p_1(s_h) = 6/25,\ p_1(s_m) = 14/25,\ p_1(s_l) = 1/5,$$
$$p_2(s_h) = 2/3,\ p_2(s_m) = 1/6,\ p_2(s_l) = 1/6,$$
$$p_3(s_h) = 1/20,\ p_3(s_m) = 1/5,\ p_3(s_l) = 3/4.$$

The employer is uncertain about the prior distribution of the three types, and as a result, her prior beliefs over the job seeker's ability are given by $P = co(\{p_1, p_2, p_3\})$. After interviewing the job seeker, the employer finds the ability of the job seeker to be at least of medium level, i.e., the information obtained by the employer through the interview is $E = \{s_h, s_m\}$. If the employer updates with the contraction rule, FB, or ML, then her sets of posteriors are given respectively by $Q^c(P, E) = co(\{q_1, q_2\})$, $Q^{fb}(P, E) = co(\{\hat{q}_1, \hat{q}_2\})$, or $Q^{ml}(P, E) = \{\hat{q}_2\}$, where $q_1(s_h) = 1 - q_1(s_m) = 11/25, q_2(s_h) = 1 - q_2(s_m) = 2/3$, $\hat{q}_1(s_h) = 1 - \hat{q}_1(s_m) = 1/5, \hat{q}_2(s_h) = 1 - \hat{q}_2(s_m) = 4/5$. We depict them in Figure 2. □



(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.

If the employer updates her beliefs with FB, she updates every prior to some posterior *regardless of* the probability it assigns to the realized event. Thus, one consequence of adopting FB is that the employer's posterior belief set can be sensitive to priors that are less likely. In Example 2, p_3 assigns a low probability to event E, meaning that p_3 is unlikely to be the true prior given the information. Despite this, its Bayes' posterior $\hat{q}_1$—according to which the job seeker is very unlikely to have high ability—is maintained in the posterior belief set. This posterior, which is significantly different from the Bayes' posterior of p_2 and more extreme than that of p_1, may affect the employer's final decision.[10] Furthermore, priors that assign lower probabilities to the realized event are amplified more by the Bayes' rule, and thus changes among those priors can result in more significant changes of the FB posterior set. To see this, suppose that p_3 shifts slightly to p_4 as depicted in Figure 2(b). Then the prior set changes from $co(\{p_1, p_2, p_3\})$ to $co(\{p_1, p_2, p_4\})$, resulting in the FB posterior set being enlarged from $co(\{\hat{q}_1, \hat{q}_2\})$ to $co(\{\hat{q}_4, \hat{q}_2\})$. Note that p_4 is even less likely than p_3 given E, but it leads to a more extreme posterior $\hat{q}_4$ that may affect the employer's decision even more.

If the employer updates her beliefs with ML, her posterior set becomes the singleton set $\{\hat{q}_2\}$, which is shown in Figure 2(c). Unlike FB, ML only updates priors that are most consistent with the information. Consequently, ML may rule out some reasonable priors. In Example 2, prior p_1 only assigns slightly less probability to E than p_2 does, but is disregarded by ML when E occurs. Since the Bayes' posterior of p_2 yields a high probability for the job seeker to have high ability, the employer will become over-optimistic about the ability of job seeker if she updates with ML.

The contraction rule moderates FB and ML in the sense that it updates all priors but puts less weights to those that are less consistent with the information. As shown in Figure 2(a), with the contraction rule, the employer's posterior set $co(\{q_1, q_2\})$ is determined by the more likely priors p_1 and p_2, and is unaffected by the less likely prior p_3.

Based on Example 2, it appears that the contraction rule, as a less extreme rule compared to FB and ML, is a more appropriate choice for situations where the DM wants to consider all priors but also aims to minimize the impact of unlikely ones. Relevant applications include the belief updating of judges, physicians, managers, etc. In those scenarios, DMs do need to take into account unlikely priors but also need to discount them properly either for fairness (judges) or to avoid over- or under-treatments (physicians).

The contraction rule is also useful in scenarios where there is large ambiguity on the payoff-relevant states, but the available information is relatively informative. Formally, let the state space be given by $S = D \times \Theta$, where D contains all payoff-relevant states, and Θ contains all signals. Suppose that the DM is fully uncertain about the prior distributions over D but is confident that the conditional distribution of Θ on D is given by χ.[11] For simplicity, assume that for all $\theta \in \Theta$ and $d \in D$, $\chi(\theta|d) > 0$. With FB, any signal

[10] For instance, the employer may evaluate the job seeker's ability according to the minimal probability for him to have high ability.

[11] That is, for each $d \in D$, $\chi(\cdot|d)$ is a distribution over Θ

realization does not change the the DM's prior beliefs over D, i.e., she remains to be fully uncertain about the payoff-relevant states. With ML, if $\theta \in \Theta$ is observed, the DM will rule out states in D that do not maximize $\chi(\theta|d)$ all of a sudden, regardless of how close the conditional probabilities of θ on those states are to the maximal one. By contrast, the contraction rule suggests that the DM maintains all states in D and updates her beliefs over D proportionally according to $(\chi(\theta|d))_{d\in D}$.

The contraction rule can also be useful when the ambiguity mainly comes from the new information, i.e., the DM has small ambiguity about the prior distribution of the payoff-relevant states in D, but there is a significant level of ambiguity regarding how states in D are correlated with signals in Θ. As we have illustrated in the Introduction, the contraction rule leads to more justifiable posterior beliefs in such circumstances. In Section 4, we further show that the predictions of the contraction rule are largely consistent with the experimental evidence in these scenarios.

Finally, we highlight one scenario in which the contraction rule may not be the preferred rule for belief updating. Consider an urn containing a total of 90 balls, with 30 red balls and 60 balls with an unknown composition of blue and white balls. Let a random ball be drawn from the urn, and assume that the DM believes that the ball has a probability of $\frac{1}{3}$ of being red and a probability ranging from 0 to $\frac{2}{3}$ of being blue (white). In such a situation, if the DM is informed that the ball is not red, it appears more likely that she will update her beliefs using FB instead of the contraction rule. That is, she will revise her beliefs to that the ball has a probability ranging from 0 to 1 of being blue (white). In this example, the source of ambiguity—the unknown composition of the blue and while balls—is evident, and thus it seems to be more natural for the DM to backtrack the source of ambiguity and revise her beliefs accordingly.

2.4 Divisibility

When a DM has multiple pieces of information, the order in which the information arrives may affect the DM's final beliefs. We say that an updating rule is divisible, or path-independent, if the posterior set of the DM is unaffected by the arrival order of multiple pieces of information. A divisible rule delivers unique predictions on the DM's final beliefs regardless of her information acquisition order. When there is no ambiguity, Cripps (2019) provides a characterization for divisible rules. The next proposition demonstrates that the contraction rule also satisfies divisibility.

Proposition 3. *For all $P \in \mathscr{P}$ and $E, F \in \mathcal{S}_P$ with $F \subseteq E$, $Q^c(Q^c(P,E),F) = Q^c(P,F)$.*

The key observation that leads to Proposition 3 is that the contraction measure remains the same if ambiguity remains unresolved, i.e., $\mu_{Q^c(P,E)}|F = \mu_P|F$ (which is implied by Proposition 1). Recall that with the contraction rule, the DM updates her priors towards the contraction measure. Therefore, as long as the contraction measure on F remains unchanged with the arrival order of information, the DM's final posterior set on F also remains invariant. Essentially, one can show that FB is also a divisible rule by the same argument: The origin—the measure that assigns each state a measure of zero—serves as a contraction point such that each prior is updated *away* from it following the Bayes' rule. Apparently, the contraction point of FB—i.e., the origin—remains unchanged regardless of the arrival order of multiple pieces of information.

3 Axiomatic Foundation

In this section, we present the MEU model introduced by GS. The model provides a unique identification of the DM's belief set. We then axiomatize the contraction rule based on the MEU framework.

3.1 Maxmin Expected Utility

A preference of the DM is a binary relation $\succsim$ over the set of acts $\mathcal{F}$. We say that a preference $\succsim$ admits a simple MEU representation if there exists $P \in \mathscr{P}$ and a non-constant and affine utility function $u : X \to \mathbb{R}$ such that for all $f, g \in \mathcal{F}$, $f \succsim g$ if and only if $u^{\downarrow}(f; P) \geq u^{\downarrow}(g; P)$.[12] We refer to such a preference as a simple maxmin preference and say that it is represented by (u, P). Denote by $\mathscr{R}^*$ the set of all simple maxmin preferences. The following axioms characterize $\mathscr{R}^*$.

Axiom M1—Weak Order: The preference $\succsim$ is complete and transitive.

Axiom M2—Certainty-Independence: For all $f, g \in \mathcal{F}$, $x \in X$, and $\alpha \in (0, 1)$, $f \succ g$ if and only if $\alpha f + (1 - \alpha)x \succ \alpha g + (1 - \alpha)x$.

Axiom M3—Weak Continuity: For all $f, g, h \in \mathcal{F}$, if $f \succ g$ and $g \succ h$, then there exist α and β such that $\alpha f + (1 - \alpha)h \succ g$ and $g \succ \beta f + (1 - \beta)h$.

Axiom M4—Monotonicity: For all $f, g \in \mathcal{F}$, if $f \succ g$, then there exists $s \in S$ such that $f(s) \succ g(s)$.

Axiom M5—Uncertainty Aversion: For all $f, g \in \mathcal{F}$ and $\alpha \in (0, 1)$, $f \sim g$ implies $\alpha f + (1 - \alpha)g \succsim f$.

Axiom M6—Non-degeneracy: There exist $f, g \in \mathcal{F}$ such that $f \succ g$.

Axiom M7—Finiteness: There exists a nonempty and finite subset $\hat{S} \subseteq S$ such that for all $f, g \in \mathcal{F}$, if for all $s \in \hat{S}$, $f(s) = g(s)$, then $f \sim g$.

Axioms M1-M6 are introduced by GS to characterize the set of maxmin preferences. The additional axiom, Axiom M7, ensures that the set of priors P in the representation is

[12] A function u is affine if for all $x, y \in X$ and $\alpha \in (0, 1)$, $u(\alpha x + (1 - \alpha)y) = \alpha u(x) + (1 - \alpha)u(y)$.

finitely supported. Together, Axioms M1-M7 are sufficient and necessary for a preference to be a simple maxmin preference. The MEU representation allows a unique identification of the DM's set of beliefs: If a preference $\succsim$ is represented by both (u, P) and $(\hat{u}, \hat{P})$, then $P = \hat{P}$, and u is an affine transformation of $\hat{u}$.[13]

3.2 Updating rules

Collection of preferences. An updating rule specifies how a DM's preference changes with new information in *all possible choice scenarios*. To define updating rules, we consider a collection of preferences $\mathscr{R} \subseteq \mathscr{R}^*$ that satisfy the following two properties:

(1) for all $\succsim^1, \succsim^2 \in \mathscr{R}$ and $x, y \in X$, $x \succsim^1 y$ if and only if $x \succsim^2 y$, and
(2) for all $\bar{\mathscr{R}} \subseteq \mathscr{R}^*$ such that $\mathscr{R} \subsetneq \bar{\mathscr{R}}$, the set $\bar{\mathscr{R}}$ violates condition (1).

We interpret $\mathscr{R}$ as a data set that includes the DM's preferences in various choice scenarios, and interpret the state space S as an abstract space used for storing those preferences. To understand this interpretation, consider a choice scenario $(H, \trianglerighteq)$, where the nonempty and finite set H is the *actual* state space, and $\trianglerighteq$ is the DM's *actual* maxmin preference over the set of *actual* acts X^H. We can transfer the preference $\trianglerighteq$ to elements in $\mathscr{R}$ as follows. For all injection $\gamma : H \to S$, we can define $\succsim_\gamma \in \mathscr{R}$ such that for all $f, g \in \mathcal{F}$, $f \succsim_\gamma g$ if and only if $f \circ \gamma \trianglerighteq g \circ \gamma$.[14] Following this construction procedure, we can consider many choice scenarios $\{(H_i, \trianglerighteq^i)\}_{i \in I}$ of the DM, and transfer each one of them to elements in $\mathscr{R}$.

With the above interpretation in mind, property (1) states that the DM's preference over the set of consequences X remains the same in different choice scenarios, and the variation of the DM's preferences across different scenarios is driven by the change of her beliefs. Property (2) is a richness condition, which states that our data set contains all simple maxmin preferences that induce the same preference over X. Essentially, this means that for every $P \in \mathscr{P}$, we can find some choice scenario in which the DM's set of beliefs is given by P (under some properly defined injection γ).

New information and updating rules. The new information takes the form of events that happen with non-zero probabilities.[15] Those events are referred to as non-null events. Specifically, for a given simple maxmin preference $\succsim$, a set $E \subseteq S$ is said to be $\succsim$-null if for all $f, g \in \mathcal{F}$ such that $f \overset{S \setminus E}{=} g$, we have $f \sim g$; otherwise, E is said to be $\succsim$-non-null.

[13] The function u is an affine transformation of $\hat{u}$ if there exist real numbers $\lambda > 0$ and β such that for all $x \in X$, $u(x) = \lambda \hat{u}(x) + \beta$.

[14] Note that $f \circ \gamma$ (or $g \circ \gamma$) is an element of X^H satisfying that for all $h \in H$, $f \circ \gamma(h) = f(\gamma(h))$. Since $\trianglerighteq$ allows for a maxmin representation and H is finite, the induced preference $\succsim_\gamma$ is a simple maxmin preference.

[15] 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).

Clearly, if $\succsim$ is represented by (u, P), then E is $\succsim$-null if and only if E is P-null. We denote by $\mathcal{S}(\succsim)$ the set of all $\succsim$-non-null events, and define the set $\mathcal{G} \subseteq \mathscr{R} \times \mathcal{S}$ such that $(\succsim, E) \in \mathcal{G}$ if and only if $E \in \mathcal{S}(\succsim)$.

Definition 1. *An updating rule (over preferences) is a function $\Gamma : \mathcal{G} \to \mathscr{R}$ such that for all $(\succsim, E) \in \mathcal{G}$,* (i) *$S \backslash E$ is $\Gamma(\succsim, E)$-null, and* (ii) *if $S \backslash E$ is $\succsim$-null, then $\Gamma(\succsim, E) = \succsim$.*

We denote $\Gamma(\succsim, E)$ by $\succsim_{E,\Gamma}$, and by $\succsim_E$ whenever the updating rule is clearly specified. An updating rule maps the DM's ex-ante preference $\succsim$ and information E to her ex-post preference $\succsim_E$. Condition (i) states that given event E, the DM will ignore states outside of E. Condition (ii) states that the DM does not change her preference if the new information is not informative.

Definition 2. *An updating rule Γ is the contraction rule (respectively FB and ML) if for all $(\succsim, E) \in \mathcal{G}$ with $\succsim$ being represented by (u, P), $\succsim_E$ can be represented by $(u, Q^c(P, E))$ (respectively $(u, Q^{fb}(P, E))$ and $(u, Q^{ml}(P, E))$).*

An implicit assumption on updating rules. Definition 1 entails an implicit assumption on the DM's updating. Recall that we interpret S as an abstract space for storing the DM's actual preferences in different choice scenarios. Consider two choice scenarios $(H_1, \trianglerighteq^1)$ and $(H_2, \trianglerighteq^2)$, where both H_1 and H_2 are nonempty and finite, and both $\trianglerighteq^1$ and $\trianglerighteq^2$ allow for maxmin representations. The two choice scenarios are said to be isomorphic if there is a bijection $\phi : H_1 \to H_2$ such that for all $f, g \in X^{H_2}$, $f \trianglerighteq^2 g$ if and only if $f \circ \phi \trianglerighteq^1 g \circ \phi$. In such a case, the two actual preferences $\trianglerighteq^1$ and $\trianglerighteq^2$ are said to be ϕ-identical. Note that two isomorphic choice scenarios can induce the same preference $\succsim$ in $\mathscr{R}$ through some properly defined injections $\gamma_1 : H_1 \to S$ and $\gamma_2 : H_2 \to S$.[16] However, irrespective of whether $\succsim$ is induced by $\trianglerighteq^1$ or $\trianglerighteq^2$, the given updating rule always updates $\succsim$ to $\succsim_E$ when event E occurs. Hence, the implicit assumption here is that identical preferences are updated to identical ex-post preferences under identical information. More specifically, consider a $\succsim$-non-null event E. For each $i \in \{1, 2\}$, define $E_i = \gamma_i^{-1}(E) \subseteq H_i$, and let $\trianglerighteq^i_{E_i}$ be the DM's ex-post preference in choice scenario $(H_i, \trianglerighteq^i)$ after E_i is realized. The implicit assumption essentially requires that if $\trianglerighteq^1$ and $\trianglerighteq^2$ are ϕ-identical, then $\trianglerighteq^1_{E_1}$ and $\trianglerighteq^2_{E_2}$ must also be ϕ-identical, ensuring that they can induce the same ex-post preference $\succsim_E$ in $\mathscr{R}$ through the injections γ_1 and γ_2, respectively.

3.3 Characterization

In this section, we characterize the contraction rule. Throughout the section, all preferences are assumed to be in $\mathscr{R}$, and thus they share the same ranking for constant acts.

[16]The two injections γ_1 and γ_2 should satisfy $\gamma_1 = \gamma_2 \circ \phi$ and $\gamma_2 = \gamma_1 \circ \phi^{-1}$, where $\phi^{-1} : H_2 \to H_1$ is the inverse of ϕ.

Definition 3. *For all* $\succsim \in \mathscr{R}$ *and partition* $\Pi = \{S_i\}_{i=1}^n$ *of* S, $\succsim$ *is* Π*-unambiguous if for all* $\alpha \in (0,1)$ *and* $f, g \in \mathcal{F}$ *that are measurable with respect to* Π, $f \sim g$ *implies* $\alpha f + (1-\alpha) g \sim f$*; otherwise,* $\succsim$ *is* Π*-ambiguous. The preference* $\succsim$ *is unambiguous if for all partition* Π' *of* S, $\succsim$ *is* Π'*-unambiguous; otherwise,* $\succsim$ *is ambiguous.*

Definition 3 divides all preferences into two categories: unambiguous and ambiguous. A preference is considered unambiguous if it is not in favor of or against randomization. For any given maxmin preference, it is unambiguous if and only if it can be represented by some (u, P) such that P is a singleton set.

Definition 4. *For all* $\{\succsim^k\}_{k=1}^3 \subseteq \mathscr{R}$ *and* $E \in \cap_{i=1}^3 \mathcal{S}(\succsim^i)$, $\succsim^3$ *is* E*-aligned with* $(\succsim^1, \succsim^2)$ *if for all* $x, y \in X$ *and* $f \in \mathcal{F}$, $fEx \succsim^1 y$ *and* $fEx \succsim^2 y$ *imply* $fEx \succsim^3 y$*, and* $y \succsim^1 fEx$ *and* $y \succsim^2 fEx$ *imply* $y \succsim^3 fEx$*.*

Definition 4 introduces a new alignment relation that only involves acts taking constant values on the complement of E. The restriction of a given preference on those acts can be regarded as the *preference on* E, since the details of the preference outside of E are ignored. By Definition 4, a given preference is E-aligned with two other preferences if, for every act fEx and every constant act y, the given preference ranks the two acts consistently with the other two preferences whenever the latter two share the same ranking for the two acts. Lemma 3 in the Appendix characterizes this E-alignment relation.

The E-alignment of $\succsim^3$ with $(\succsim^1, \succsim^2)$ also means that $\succsim^3$ lies between the latter two preferences on E. More specifically, it means that the evaluation of any act fEx under $\succsim^3$ is always between its evaluations under $\succsim^1$ and $\succsim^2$. To see this, consider act fEx and three constant acts y_1, y_2, and y_3 that are equally good as fEx under $\succsim^1, \succsim^2$, and $\succsim^3$, respectively. These constant acts can be viewed as the evaluations of fEx under the three preferences. Without loss of generality (WLOG), assume that y_1 is better than y_2. Since both $\succsim^1$ and $\succsim^2$ rank fEx weakly below y_1 and weakly above y_2, $\succsim^3$ should also do so, indicating that y_3—which is as good as fEx under $\succsim^3$—is ranked between y_1 and y_2.

Axiom 1—Alignment Consistency: For all $\{\succsim^k\}_{k=1}^3 \subseteq \mathscr{R}$ and $E \in \cap_{k=1}^3 \mathcal{S}(\succsim^k)$, if $\succsim^3$ is E-aligned with $(\succsim^1, \succsim^2)$, $\succsim^1_E = \succsim^2_E$, and $\succsim^1_E$ and $\succsim^2_E$ are ambiguous, then $\succsim^1_E = \succsim^2_E = \succsim^3_E$.

Consider three choice scenarios such for every $k \in \{1, 2, 3\}$, the DM's preference in scenario k is given by $\succsim^k$. Assume that the DM's preference on E in scenario 3 falls between her preferences on E in scenarios 1 and 2. Axiom 1 applies to this situation and states that if the DM updates her preferences in scenarios 1 and 2 to the same ex-post preference after E occurs, then she should also do so in scenario 1, provided that ambiguity is unresolved after updating. Intuitively, since the difference between $\succsim^1$ and $\succsim^2$ on E disappears after E occurs, the difference between $\succsim^1$ and $\succsim^3$ on E, which is even smaller, should also disappear after E occurs. Therefore, Axiom 1 posits that the more similar two ex-ante preferences are, the more similar their ex-post preferences should be.

Notably, Axiom 1 requires $\succsim_E^1$ and $\succsim_E^2$ to be ambiguous. This means that the consistency condition, which stipulates that more similar ex-ante preferences are updated to more similar ex-post preferences, only applies to scenarios where ambiguity remains unresolved after updating. Intuitively, similar ex-ante preferences entail similar informational contents on the realized event, indicating that the consistency condition is naturally fulfilled when all available information is used for belief updating and can be violated when only partial information is utilized. Therefore, Axiom 1 suggests that the DM should utilize all available information for updating whenever she chooses not to disregard the ambiguity. This is precisely the case with the contraction rule: When the degree of ambiguity on the realized event is large, the DM utilizes all available information by updating *every* prior towards the contraction measure to form its corresponding posterior, resulting in a non-singleton set of posteriors.

The aforementioned discussion also indicates that FB generally satisfies the consistency condition in Axiom 1. This is because with FB, the DM utilizes all information by updating every prior to a posterior following the Bayes' rule. Indeed, FB satisfies the following stronger version of Axiom 1, which states that the consistency condition always holds.[17]

Axiom 1*—Alignment Consistency*: For all $\{\succsim^k\}_{k=1}^3 \subseteq \mathscr{R}$ and $E \in \cap_{k=1}^3 \mathcal{S}(\succsim^k)$, if $\succsim^3$ is E-aligned with $(\succsim^1, \succsim^2)$, and $\succsim_E^1 = \succsim_E^2$, then $\succsim_E^1 = \succsim_E^2 = \succsim_E^3$.

Next, we define the comparative sensitivity of two preferences on a given event E.

Definition 5. *For all $\succsim^1, \succsim^2 \in \mathscr{R}$, $E \in \mathcal{S}(\succsim^1) \cap \mathcal{S}(\succsim^2)$, and $\lambda \in [1, +\infty)$, $\succsim^2$ is λ times more E-sensitive than $\succsim^1$, denoted by $\succsim^1 \overset{(\lambda,E)}{\rightsquigarrow} \succsim^2$, if for all $f \in \mathcal{F}$ and $x, y \in X$, $fEx \sim^1 y$ implies $\frac{1}{\lambda} fEx + (1 - \frac{1}{\lambda})x \sim^2 y$.*

To understand Definition 5, first note that when $\succsim^1 \overset{(1,E)}{\rightsquigarrow} \succsim^2$, the two preference are identical on E. When $\succsim^1 \overset{(\lambda,E)}{\rightsquigarrow} \succsim^2$ with $\lambda > 1$, the only difference between the two preferences on E is that the decision weight assigned by preference $\succsim^2$ to E is λ times more than that by $\succsim^1$. Thus, whenever fEx is as good as y under $\succsim^1$, $\frac{1}{\lambda} fEx + (1 - \frac{1}{\lambda})x$ is as good as y under $\succsim^2$. In Lemma 5, we provide a characterization result for this comparative sensitivity relation.

Axiom 2—Sensitivity Congruence: For all $\{\succsim^k\}_{k=1}^4 \subseteq \mathscr{R}$, $E \in \cap_{k=1}^4 \mathcal{S}(\succsim^k)$, and $\lambda \in [1, +\infty)$, if $\succsim^3 \overset{(\lambda,E)}{\rightsquigarrow} \succsim^1$, $\succsim^4 \overset{(\lambda,E)}{\rightsquigarrow} \succsim^2$, $\succsim_E^3 = \succsim_E^4$, and $\succsim_E^k$ is ambiguous for all $k \in \{1, 2, 3, 4\}$, then $\succsim_E^1 = \succsim_E^2$.

Like Axiom 1, Axiom 2 applies to situations where ambiguity remains unresolved after updating. We interpret this axiom as follows: If two preferences are updated to identical ex-post preferences after the realization of event E, they should contain similar

[17] All claims regarding FB and their proofs can be found in the Online Appendix.

information on E. Therefore, if two other preferences are λ times more E-sensitive than the two given preferences respectively, they should also contain similar information on E and be updated to identical ex-post preferences after E occurs.

Our next axiom concerns situations in which ambiguity is resolved after updating.

Axiom 3—Sensitivity Independence: For all $\succsim^1, \succsim^2 \in \mathscr{R}$, $E \in \mathcal{S}(\succsim^1) \cap \mathcal{S}(\succsim^2)$, and $\lambda \in [1, +\infty)$, if $\succsim^1 \overset{(\lambda,E)}{\rightsquigarrow} \succsim^2$, and $\succsim^2_E$ is unambiguous, then $\succsim^1_E = \succsim^2_E$.

Axiom 3 states that if the only difference between two preferences is their decision weights assigned to E, then they should be updated to the same ex-post preference when E occurs, provided that ambiguity is resolved after updating. We note that if there is ex-ante ambiguity on E, then the decision weight assigned to E is positively correlated with the degree of ambiguity on E. Thus, Axiom 3 essentially posits that if ambiguity is already resolved after updating, then the degree of ambiguity on the realized event will no longer affect the DM's ex-post preference.

By comparison, FB satisfies the following axiom that strengthens Axioms 2 and 3.

Axiom 3*—Sensitivity Independence*: For all $\succsim^1, \succsim^2 \in \mathscr{R}$, $E \in \mathcal{S}(\succsim^1) \cap \mathcal{S}(\succsim^2)$, and $\lambda \in [1, +\infty)$, if $\succsim^1 \overset{(\lambda,E)}{\rightsquigarrow} \succsim^2$, then $\succsim^1_E = \succsim^2_E$.

The next axiom is the key departure of the contraction rule from FB.

Axiom 4—Non-Ambiguity Persistence: For all $\succsim \in \mathscr{R}$, $E \in \mathcal{S}(\succsim)$, and partition $\Pi = \{S_i\}_{i=1}^n$ of S, if $\succsim$ is Π-unambiguous, and $|E \cap S_i| \leq 1$ for all $i \in \{1, ..., n\}$, then $\succsim_E$ is unambiguous.

We interpret each block S_i in Axiom 4 as a payoff-relevant state and the realized event E as a signal. With this interpretation, Axiom 4 states that if the DM is unambiguous with respect to the payoff-relevant states at the ex-ante stage, then any information does not render her ex-post preference ambiguous. Clearly, this axiom is violated by FB, as shown by our motivating example in the Introduction.

The next axiom is the key behavior axiom of the contraction rule.

Axiom 5—Increased Sensitivity after Updating: For all $\succsim \in \mathscr{R}$ and $E \in \mathcal{S}(\succsim)$, if $\succsim_E$ is unambiguous, then for all $s \in E$, $x, y \in X$, and $f, g \in \mathcal{F}$, if $f \overset{S \backslash s}{=} g$, $g(s) \succ f(s)$, $f \sim x$, $f \sim_E x$ and $g \sim y$, then $g \succsim_E y$.

To understand Axiom 5, note that the act f serves as a benchmark: Its evaluation remains to be x before and after the information. Since $f \overset{S \backslash s}{=} g$, the difference between the DM's evaluations of f and g is driven by their differences on state s. The preference relations $g \sim y$ and $g \succsim_E y$ then reveal that the DM increases the evaluation of g relative to f more after the information than before. Therefore, we interpret this axiom as postulating that the DM becomes more sensitive to payoff differences on state s if the state is not ruled out by the information, provided that ambiguity is resolved after updating.

For any given preference $\succsim$ and event E that is $\succsim$-non-null, we say that $\succsim$ satisfies *updating monotonicity* on E if for all $s \in E$, $x, y \in X$, and $f, g \in \mathcal{F}$, if $f \overset{S\setminus s}{=} g$, $g(s) \succ f(s)$, $f \sim x$, $f \sim_E x$ and $g \sim y$, then $g \succsim_E y$. It seems natural to require *every* preference to satisfy updating monotonicity on *every* non-null event of it: The event, if occurs, would rule out outside states, and thus each remaining state should play a more important role in decision-making. In fact, updating monotonicity can be implied by dynamic consistency under mild conditions.[18] However, Axiom 5 only imposes this property on situations where the updated preference becomes unambiguous. We justify this restriction via the following proposition.

Proposition 4. *Consider an arbitrary updating rule Γ that satisfies Axiom 4, and suppose that for all $\succsim^1, \succsim^2 \in \mathscr{R}$ and $E \in \mathcal{S}(\succsim^1) \cap \mathcal{S}(\succsim^2)$, $\succsim^1 \overset{1,E}{\rightsquigarrow} \succsim^2$ implies $\succsim^1_E = \succsim^2_E$. Then for all $\succsim \in \mathscr{R}$ and $E \in \mathcal{S}(\succsim)$ such that $\succsim_E$ is ambiguous, there exists $\succsim' \in \mathscr{R}$ such that $\succsim \overset{1,E}{\rightsquigarrow} \succsim'$, and $\succsim'$ violates updating monotonicity on E.*

Recall that if $\succsim \overset{1,E}{\rightsquigarrow} \succsim'$, then the two preferences are identical on E. Proposition 4 considers an updating rule that updates preferences that are identical on E to the same ex-post preference when E occurs.[19] Additionally, the proposition requires the updating rule to satisfy Axiom 4, which is the crucial axiom that is satisfied by the contraction rule but violated by other updating rules. By Proposition 4, the largest collection of classes of identical preferences on which we can impose the property of updating monotonicity is the collection of those that are updated to unambiguous preferences. This is precisely what Axiom 5 posits.

Definition 6. *For all $\succsim \in \mathscr{R}$, sequence $(\succsim^k)_{k=1}^{+\infty}$ in $\mathscr{R}$, and event E, $(\succsim^k)_{k=1}^{+\infty}$ converges to $\succsim$ on E if for all $f \in \mathcal{F}$ and $x, y \in X$, $fEx \succsim^k y$ for all $k \in \mathbb{N}_+$ implies $fEx \succsim y$, and $y \succsim^k fEx$ for all $k \in \mathbb{N}_+$ implies $y \succsim fEx$.*

Our next axiom is the Continuity axiom.

Axiom 6—Continuity: For all $\succsim^0 \in \mathscr{R}$, sequence $(\succsim^k)_{k=1}^{+\infty}$ in $\mathscr{R}$, and $E \in \cap_{k=0}^{+\infty} \mathcal{S}(\succsim^k)$, if $(\succsim^k)_{k=1}^{+\infty}$ converges to $\succsim^0$ on E, then $(\succsim^k_E)_{k=1}^{+\infty}$ converges to $\succsim^0_E$ on E.

Together, Axioms 1-6 characterize the contraction rule.

---

[18] In the Online Appendix, we show that in a more general preference domain in which the preferences are assumed to satisfy Axioms M1, M2 and M4, dynamic consistency implies updating monotonicity. This preference domain includes the class of preferences that allow for dual-self expected utility representations (Chandrasekher, Frick, Iijima, and Le Yaouanq, 2022) which generalize the maxmin representation.

[19] In fact, for any preferences $\succsim$ and $\succsim'$ and event E such that $\succsim \overset{1,E}{\rightsquigarrow} \succsim'$, FB, ML, and the contraction rule all update them to identical preferences when E occurs. By Lemma 5, this requirement is essentially equivalent to postulate 1 that we discussed in Section 2.2.

Theorem 1. *An updating rule is the contraction rule if and only if it satisfies Axioms 1-6.*

Sketch of the sufficiency part of the proof. Consider an updating rule Γ that satisfies Axioms 1-6 and an arbitrary preference $\succsim$ in $\mathscr{R}$ that is represented by (u, P). Our first step is to show that if $\succsim$ is updated by Γ to some ex-post preference $\succsim_E$ when E occurs, then $\succsim_E$ is unambiguous if and only if $\mu_P(E) \leq 1$. It then follows from Increased Sensitivity after Updating that $\mu_P(E) = 1$ implies $\succsim_E$ to be represented by $(u, \{\mu_P|E\})$.

The remaining two cases are $\mu_P(E) < 1$ and $\mu_P(E) > 1$. For the case $\mu_P(E) < 1$, we want to show that $\succsim_E$ is represented by $(u, \{\overline{\mu_P|E}\})$. We show this by considering another preference $\succsim'$ that is represented by (u, P') such that $P'|E = \lambda P|E$ where $\lambda = \frac{1}{\mu_P(E)}$. By Lemma 5, we have $\succsim \overset{(\lambda,E)}{\rightsquigarrow} \succsim'$. Therefore, by Sensitivity Independence, we have $\succsim_E = \succsim'_E$. Since $\mu_{P'}(E) = \lambda\mu_P(E) = 1$, $\succsim'_E$ is represented by $(u, \{\mu_{P'}|E\})$, i.e., $(u, \{\overline{\mu_P|E}\})$. Hence, $\succsim_E$ is also represented by $(u, \{\overline{\mu_P|E}\})$.

For the case $\mu_P(E) > 1$, we sketch its proof through Figure 3. We want to show that when E occurs, $\succsim$ is updated by Γ to $\succsim'$, where $\succsim'$ is represented by (u, Q) with $Q = Q^c(P, E)$ (as shown in the figure). Let $\lambda = \mu_P(E)$, and consider preferences $\succsim^1$, $\succsim^2$ and $\succsim^3$ such that there are represented by $\hat{P}$, $\hat{Q}$ and $\{\mu_{\hat{P}}|E\}$ respectively, where $\lambda\hat{P}|E = P$ and $\lambda\hat{Q}|E = Q$. By Lemma 3, $\succsim^2$ is E-aligned with $(\succsim^1, \succsim^3)$. This enables us to apply some continuity arguments, together with Alignment Consistency and Sensitivity Congruence, to show that the two preferences $\succsim$ and $\succsim'$, which satisfy $\succsim^1 \overset{(\lambda,E)}{\rightsquigarrow} \succsim$ and $\succsim^2 \overset{(\lambda,E)}{\rightsquigarrow} \succsim'$, should be updated by Γ to the same ex-post preference when E occurs.[20] Clearly, since $S \backslash E$ is $\succsim'$-null, $\succsim'$ is updated to $\succsim'$ by Γ when E occurs, and so is $\succsim$. $\square$

Regarding Theorem 1, we have two remarks. First, our characterization result is not only applicable to maxmin preferences, but also to maxmax preferences—the preferences that capture ambiguity seeking behavior. Formally, a preference $\succsim$ is a simple maxmax preference if there is a non-constant and affine utility function $u : X \to \mathbb{R}$ and a set of probability distributions $P \in \mathscr{P}$ such that for all $f, g \in \mathcal{F}$, $f \succsim g$ if and only if $\max_{p \in P} u(f; p) \geq \max_{p \in P} u(g; p)$. Given a rich set of maxmax preferences $\mathscr{R}^+$ that satisfies similar conditions as $\mathscr{R}$, we can show that the contraction rule, defined on $\mathscr{R}^+$, can be characterized by the same set of axioms, with a slight revision of Increased Sensitivity after Updating.[21] The proof strategy is the same as that for Theorem 1.

Second, among the axioms that characterize the contraction rule, two are not satisfied by FB: Non-Ambiguity Persistence and Continuity. However, in the Online Appendix,

[20] Note that Alignment Consistency and Sensitivity Congruence cannot be directly applied here, as $\succsim^1$ and $\succsim^2$ are updated by Γ to an *unambiguous* preference when E occurs. See Lemma 19 for more details.

[21] All our definitions and notations can be directly applied to preferences in $\mathscr{R}^+$ without any further modification. The revised version of the axiom of Increased Sensitivity after Updating states that for all $\succsim \in \mathscr{R}^+$ and $E \in \mathcal{S}(\succsim)$, if $\succsim_E$ is unambiguous, then for all $s \in E$, $x, y \in X$, and $f, g \in \mathcal{F}$, if $f \overset{S \backslash s}{=} g$, $f(s) \succ g(s)$, $f \sim x$, $f \sim_E x$ and $g \sim y$, then $y \succsim_E g$.

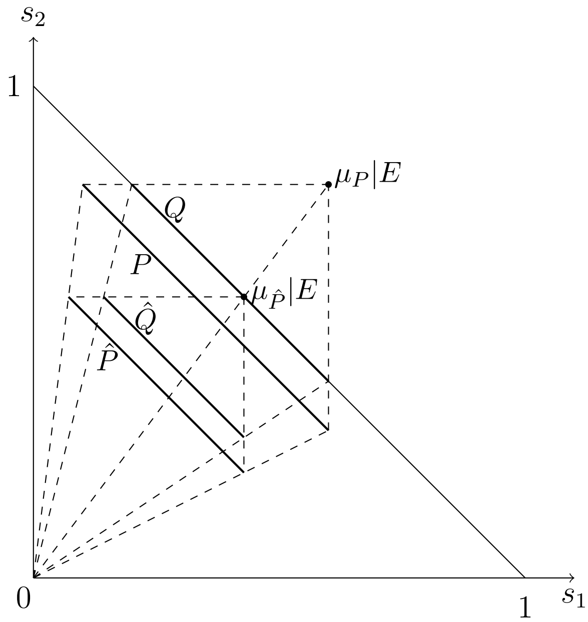


Figure 3: Graphic Illustration for the Case $\mu_P(E) > 1$

we show that FB *generically* satisfies the Continuity axiom. Therefore, essentially, FB only violates the axiom of Non-Ambiguity Persistence. We also show that FB can be characterized by Axioms 1^* and 3^* in the Online Appendix.

4 Applications

Throughout this section, we assume that each prior of the DM is supported in $\hat{S} \subseteq S$ where $\hat{S} = D \times \Theta$. The set 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 $\hat{S}$.[22]

We consider three applications of the contraction rule. First, we connect the contraction rule with the experimental findings by SO23 and show that the rule does not create belief dilation under mild conditions. Second, we study how DMs update their beliefs with information of unknown accuracy by associating the contraction rule with the findings by L24. Finally, we consider an individual whose current and future preferences are not aligned, and show that the individual can maximally align her future choices with her current preference via ambiguous information if she adopts the contraction rule for belief updating.

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

4.1 Dilation

SO23 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 belief set P over $D \times \Theta$. A signal $\theta \in \Theta$ is said to dilate 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 belief set over D strictly contains her prior one.[23] 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 SO23. 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 SO23, the realization of any signal dilates the DM's belief set over D if she updates with 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 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 SO23.

When $a > 0$, there is ex-ante payoff-relevant ambiguity: The DM believes that the probability of d_1 ranges from $1/2 - a$ to $1/2 + a$. For any given signal, there are two cases to be considered with the contraction rule. If $(1/2 + a)(1/2 + b) \le 1/2$, 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.

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

This is indeed the case for a non-negligible proportion of ambiguity averse subjects in SO23'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 belief set is contained in the interior of $\Delta(D \times \Theta)$.

Proposition 6. *If either (i) $P_{\hat{\Pi}}$ is a singleton, or (ii) for all $p \in P$, $d \in D$ and $\theta \in \Theta$, $p(d, \theta) > 0$, 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 L24. 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 $[\underline{\kappa}, \overline{\kappa}] \subseteq (0, 1)$, i.e., she believes that the probability of d_1 ranges from $\underline{\kappa}$ to $\overline{\kappa}$. Following L24, 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 $\hat{d}_2$, where $\hat{d}_1$ and $\hat{d}_2$ refer to the prediction of d_1 and d_2, respectively. 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: $\chi^H(\hat{d}_1|d_1) = \chi^H(\hat{d}_2|d_2) = H$ and $\chi^L(\hat{d}_1|d_1) = \chi^L(\hat{d}_2|d_2) = L$. We require that $1 > H > L > 0$ and $H + L > 1$.[24] 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 $co(\{p_{\underline{\kappa},L}, p_{\underline{\kappa},H}, p_{\overline{\kappa},L}, p_{\overline{\kappa},H}\})$, where for each $\kappa \in \{\underline{\kappa}, \overline{\kappa}\}$ and $J \in \{L, H\}$, we have

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

In scenario 2, there is an expert with accuracy $\frac{H+L}{2}$. 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 given by $co(\{p_{\overline{\kappa}}, p_{\underline{\kappa}}\})$ where for each $\kappa \in \{\overline{\kappa}, \underline{\kappa}\}$,

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

To proceed, 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 $f(d_2, \hat{d}_1) = f(d_2, \hat{d}_2) = g(d_1, \hat{d}_1) = g(d_1, \hat{d}_2) = 0$. That is, f yields payoff

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

1 on d_1 and 0 on d_2, and g yields 0 on d_1 and 1 on d_2.[25] We compare the DM's ex-post evaluations over the two acts in the two scenarios. Let v_f^a and v_f^u be the DM's ex-post evaluations of f after observing $\hat{d}_1$ in scenarios 1 and 2, respectively. Let v_g^a and v_g^u be the DM's ex-post evaluations of g after observing $\hat{d}_1$ in scenarios 1 and 2, respectively.

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 the value of $\frac{1-\overline{\kappa}}{1-\underline{\kappa}}$:*
(1) *if $\frac{1-\overline{\kappa}}{1-\underline{\kappa}} < \frac{1-\underline{\kappa}}{1-\underline{\kappa}H}\frac{2-H-L}{H+L}$, then $v_g^a < v_g^u$,*
(2) *if $\frac{1-\overline{\kappa}}{1-\underline{\kappa}} = \frac{1-\underline{\kappa}}{1-\underline{\kappa}H}\frac{2-H-L}{H+L}$, then $v_g^a = v_g^u$,*
(3) *if $\frac{1-\underline{\kappa}}{1-\underline{\kappa}H}\frac{2-H-L}{H+L} < \frac{1-\overline{\kappa}}{1-\underline{\kappa}}$, 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 in the ambiguous scenario (scenario 1).

However, contraction rule does not always predict under-reaction to ambiguously bad news. Bad news not only pushes the DM's priors towards the bad state but also partially resolves the payoff-relevant ambiguity. Thus, a DM who exhibits ambiguity aversion may increase her evaluation of the given act even when bad news is received. In the proposition, the value of $\frac{1-\overline{\kappa}}{1-\underline{\kappa}}$ captures the degree of the DM's ex-ante ambiguity on D. A larger value of $\frac{1-\overline{\kappa}}{1-\underline{\kappa}}$ corresponds to less ex-ante ambiguity. When $\frac{1-\overline{\kappa}}{1-\underline{\kappa}}$ is small (case (1)), the DM's ex-ante ambiguity is large, and she benefits more from the resolution of ambiguity. Consequently, the DM decreases her evaluation of g more in scenario 1 than in scenario 2, as unambiguous information helps to resolve more of her ex-ante ambiguity. As the ex-ante ambiguity decreases, the effect of ambiguity resolution 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 offered by L24, 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 in cases where the ex-ante ambiguity is large, while a certain proportion of subjects in L24's experiments exhibit pessimism for such news even when there is no ex-ante ambiguity. Nevertheless, our results are relevant in the empirical analysis of the stock price reactions by L24, since in these scenarios, the subjects are shown to exhibit pessimism to ambiguously bad news, and the ex-ante ambiguity of stock prices is typically large.

[25] We assume for simplicity that the consequence space is given by $X = \mathbb{R}$, and the DM's utility function $u : X \to \mathbb{R}$ satisfies $u(x) = x$ for all $x \in X$.

4.3 Contraction Rule as a Tool for Self-Control

In this section, we study optimal information design with the contraction rule in situations where ambiguous information can be used. The concept of information design was initially introduced and explored by Kamenica and Gentzkow (2011) in a purely Bayesian setting, and was investigated by Beauchêne, Li, and Li (2019) in the context where agents are assumed to be ambiguity averse and update with FB.

To provide a context for our exploration of the optimal information design problem, consider a scenario in which an individual's current preference differs from her future preference. The individual is aware that her future-self will choose from a set of feasible actions, and she can design an information structure that influences her future-self's beliefs in order to align her future choices with her current preference. We demonstrate that if the individual uses the contraction rule for belief updating, then she can maximally align her future choices with her current preference through appropriate information structures.

Formally, let $R \subseteq \Delta(D)$ be the individual's prior belief set, where D contains all payoff-relevant states. WLOG, assume that for every $d \in D$, there exists $r \in R$ such that $r(d) > 0$.[26] An (ambiguous) information structure is a tuple $\mathcal{I} = (\Theta, \{\chi_k\}_{k=1}^n)$ where Θ is a nonempty and finite set of signals, and for every k, χ_k denotes the distributions over Θ conditional on every $d \in D$, i.e., for every $d \in D$, $\chi_k(\cdot|d) \in \Delta(\Theta)$. For a given information structure $\mathcal{I} = (\Theta, \{\chi_k\}_{k=1}^n)$, we denote by $P_{\mathcal{I}} \subseteq \Delta(D \times \Theta)$ the set of joint distributions induced by $\mathcal{I}$ and the prior set R such that $p \in P_{\mathcal{I}}$ if and only if there exists $r \in R$ and $k \in \{1, ..., n\}$ such that for all $(d, \theta) \in D \times \Theta$, $p(d, \theta) = r(d)\chi_k(\theta|d)$.

Let A be a nonempty and finite set of actions from which the individual's future-self will make a choice. Let $w : A \times D \to \mathbb{R}$ and $v : A \times D \to \mathbb{R}$ be the individual's current and future utility functions, respectively. We denote by $A_v^* \subseteq A$ the set of actions that are optimal for the future-self under some belief, i.e.,

$$A_v^* = \{a \in A : \exists r \in \Delta(D) \text{ such that } \forall \hat{a} \in A, \sum_{d \in D} r(d)v(a, d) \geq \sum_{d \in D} r(d)v(\hat{a}, d)\}.$$

Note that the future-self will only choose actions from A_v^* even if she has ambiguous beliefs and adopts maxmin or maxmax criterion.[27] This is because if an action a is not in A_v^*, then by a simple proof similar to that of Lemma 3 in Pearce (1984), we can show that a is strictly dominated by some $\sigma \in \Delta(A)$ with $\sigma(a) = 0$ in the sense that for all $r \in \Delta(D)$,

$$\sum_{d \in D} \sum_{\hat{a} \in A} r(d)\sigma(\hat{a})v(\hat{a}, d) > \sum_{d \in D} r(d)v(a, d).$$

[26] If for some $d \in D$, we have $r(d) = 0$ for all $r \in R$, then we can rule out the state d and consider the state space $D \backslash \{d\}$.

[27] I thank Xiaoyu Cheng for pointing this out.

Let $\Delta^o(D)$ be the set of distributions over D that assign each state in D a positive probability. For the set A_v^*, we impose the following generic assumption, which requires that each $a \in A_v^*$ is optimal under some distribution in $\Delta^o(D)$.

Assumption 1. For all $a \in A_v^*$, there exists $r \in \Delta^o(D)$ such that for all $\hat{a} \in A$, $\sum_{d\in D} r(d)v(a,d) \geq \sum_{d\in D} r(d)v(\hat{a},d)$.

For a given information structure $\mathcal{I} = (\Theta, \{\chi_k\}_{k=1}^n)$, we say that the function $\zeta : \Theta \to A_v^*$ is $\mathcal{I}$-implementable if for all $\theta \in \Theta$, $q \in Q^c(P_{\mathcal{I}}, D \times \{\theta\})$ and $a \in A$, we have $\sum_{d\in D} q(d)v(\zeta(\theta),d) \geq \sum_{d\in D} q(d)v(a,d)$.[28] Note that if ζ is $\mathcal{I}$-implementable, then for every $\theta \in \Theta$, $\zeta(\theta)$ is optional for the future-self when θ is observed, regardless of whether she adopts the maxmin or maxmin criterion. With Assumption 1, the following theorem demonstrates that if the individual updates with the contraction rule, then she can manipulate her future choices within A_v^* almost arbitrarily.

Theorem 2. *Under Assumption 1, for every function $\tau : D \to A_v^*$ and $\lambda \in (0,1)$, there is an information structure $\mathcal{I} = (\Theta, \{\chi_k\}_{k=1}^n)$ and an $\mathcal{I}$-implementable function $\zeta : \Theta \to A_v^*$ such that for all $d \in D$ and $k \in \{1, ..., n\}$, $\sum_{\theta\in\Theta:\, \zeta(\theta)=\tau(d)} \Big(\chi_k(\theta|d)\Big) > \lambda$.*

According to Theorem 2, any function $\tau : D \to A_v^*$ that maps every payoff-relevant state to an action that the individual desires for her future-self to take under that state can be almost achieved through an appropriate information structure. We can interpret this theorem in two ways. First, the individual is aware that her future-self updates beliefs using the contraction rule. Therefore, the individual can control her future choices by committing to an ambiguous information structure. Specifically, she can achieve near-optimal outcomes by setting λ close to 1 and choosing τ as the function that maps each payoff-relevant state to the w-optimal action in A_v^* under that state. Second, Theorem 2 can be viewed as a desirable *normative feature* of the contraction rule for individuals who seek to exercise greater control over their future-selves. These individuals can *choose* to update their beliefs using the contraction rule, thereby enabling them to regulate their future choices through suitable information structures.

In what follows, we provide an example to demonstrate the key idea of Theorem 2. Our example also serves to demonstrate that FB and ML are incapable of achieving the same level of near-optimal self-control as the contraction rule.

Example 3. Consider an individual who has just undergone a cholesterol examination and is expecting the results to be released in two days. The individual is anticipating two possible results: Either all indexes will be normal (d_0) or some index will be abnormal (d_1). Due to the lack of information regarding the likelihood of either outcome, the individual's prior belief set is $R = \Delta(D)$, where $D = \{d_0, d_1\}$. Upon receiving relevant

[28] Each q, which should be a distribution over $D \times \{\theta\}$, is treated as a distribution over D for simplicity.

information about the examination result, the individual's future-self can respond to the information by adjusting her diet plan. Denote by $A = \{0, \frac{1}{3}, \frac{1}{2}, \frac{2}{3}, 1\}$ the set of diet plans, where a larger $a \in A$ indicates a more healthier diet plan for the next six months. The individual's future-self has preference v which satisfies that for all $a \in A$, $v(a, d_0) = -a^2$ and $v(a, d_1) = -(1-a)^2$. According to this preference, the future-self is highly responsive to the examination result: If there is an abnormality, she will choose to eat extremely healthy ($a = 1$); otherwise, she will opt for an extremely unhealthy diet ($a = 0$). However, the individual desires for her future-self to respond in a more gradual manner to the examination result, as a sudden and drastic change in dietary habits may have adverse effects on her health. In particular, she wants her future-self to choose $a = \frac{1}{3}$ at d_0 and $a = \frac{2}{3}$ at d_1.

Our first observation is that if the future-self updates her beliefs using either FB or ML, then the current-self cannot achieve her desired future choices through the use of ambiguous information. This is because, with FB or ML, regardless of the information structure that the individual commits to, the future-self would either maintain her prior belief set R or update her beliefs to a degenerate belief on either d_0 or d_1, upon receiving any signal. In the former case, the future-self would choose $a = \frac{1}{2}$, while in the latter case, she would choose $a = 0$ or $a = 1$.

In contrast, if the individual's future-self uses the contraction rule, then her current-self can design the following information structure to regulate future choices. The individual can send the examination result to 20 friends, each labeled from 1 to 20, without revealing the result to herself. These friends are divided into two groups: Group 0 consists of those whose labels are from 1 to 10, and group 1 comprises the rest. There are a total of 20 signals denoted by $\{\theta_k\}_{k=1}^{20}$. For any friend k_0 in group 0 and k_1 in group 1 (i.e., $k_0 \leq 10$ and $k_1 > 10$), their signal distributions conditional on the examination results are presented in Table 3.

Friend k_0's Signal Distributions	d_0	d_1
Probability of Sending θ_{k_0}	1/5	1/10
Probability of Sending $\theta_{k'}$ if $k' \leq 10$ and $k' \neq k_0$	4/45	0
Probability of Sending $\theta_{k'}$ if $k' > 10$	0	9/100
Friend k_1's Signal Distributions	d_0	d_1
Probability of Sending θ_{k_1}	1/10	1/5
Probability of Sending $\theta_{k'}$ if $k' \leq 10$	9/100	0
Probability of Sending $\theta_{k'}$ if $k' > 10$ and $k' \neq k_1$	0	4/45

Table 3: Signal Distributions of Friends from the Two Groups

The individual can require her friends to adopt a secret rule to determine whose signal is sent to her so that upon receiving the signal, she is completely uncertain about the sender. Consider the case in which the individual receives signal θ_k. It follows that if $k \leq 10$, then the maximal probabilities for (d_0, θ_k) and (d_1, θ_k) are given respectively by

1/5 and 1/10, and if $k > 10$, they are given respectively by 1/10 and 1/5. With the contraction rule, the individual will always form a singleton posterior set upon receiving any signal θ_k, and she will choose $a = \frac{1}{3}$ when $k \leq 10$ and $a = \frac{2}{3}$ when $k > 10$. According to the information structure shown in Table 3, no matter which friend is selected for sending the signal, the probability for the chosen action to be misaligned with the individual's current preference is at most 1/10. For instance, if friend 1 is selected, then the only misalignment occurs when the true result is d_1 and signal θ_1 is sent, which happens with at most 1/10 chance.

Notably, when signal θ_k is received, the individual will rely on the signal distributions of friend k to update her beliefs, as friend k's signal distributions assign the highest conditional probabilities to signal θ_k compared to other friends' signal distributions. While this feature ensures that the posterior belief of the individual can be easily manipulated, it also accounts for the misalignment between the chosen action and the individual's current preference, as the conditional probabilities of θ_k of friend k are higher than those of others and thus are bounded below. One way to decrease the probability of such misalignment is to lower the conditional probabilities of every signal for every friend, which can be implemented by increasing the number of friends invited as well as the number of signals. As the number of friends and signals increases, the probability of misalignment can be made arbitrarily close to 0.

5 Appendix

Proof of Proposition 1. Consider $P \in \mathscr{P}$ and $E \in \mathcal{S}_P$. If $\mu_P(E) \leq 1$, then $|Q^c(P, E)| = 1$, since $Q^c(P, E) = \{\overline{\mu_P|E}\}$. If $\mu_P(E) > 1$, then $Q^c(P, E) = \{\Phi(p|E, \mu_P|E)\}_{p \in P}$. Consider 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 $\hat{p}(s^*) < \mu_P(s^*)$. Pick $p^* \in P$ such that $p^*(s^*) = \mu_P(s^*)$. It follows that $\Phi(p^*|E, \mu_P|E)(s^*) = \mu_P(s^*) > \Phi(\hat{p}|E, \mu_P|E)(s^*)$, and we have $\Phi(p^*|E, \mu_P|E) \neq \Phi(\hat{p}|E, \mu_P|E)$. Therefore, when $\mu_P(E) > 1$, $|Q^c(P, E)| \neq 1$.

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

Proof of Proposition 2. Consider $P \in \mathscr{P}$ and P-non-null event E. If $\mu_P(E) \leq 1$, then $Q^c(P, E) = \{\overline{\mu_P|E}\}$, which is nonempty, convex, and closed. If $\mu_P(E) > 1$, then $Q^c(P, E) = \{\Phi(p|E, \mu_P|E)\}_{p \in P}$, which is nonempty. 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 see that $Q^c(P, E)$ is closed, consider a sequence of probability distributions $(q_n)_{n=1}^{+\infty}$ in $Q^c(P,E)$ 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 compact, it is without loss of generality to assume that $(p_n)_{n=1}^{+\infty}$ converges to some $p \in P$. It follows that $(\Phi(p_n|E, \mu_P|E))_{n=1}^{+\infty}$—i.e., $(q_n)_{n=1}^{+\infty}$—converges to $\Phi(p|E, \mu_P|E) = q$. 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 equation (2), we have $\Phi(p|E, \mu_P|E) = p|E = p$ for all $p \in P$. Therefore, we conclude that $Q^c(P,E) = P$. □

Proof of Proposition 3. 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 $\mu_P|E = \mu_{Q^c(P,E)}$. It follows that $\mu_P|F = \mu_{Q^c(P,E)}|F$. By a similar argument as case (i), we have $Q^c(Q^c(P,E),F) = Q^c(P,F)$. In case (iii), $\mu_P(E) > 1$ and $\mu_P(F) > 1$. We need the following claim, of which the proof is simple algebra and thus omitted.

Claim 1. *For all $\pi, \pi' \in \mathcal{M}(S)$ and $\alpha \in (0,1]$ 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 Claim 1, for every $p \in P$, we have $\Phi(\Phi(p|E, \mu_P|E)|F, \mu_P|F) = \Phi(p|F, \mu_P|F)$, and we are done. □

Proofs of Proposition 4 and Theorem 1. Given the conditions we impose on $\mathscr{R}$, there exists a non-constant and affine utility function $u : X \to \mathbb{R}$ such that (i) for every $\succsim \in \mathscr{R}$, there is $P \in \mathscr{P}$ such that (u,P) represents $\succsim$, and (ii) for every $P \in \mathscr{P}$, there is $\succsim \in \mathscr{R}$ such that (u,P) represents $\succsim$. Throughout the proof, fix the utility function u and say that P represents $\succsim$ if (u,P) represents $\succsim$. We use Π to denote the partition $\{\{s\} : s \in E\} \cup \{S \backslash E\}$ whenever E is specified. For all $f \in \mathcal{F}$, $x \in X$, and $p \in \Delta(S)$, define $u(fEx; p_\Pi) = \sum_{s \in E} p(s)u(f(s)) + p(S \backslash E)u(x)$, and for all $P \in \mathscr{P}$, define $u^\downarrow(fEx; P_\Pi) = \min_{p \in P} u(fEx; p_\Pi)$. We have $u(fEx; p) = u(fEx; p_\Pi)$ and $u^\downarrow(fEx; P_\Pi) = u^\downarrow(fEx; P)$. The first lemma is trivial, and thus we omit its proof.

Lemma 1. *For all $\succsim \in \mathscr{R}$, $P \in \mathscr{P}$, and partition Ω of S, if P represents $\succsim$, then (i) $\mathcal{S}_P = \mathcal{S}(\succsim)$, (ii) $\succsim$ is Ω-unambiguous if and only if P_Ω is a singleton, and (iii) $\succsim$ is unambiguous if and only if P is a singleton.*

Lemma 2. *For all $P^1, P^2 \in \mathscr{P}$ and event E with $\mu_{P^1}(E) > 1$ and $\mu_{P^2}(E) > 1$, if $Q^c(P^1,E) = Q^c(P^2,E)$, then $Q^c(co(P^1 \cup P^2),E) = Q^c(P^1,E)$.*

Proof. Since $\mu_{P^1}(E) > 1$, $\mu_{P^2}(E) > 1$ and $Q^c(P^1, E) = Q^c(P^2, E)$, by Proposition 1, we have $\mu_{P^1}|E = \mu_{Q^c(P^1,E)} = \mu_{Q^c(P^2,E)} = \mu_{P^2}|E$. Therefore, $\mu_{P^1}|E = \mu_{P^2}|E = \mu_{co(P^1\cup P^2)}|E$, and it follows that $Q^c(P^1, E) = \{\Phi(p|E, \mu_{P^1}|E)\}_{p\in P^1} \subseteq \{\Phi(q|E, \mu_{P^1}|E)\}_{q\in co(P^1\cup P^2)} = Q^c(co(P^1 \cup P^2), E)$. It remains to show $Q^c(co(P^1 \cup P^2), E) \subseteq Q^c(P^1, E)$. Note that measures over E can be regarded as elements in $\mathbb{R}^E$. Define $G = \{\alpha q + (1-\alpha)\mu_{P^1}|E : \alpha \in [0, +\infty), q \in Q^c(P^1, E)\} \subseteq \mathbb{R}^E$. Since $Q^c(P^1, E)$ is convex (by Proposition 2), G is also convex. By the construction of G, we have $P^1|E \subseteq G$ and $P^2|E \subseteq G$, and thus $co(P^1|E \cup P^2|E) \subseteq G$. It then follows that $Q^c(co(P^1 \cup P^2), E) \subseteq Q^c(P^1, E)$. □

Lemma 3. *For all $\{\succsim^k\}_{k=1}^3 \subseteq \mathscr{R}$, $\{P^k\}_{k=1}^3 \subseteq \mathscr{P}$, and $E \in \cap_{k=1}^3 \mathcal{S}(\succsim^k)$ such that P^k represents $\succsim^k$ for all $k \in \{1,2,3\}$, $\succsim^3$ is E-aligned with $(\succsim^1, \succsim^2)$ if and only if (i) $P^3_\Pi \subseteq co(P^1_\Pi \cup P^2_\Pi)$, and (ii) for every $p \in P^1$ and $q \in P^2$, there exists $\alpha \in [0,1]$ such that $\alpha p_\Pi + (1-\alpha)q_\Pi \in P^3_\Pi$.*

Proof. For sufficiency, consider arbitrary $f \in \mathcal{F}$ and $x, y \in X$. If $fEx \succsim^1 y$ and $fEx \succsim^2 y$, then we have $u^\downarrow(fEx; P^1_\Pi) \geq u(y)$ and $u^\downarrow(fEx; P^2_\Pi) \geq u(y)$. It follows that $u^\downarrow(fEx; co(P^1_\Pi \cup P^2_\Pi)) \geq u(y)$. By condition (i), we have $u^\downarrow(fEx; P^3_\Pi) \geq u^\downarrow(fEx; co(P^1_\Pi \cup P^2_\Pi)) \geq u(y)$. That is, $fEx \succsim^3 y$. If $y \succsim^1 fEx$ and $y \succsim^2 fEx$, then there exists $p \in P^1$ and $q \in P^2$ such that $u(y) \geq u(fEx; p_\Pi)$ and $u(y) \geq u(fEx; q_\Pi)$. By condition (ii), there exists $\alpha \in [0,1]$ such that $r_\Pi = \alpha p_\Pi + (1-\alpha)q_\Pi \in P^3_\Pi$. Thus, we have $u^\downarrow(fEx; P^3_\Pi) \leq u(fEx; r_\Pi) \leq u(y)$. That is, $y \succsim^3 fEx$.

For necessity, we first show condition (i). Suppose to the contrary that there exists $p \in P^3$ such that $p_\Pi \notin co(P^1_\Pi \cup P^2_\Pi)$. By the separating hyperplane theorem (SHT), we can find $f \in \mathcal{F}$ and $x, y \in X$ such that $u(fEx; p_\Pi) < u(y) < u^\downarrow(fEx; co(P^1_\Pi \cup P^2_\Pi))$. It follows that $u^\downarrow(fEx; P^3_\Pi) < u(y) < u^\downarrow(fEx; co(P^1_\Pi \cup P^2_\Pi))$. That is, $fEx \succ^1 y$, $fEx \succ^2 y$ and $y \succ^3 fEx$, contradicting with the E-alignment relation. To see condition (ii), suppose to the contrary that there exists $p \in P^1$ and $q \in P^2$ such that $co(\{p_\Pi, q_\Pi\}) \cap P^3_\Pi = \emptyset$. By SHT, there exists $f \in \mathcal{F}$ and $x, y \in X$ such that $u^\downarrow(fEx; P^3_\Pi) > u(y) > \max\{u(fEx; p_\Pi), u(fEx; q_\Pi)\}$. It then follows that $fEx \succ^3 y$, $y \succ^1 fEx$ and $y \succ^2 fEx$, which is a contradiction. □

For all $\{P^k\}_{k=1}^3 \subseteq \mathscr{P}$ and event E, if conditions (i) and (ii) in the statement of Lemma 3 hold, then we say that P^3 is E-aligned with (P^1, P^2).

Lemma 4. *For all $\{P^k\}_{k=1}^3 \subseteq \mathscr{P}$ and $E \in \cap_{k=1}^3 \mathcal{S}_{P^k}$, if P^3 is E-aligned with (P^1, P^2) and $\mu_{P^k}(E) > 1$ for all $k \in \{1,2\}$, then $Q^c(P^1, E) = Q^c(P^2, E)$ implies $Q^c(P^1, E) = Q^c(P^2, E) = Q^c(P^3, E)$.*

Proof. We first show $\mu_{P^1}|E = \mu_{P^2}|E = \mu_{P^3}|E$. Since $\mu_{P^1}(E) > 1$ and $\mu_{P^2}(E) > 1$, by Proposition 1, $Q^c(P^1, E) = Q^c(P^2, E)$ implies $\mu_{P^1}|E = \mu_{P^2}|E$. Since P^3 is E-aligned with (P^1, P^2), $P^3_\Pi \subseteq co(P^1_\Pi \cup P^2_\Pi)$. Thus for every $s \in E$, we have $\mu_{P^3}(s) \leq \mu_{P^1}(s)$. For each $s \in E$, there exists $p^1 \in P^1$ and $p^2 \in P^2$ such that $p^1(s) = p^2(s) = \mu_{P^1}(s)$. By the

E-alignment relation, there exists $\alpha \in [0,1]$ such that $\alpha p^1_\Pi + (1-\alpha)p^2_\Pi \in P^3_\Pi$, indicating that $\mu_{P^3}(s) \geq \mu_{P^1}(s)$. Thus $\mu_{P^1}|E = \mu_{P^3}|E$.

Since $P^3_\Pi \subseteq co(P^1_\Pi \cup P^2_\Pi)$, by Lemma 2 and $\mu_{P^1}|E = \mu_{P^3}|E$, we have $Q^c(P^3, E) \subseteq Q^c(co(P^1 \cup P^2), E) = Q^c(P^1, E)$. It remains to show $Q^c(P^1, E) \subseteq Q^c(P^3, E)$. For every $q \in Q^c(P^1, E) = Q^c(P^2, E)$, there exists $p^1 \in P^1$ and $p^2 \in P^2$ such that $\Phi(p^1|E, \mu_{P^1}|E) = \Phi(p^2|E, \mu_{P^2}|E) = q$. By the E-alignment relation, there exists $\alpha \in [0,1]$ such that $\alpha p^1_\Pi + (1-\alpha)p^2_\Pi \in P^3_\Pi$. Thus $q = \Phi(\alpha p^1|E + (1-\alpha)p^2|E, \mu_{P^1}|E) = \Phi(\alpha p^1|E + (1-\alpha)p^2|E, \mu_{P^3}|E) \in Q^c(P^3, E)$. We are done. $\square$

Lemma 5. *For all $\succsim^1, \succsim^2 \in \mathscr{R}$, $P^1, P^2 \in \mathscr{P}$, $E \in \mathcal{S}(\succsim^1) \cap \mathcal{S}(\succsim^2)$ and $\lambda \in [1, +\infty)$ such that P^1 and P^2 represent $\succsim^1$ and $\succsim^2$ respectively, $\succsim^1 \overset{(\lambda,E)}{\rightsquigarrow} \succsim^2$ if and only if $\lambda P^1|E = P^2|E$.*

Proof. For the sufficiency part, consider arbitrary $f \in \mathcal{F}$ and $x \in X$ such that $fEx \sim^1 y$. We have

$$\begin{aligned}
u(y) = u^{\downarrow}(fEx; P^1_\Pi) &= \min_{p \in P^1} \Big(\sum_{s \in E} p(s)u(f(s)) + (1-p(E))u(x) \Big) \\
&= \min_{p \in P^1} \Big(\sum_{s \in E} \lambda p(s)(\frac{1}{\lambda}u(f(s)) + (1-\frac{1}{\lambda})u(x)) + (1-\lambda p(E))u(x) \Big) \\
&= \min_{q \in P^2} \Big(\sum_{s \in E} q(s)(\frac{1}{\lambda}u(f(s)) + (1-\frac{1}{\lambda})u(x)) + (1-q(E))u(x) \Big) \\
&= u^{\downarrow}(\frac{1}{\lambda}fEx + (1-\frac{1}{\lambda})x; P^2_\Pi)
\end{aligned}$$

It follows that $\frac{1}{\lambda}fEx + (1-\frac{1}{\lambda})x \sim^2 y$, and thus $\succsim^1 \overset{(\lambda,E)}{\rightsquigarrow} \succsim^2$. The necessity part can be implied by the uniqueness of the maxmin representation. $\square$

Lemma 6. *For all $\{P^k\}_{k=1}^4 \subseteq \mathscr{P}$, event E and $\lambda \in [1, +\infty)$ such that $\mu_{P^k}(E) > 1$ for all $k \in \{1,2,3,4\}$, if $P^1|E = \lambda P^3|E$ and $P^2|E = \lambda P^4|E$, then $Q^c(P^3, E) = Q^c(P^4, E)$ implies $Q^c(P^1, E) = Q^c(P^2, E)$.*

Proof. Since $Q^c(P^3, E) = Q^c(P^4, E)$, $\mu_{P^3}(E) > 1$ and $\mu_{P^4}(E) > 1$, we have $\mu_{P^3}|E = \mu_{P^4}|E$. Since $P^1 = \lambda P^3|E$ and $P^2 = \lambda P^4|E$, we have $\mu_{P^1}|E = \mu_{P^2}|E = \lambda \mu_{P^3}|E$. For every $p^1 \in P^1$, there exists $p^2 \in P^2$, $p^3 \in P^3$ and $p^4 \in P^4$ such that $\lambda p^3|E = p^1|E$, $\lambda p^4|E = p^2|E$ and $\Phi(p^3|E, \mu_{P^3}|E) = \Phi(p^4|E, \mu_{P^4}|E)$. Since $\mu_{P^3}|E = \mu_{P^4}|E$, we have $\Phi(\lambda p^3|E, \lambda\mu_{P^3}|E) = \Phi(\lambda p^4|E, \lambda\mu_{P^4}|E)$, i.e., $\Phi(p^1|E, \mu_{P^1}|E) = \Phi(p^2|E, \mu_{P^2}|E)$. Thus, $Q^c(P^1, E) \subseteq Q^c(P^2, E)$. By a similar argument, we have $Q^c(P^2, E) \subseteq Q^c(P^1, E)$, and thus $Q^c(P^1, E) = Q^c(P^2, E)$. $\square$

Lemma 7. *For all $\succsim \in \mathscr{R}$, sequence $(\succsim^n)_{n=1}^{+\infty}$ in $\mathscr{R}$, $P \in \mathscr{P}$, sequence $(P^n)_{n=1}^{+\infty}$ in $\mathscr{P}$, and event E such that P represents $\succsim$ and P^n represents $\succsim^n$ for all $n \in \mathbb{N}_+$, $(\succsim^n)_{n=1}^{+\infty}$*

converges to $\succsim$ on E if and only if $(P^n_\Pi)_{n=1}^{+\infty}$ converges to P_Π.[29]

Proof. Note that the function $u^\downarrow(fEx; P_\Pi)$ is continuous with respect to P_Π. Thus, if $(P^n_\Pi)_{n=1}^{+\infty}$ converges to P_Π, then for all $f \in \mathcal{F}$ and $x \in X$, $\lim_{n\to+\infty} u^\downarrow(fEx; P^n) = \lim_{n\to+\infty} u^\downarrow(fEx; P^n_\Pi) = u^\downarrow(fEx; P_\Pi) = u^\downarrow(fEx; P)$. Thus, for all $y \in X$, $fEx \succsim^n y$ for all $n \in \mathbb{N}_+$ implies $fEx \succsim y$, and $y \succsim^n fEx$ for all $n \in \mathbb{N}_+$ implies $y \succsim fEx$. That is, $(\succsim^n)_{n=1}^{+\infty}$ converges to $\succsim$ on E. For the inverse direction, suppose to the contrary that $(P^n_\Pi)_{n=1}^{+\infty}$ does not converge to P_Π, then there is a subsequence $(P^{n_k}_\Pi)_{k=1}^{+\infty}$ that converges to $\bar{P}_\Pi \neq P_\Pi$. It follows that there exists $f \in \mathcal{F}$ and $x, y \in X$ such that $u^\downarrow(fEx; P_\Pi) > u(y) > u^\downarrow(fEx; \bar{P}_\Pi)$, which is a contradiction. $\square$

Lemma 8. *For all $P \in \mathscr{P}$, sequence $(P^n)_{n=1}^{+\infty}$ in $\mathscr{P}$, and $E \in \mathcal{S}_P \cap (\cap_{n=1}^{+\infty} \mathcal{S}_{P^n})$, if $(P^n_\Pi)_{n=1}^{+\infty}$ converges to P_Π, then $(Q^c(P^n, E))_{n=1}^{+\infty}$ converges to $Q^c(P, E)$.*

Proof. We consider three cases. In case 1, $\mu_P(E) < 1$. Since $(P^n_\Pi)_{n=1}^{+\infty}$ converges to P_Π, $(P^n|E)_{n=1}^{+\infty}$ converges to $P|E$. Thus, $(\overline{\mu_{P^n}|E})_{n=1}^{+\infty}$ converges to $\overline{\mu_P|E}$. WLOG, we can assume that for every $n \in \mathbb{N}_+$, $\mu_{P^n}(E) < 1$, and we are done. In case 2, $\mu_P(E) > 1$. WLOG, we can assume that for every $n \in \mathbb{N}_+$, $\mu_{P^n}(E) > 1$. In this case, $(P^n|E \cup \{\mu_{P^n}|E\})_{n=1}^{+\infty}$ converges to $P|E \cup \{\mu_P|E\}$. It follows that $(co(P^n|E \cup \{\mu_{P^n}|E\}))_{n=1}^{\infty}$ converges to $co(P|E \cup \{\mu_P|E\})$. Let G denote the set of probability distributions over E, and we have that $(co(P^n|E \cup \{\mu_{P^n}|E\}) \cap G)_{n=1}^{+\infty}$ (i.e., $(Q^c(P^n, E))_{n=1}^{+\infty}$) converges to $co(P|E \cup \{\mu_P|E\}) \cap G$ (i.e., $Q^c(P, E)$). In case 3, $\mu_P(E) = 1$. We divide the sequence $(P^n)_{n=1}^{+\infty}$ to two subsequences $(P^{n_k})_{k=1}^{+\infty}$ and $(P^{m_k})_{k=1}^{+\infty}$ such that for all $k \in \mathbb{N}_+$, $\mu_{P^{n_k}}(E) \le 1 < \mu_{P^{m_k}}(E)$. It follows from the previous arguments that $(\mu_{P^{n_k}}|E)_{k=1}^{+\infty}$ converges to $\mu_P|E$ and $(co(P^{m_k}|E \cup \{\mu_{P^{m_k}}|E\}) \cap G)_{k=1}^{+\infty}$ converges to $co(P|E \cup \{\mu_P|E\}) \cap G$. Note that $Q^c(P, E) = \{\mu_P|E\} = co(P|E \cup \{\mu_P|E\}) \cap G$. Therefore, both sequences converge to $Q^c(P, E)$ in this case. $\square$

Lemma 9. *The contraction rule satisfies Alignment Consistency.*

Proof. Consider arbitrary $\{\succsim^k\}_{k=1}^3 \subseteq \mathscr{R}$ and event E that satisfy all the primitive conditions stated in the axiom. For every $k \in \{1, 2, 3\}$, let P^k represents $\succsim^k$. Since $\succsim^3$ is E-aligned with $(\succsim^1, \succsim^2)$, by Lemma 3, P^3 is E-aligned with (P^1, P^2). Since for every $k \in \{1, 2\}$, $\succsim^k_E$ is ambiguous, by Lemma 1, we have for every $k \in \{1, 2\}$, $\mu_{P^k}(E) > 1$. Since $\succsim^1_E = \succsim^2_E$, we have $Q^c(P^1, E) = Q^c(P^2, E)$. By Lemma 4, we have $Q^c(P^1, E) = Q^c(P^2, E) = Q^3(P^3, E)$, which implies $\succsim^1_E = \succsim^2_E = \succsim^3_E$. $\square$

Lemma 10. *The contraction rule satisfies Sensitivity Congruence.*

[29] For all $P \in \mathscr{P}$ and event E, P_Π can be viewed as a subset of $\mathbb{R}^{|E|+1}$. Let d be the Euclidean metric on $\mathbb{R}^{|E|+1}$. The metric we use to measure the distance between two sets is the Hausdorff metric, denoted by d_h, which is defined such that for any two closed sets A and B in $\mathbb{R}^{|E|+1}$, $d_h(A, B) = \max\{\max_{a\in A} \min_{b\in B} d(a, b), \max_{a'\in B} \min_{b'\in A} d(a', b')\}$.

Proof. Consider arbitrary $\{\succsim^k\}_{k=1}^4 \subseteq \mathscr{R}$, event E, and $\lambda \in [1, +\infty)$ that satisfy all the primitive conditions stated in the axiom. For every $k \in \{1,2,3,4\}$, let P^k represent $\succsim^k$. Since $\succsim^3 \overset{(\lambda,E)}{\rightsquigarrow} \succsim^1$ and $\succsim^4 \overset{(\lambda,E)}{\rightsquigarrow} \succsim^2$, by Lemma 5, we have $P^1|E = \lambda P^3|E$ and $P^2|E = \lambda P^4|E$. Since $\succsim_E^k$ is ambiguous, by Lemma 1, we have $\mu_{P^k}(E) > 1$ for every k. Since $Q^c(P^3, E) = Q^c(P^4, E)$ (which is implied by $\succsim_E^3 = \succsim_E^4$), by Lemma 6, we have $Q^c(P^1, E) = Q^c(P^2, E)$, i.e., $\succsim_E^1 = \succsim_E^2$. □

Lemma 11. *The contraction rule satisfies Sensitivity Independence.*

Proof. Consider arbitrary $\succsim^1, \succsim^2 \in \mathscr{R}$, event E, and $\lambda \in [1, +\infty)$ that satisfy all the primitive conditions stated in the axiom. Let P^1 and P^2 represent the two preferences respectively. Since $\succsim^1 \overset{(\lambda,E)}{\rightsquigarrow} \succsim^2$, by Lemma 5, we have $P^2|E = \lambda P^1|E$. It follows that $\lambda \mu_{P^1}|E = \mu_{P^2}|E$, which implies $\overline{\mu_{P^1}|E} = \overline{\mu_{P^2}|E}$. Since both $\succsim_E^1$ and $\succsim_E^2$ are unambiguous, they are represented by $\{\overline{\mu_{P^1}|E}\}$ and $\{\overline{\mu_{P^2}|E}\}$ respectively. Therefore, $\succsim_E^1 = \succsim_E^2$. □

Lemma 12. *The contraction rule satisfies Non-Ambiguity Persistence*

Proof. Consider arbitrary $\succsim \in \mathscr{R}$, partition $\Omega = \{S_i\}_{i=1}^n$ and event E that satisfy all the conditions stated in the axiom. Let P represent $\succsim$. Since $\succsim$ is Ω-unambiguous, by Lemma 1, P_Ω is a singleton. Thus, $\mu_P(E) \leq \sum_{i=1}^n (\max_{p \in P} p(S_i)) = 1$. It follows that $Q^c(P, E) = \{\overline{\mu_P|E}\}$, i.e., $\succsim_E$ is unambiguous. □

Lemma 13. *The contraction rule satisfies Increased Sensitivity after Updating.*

Proof. Consider arbitrary preference $\succsim \in \mathscr{R}$, event E, $s \in E$, $x, y \in X$, and $f, g \in \mathcal{F}$ that satisfy all the primitive conditions stated in the axiom. Let P represent $\succsim$ and $p^* = \overline{\mu_P|E}$. Since $\succsim_E$ is unambiguous, $\{p^*\}$ represents $\succsim_E$. Since $f \sim x$, $f \sim_E x$, and $g \sim y$, to show $g \succsim_E y$, it suffices to show that $u(g; p^*) - u(f; p^*) \geq u^\downarrow(g; P) - u^\downarrow(f; P)$. Since $f \overset{S \setminus s}{=} g$, we have $u(g; p^*) - u(f; p^*) = p^*(s)[u(g(s)) - u(f(s))] \geq \mu_P(s)[u(g(s)) - u(f(s))] \geq p(s)[u(g(s)) - u(f(s))] \geq u^\downarrow(g; P) - u^\downarrow(f; P)$, where $p \in P$ satisfies $u^\downarrow(f; P) = u(f; p)$. □

Lemma 14. *The contraction rule satisfies Continuity.*

Proof. The lemma follows directly from Lemmas 7 and 8. □

Definition 7. *An updating rule Γ satisfies Independence of Irrelevant States (IIS) if for all $\succsim^1, \succsim^2 \in \mathscr{R}$ and $E \in \mathcal{S}(\succsim^1) \cap \mathcal{S}(\succsim^2)$, if $\succsim^1 \overset{(1,E)}{\rightsquigarrow} \succsim^2$, then $\succsim_E^1 = \succsim_E^2$.*

Lemma 15. *Consider an updating rule that satisfies IIS. With this rule, for all $\succsim \in \mathscr{R}$, $P \in \mathscr{P}$, and event E such that P represents $\succsim$ and $\mu_P(E) > 1$, there exists $\succsim^* \in \mathscr{R}$ that satisfies $\succsim \overset{(1,E)}{\rightsquigarrow} \succsim^*$ and violates updating monotonicity on E.*

Proof. Since E is finite and S is infinite, there exists finite $\hat{E} \subseteq S$ that satisfies $|E| = |\hat{E}|$ and $E \cap \hat{E} = \emptyset$. Consider an arbitrary bijection $\tau : E \to \hat{E}$ and denote $\hat{s} = \tau(s)$ for every $s \in E$. For every $p \in P$, define $\hat{p}$ such that for all $s \in E$, $\hat{p}(s) = p(s)$, and for all $\hat{s} \in \hat{E}$, $\hat{p}(\hat{s}) = \Phi(p|E, \mu_P|E)(s) - p(s)$. 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 $P_\Pi = \hat{P}_\Pi$ and for every $\hat{p} \in \hat{P}$, $\hat{p}(E \cup \hat{E}) = 1$. Since $\hat{P}$ is closed, we have $co(\hat{P}) \in \mathscr{P}$. Let $\succsim^* \in \mathscr{R}$ be represented by $co(\hat{P})$. Since $\hat{P}_\Pi = P_\Pi$, we have $\succsim \overset{(1,E)}{\rightsquigarrow} \succsim^*$. Since Γ satisfies IIS, $\succsim$ and $\succsim^*$ are updated to the same preference, denoted by $\succsim_E$, when E occurs. Let $\succsim_E$ be represented by $Q \in \mathscr{P}$. It remains to show that $\succsim^*$ violates updating monotonicity on E.

Suppose to the contrary that $\succsim^*$ satisfies updating monotonicity on E. Fix some $s \in E$. Consider act f such that $f(s) = x$, $f(\hat{s}) = y$, and for all $\tilde{s} \in S \backslash \{s, \hat{s}\}$, $f(\tilde{s}) = x$, where $u(y) > u(x)$. WLOG, assume $u(x) = 0$ and $u(y) = 1$. Clearly, $f \sim_E x$, since f equals x on E. We argue that $f \sim^* x$. Note that $u^\downarrow(f; co(\hat{P})) = \min_{\hat{p} \in \hat{P}} \hat{p}(\hat{s}) = 0$, where the second equality holds since there exists $p \in P$ such that $p(s) = \mu_P(s)$, implying $\hat{p}(\hat{s}) = \Phi(p|E, \mu_P|E)(s) - p(s) = 0$. Define $\hat{P}^* \subseteq \hat{P}$ such that $\hat{p} \in \hat{P}^*$ if and only if $p(s) = \mu_P(s)$, and we have $\arg\min_{\tilde{p} \in co(\hat{P})} \tilde{p}(\hat{s}) = co(\hat{P}^*)$. For every $\epsilon \in (0, 1)$, define act f^ϵ such that $f^\epsilon(s) = \epsilon y + (1 - \epsilon)x$, $f^\epsilon(\hat{s}) = y$, and for all $\tilde{s} \in S \backslash \{s, \hat{s}\}$, $f^\epsilon(\tilde{s}) = x$. It follows that when ϵ is sufficiently small,

$$u^\downarrow(f^\epsilon; co(\hat{P})) - u^\downarrow(f; co(\hat{P})) = \min_{\tilde{p} \in co(\hat{P}^*)} (u(f^\epsilon; \tilde{p}) - u(f; \tilde{p})) + o(\epsilon),$$

where $o(\epsilon)$ denotes an infinitesimal term of ϵ. The equation holds since the maxmin representation is essentially a concave function over $(u(X))^m$ (where m is the cardinality of the support of $co(\hat{P})$), and $co(\hat{P}^*)$ is the set of supergradients of the function at f. Since for each $\tilde{p} \in co(\hat{P}^*)$, $\tilde{p}(s) = \mu_P(s)$, we can re-write the above equation as

$$u^\downarrow(f^\epsilon; co(\hat{P})) - u^\downarrow(f; co(\hat{P})) = \mu_P(s)\epsilon + o(\epsilon).$$

Hence, for any $k < \mu_P(s)$, we can find small enough ϵ such that $u^\downarrow(f^\epsilon; co(\hat{P})) - u^\downarrow(f; co(\hat{P})) > k\epsilon$. Since $\succsim_E$ is represented by Q, which is supported in E, we have $u^\downarrow(f^\epsilon; Q) - u^\downarrow(f; Q) = \min_{q \in Q} q(s)\epsilon$. Since $f \sim^* x$ and $f \sim_E x$, to ensure that $\succsim^*$ satisfies updating monotonicity on E, we need to ensure that for all ϵ close to 0 and $z \in X$, $f^\epsilon \sim^* z$ implies $f^\epsilon \succsim_E z$, i.e., for all $k < \mu_P(s)$, $\min_{q \in Q} q(s)\epsilon > k\epsilon$. Thus, $\min_{q \in Q} q(s) \geq \mu_P(s)$. Similarly, we can show that for all $\tilde{s} \in E$, $\min_{q \in Q} q(\tilde{s}) \geq \mu_P(\tilde{s})$, which is impossible since $\mu_P(E) > 1$. $\square$

Lemma 16. *Consider an updating rule that satisfies IIS and Non-Ambiguity Persistence. With this rule, for all $\succsim \in \mathscr{R}$, $P \in \mathscr{P}$, and event E such that P represents $\succsim$, if $\mu_P(E) \leq 1$, then $\succsim_E$ is unambiguous.*

Proof. Since $\mu_P(E) \leq 1$, there exists $p^* \in \Delta(S)$ such that $p^*(E) = 1$ and for every $s \in E$, $p^*(s) \geq \mu_P(s)$. Fix such p^*. 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 bijection $\tau : E \to \hat{E}$ and denote $\hat{s} = \tau(s)$ for each $s \in E$. For each $p \in P$, define $\hat{p} \in \Delta(S)$ such that for every $s \in E$, $\hat{p}(s) = p(s)$ and $\hat{p}(\hat{s}) = p^*(s) - p(s)$. Let $\hat{P} = \{\hat{p}\}_{p \in P}$, and we have $\hat{P} \in \mathscr{P}$ and $\hat{P}_\Pi = P_\Pi$. Let $\succsim^* \in \mathscr{R}$ be represented by $\hat{P}$. For every $s \in E$ and every $\hat{p} \in \hat{P}$, we have $\hat{p}(\{s, \hat{s}\}) = p^*(s)$. Thus, $\succsim^*$ is Ω-unambiguous, where $\Omega = \{\{s, \hat{s}\}\}_{s \in E} \cup \{S \backslash (\cup_{s \in E}\{s, \hat{s}\})\}$. By Non-Ambiguity Persistence, $\succsim^*_E$ is unambiguous, and so is $\succsim_E$ by IIS. $\square$

Lemma 17. *Consider an updating rule that satisfies Sensitivity Independence, Non-Ambiguity Persistence, and Increased Sensitivity after Updating. With this rule, for all* $\succsim \in \mathscr{R}$, $E \in \mathcal{S}(\succsim)$, *and* $P \in \mathscr{P}$ *such that* P *represents* $\succsim$, $\mu_P(E) \leq 1$ *is sufficient and necessary for* $\succsim_E$ *to be unambiguous.*

Proof. By a similar argument to the proof of Lemma 16, Sensitivity Independence and Non-Ambiguity Persistence ensure that if $\mu_P(E) \leq 1$, then $\succsim_E$ is unambiguous. It remains to prove that if $\mu_P(E) > 1$, then $\succsim_E$ is ambiguous. Suppose to the contrary that $\mu_P(E) > 1$, and $\succsim_E$ is unambiguous. By Sensitivity Independence, for all $\succsim^* \in \mathscr{R}$ with $\succsim \overset{(1,E)}{\rightsquigarrow} \succsim^*$, we have $\succsim^*_E = \succsim_E$, and thus $\succsim^*_E$ is unambiguous. This contradicts with Increased Sensitivity after Updating by the proof of Lemma 15. $\square$

Lemma 18. *Consider an updating rule that satisfies Sensitivity Independence, Non-Ambiguity Persistence, and Increased Sensitivity after Updating. With this rule, for all* $\succsim \in \mathscr{R}$, $E \in \mathcal{S}(\succsim)$, *and* $P \in \mathscr{P}$ *such that* P *represents* $\succsim$, *if* $\mu_P(E) \leq 1$, *then* $\succsim_E$ *is represented by* $\{\overline{\mu_P|E}\}$.

Proof. Let $\lambda = \frac{1}{\mu_P(E)}$. Since $\mu_P(E) \leq 1$, we have $\lambda \geq 1$. Let $\hat{E} \subseteq S$ be such that $|E| = |\hat{E}|$ and $E \cap \hat{E} = \emptyset$. Consider an arbitrary bijection $\tau : E \to \hat{E}$ and let $\hat{s} = \tau(s)$ for every $s \in E$. For every $p \in P$, define $\hat{p} \in \Delta(S)$ such that for every $s \in E$, $\hat{p}(s) = \lambda p(s)$, and for every $\hat{s} \in \hat{E}$, $\hat{p}(\hat{s}) = \lambda\mu_P(s) - \lambda p(s)$. Let $\hat{P} = \{\hat{p}\}_{p \in P}$, and we have $\hat{P} \in \mathscr{P}$ and $\hat{P}|E = \lambda P|E$. Let $\succsim^* \in \mathscr{R}$ be represented by $\hat{P}$, and it follows from Lemma 5 that $\succsim \overset{(\lambda,E)}{\rightsquigarrow} \succsim^*$. Since $\mu_P(E) \leq \mu_{\hat{P}}(E) = 1$, by Lemma 17, both $\succsim_E$ and $\succsim^*_E$ are unambiguous. By Sensitivity Independence, we have $\succsim_E = \succsim^*_E$. Let $\succsim^*_E$ be represented by $\{q\}$ for some $q \in \Delta(S)$.

It remains to show that $q = \overline{\mu_P|E}$. Fix some $s \in E$. Consider acts f and g such that $f(s) = x$, $f(\hat{s}) = y$, $g(s) = \frac{1}{2}x + \frac{1}{2}y$, $g(\hat{s}) = y$, and for all $\tilde{s} \in S \backslash \{s, \hat{s}\}$, $f(\tilde{s}) = g(\tilde{s}) = x$, where $u(y) > u(x)$. We can assume WLOG that $u(x) = 0$ and $u(y) = 1$. Since for all $\hat{p} \in \hat{P}$, $\hat{p}(\{s, \hat{s}\}) = \lambda\mu_P(s)$, it follows that

$$u^{\downarrow}(g; \hat{P}) = \min_{\hat{p} \in \hat{P}} \left(\frac{1}{2}\hat{p}(s) + \hat{p}(\hat{s})\right) = \min_{\hat{p} \in \hat{P}} \left(\frac{1}{2}\hat{p}(s) + \lambda\mu_P(s) - \hat{p}(s)\right) = \frac{1}{2}\lambda\mu_P(s).$$

Similarly, $u^{\downarrow}(f;\hat{P}) = 0$. Also, we have $u(f;q) = 0$ and $u(g;q) = \frac{1}{2}q(s)$. Since $\succsim_E^*$ is unambiguous, Increased Sensitivity after Updating implies $u(g;q) \geq u^{\downarrow}(g;\hat{P})$. That is, $\lambda\mu_P(s) \leq q(s)$. By a similar argument, for every $\tilde{s} \in E$, we have $\lambda\mu_P(\tilde{s}) \leq q(\tilde{s})$. Since $\lambda\mu_P|E = \overline{\mu_P|E}$, we conclude that $q = \overline{\mu_P|E}$. □

Lemma 19. *Consider an updating rule that satisfies Axioms 1-6. With this rule, for all $\succsim \in \mathscr{R}$, $E \in \mathcal{S}(\succsim)$, and $P \in \mathscr{P}$ such that P represents $\succsim$, if $\mu_P(E) > 1$, then $\succsim_E$ is represented by $\{\Phi(p|E, \mu_P|E)\}_{p\in P}$.*

Proof. Let $\lambda = \mu_P(E) > 1$. Consider a sequence of real numbers $(\lambda_n)_{n=1}^{+\infty}$ such that $(\lambda_n)_{n=1}^{+\infty}$ converges to λ, and for all $n \in \mathbb{N}_+$, $1 < \lambda_n < \lambda$. For every $n \in \mathbb{N}_+$, let $P_n \in \mathscr{P}$ be such that $\lambda_n P_n|E = P|E$, $\succsim^n \in \mathscr{R}$ be represented by P_n, $Q_n \in \mathscr{P}$ represent $\succsim_E^n$, $R_n \in \mathscr{P}$ be such that $R_n = \{p \in co(P_n \cup Q_n) : p(E) = \frac{1}{\lambda_n}\}$, $\trianglerighteq^n \in \mathscr{R}$ be represented by R_n, and $\trianglerighteq^{*,n} \in \mathscr{R}$ be represented by $\lambda_n R_n|E$.

By our construction, for all $n \in \mathbb{N}_+$, we have $\mu_{P_n}(E) > 1$. By Lemma 17, $\succsim_E^n$ is ambiguous, and thus $\mu_{Q_n}(E) > 1$. By our construction, R_n is E-aligned with (P_n, Q_n). It follows from Lemma 3 and Alignment Consistency that $\succsim_E^n = \trianglerighteq_E^n$. By Lemma 5, we have $\succsim^n \overset{(\lambda_n,E)}{\rightsquigarrow} \succsim$ and $\trianglerighteq^n \overset{(\lambda_n,E)}{\rightsquigarrow} \trianglerighteq^{*,n}$. Thus, by Sensitivity Congruence, we have $\succsim_E = \trianglerighteq_E^{*,n}$.

Let $P^* \in \mathscr{P}$ satisfy $\lambda P^*|E = P|E$, and $\succsim^* \in \mathscr{R}$ be represented by P^*. Since $\mu_{P^*}(E) = 1$, by Lemma 18, $\succsim_E^*$ is represented by $\{q\}$, where $q = \mu_{P^*}|E$. Since $(P_n|E)_{n=1}^{+\infty}$ converges to $P^*|E$, by Lemma 7, $(\succsim^n)_{n=1}^{+\infty}$ converges to $\succsim^*$ on E. By Continuity, $(\succsim_E^n)_{n=1}^{+\infty}$ converges to $\succsim_E^*$ on E. Therefore, by Lemma 7, $(Q_n)_{n=1}^{+\infty}$ converges to $\{q\}$. It follows that $(\lambda_n R^n|E)_{n=1}^{+\infty}$ converges to $\{\Phi(p|E, \mu_P|E)\}_{p\in P}$. Let $\trianglerighteq^* \in \mathscr{R}$ be represented by $\{\Phi(p|E, \mu_P|E)\}_{p\in P}$. By Lemma 7, $(\trianglerighteq^{*,n})_{n=1}^{+\infty}$ converges to $\trianglerighteq^*$ on E. Since $S\backslash E$ is $\trianglerighteq^*$-null, we have $\trianglerighteq_E^* = \trianglerighteq^*$. By Continuity, $(\trianglerighteq_E^{*,n})_{n=1}^{+\infty}$ converges to $\trianglerighteq^*$ on E. Since for all $n \in \mathbb{N}_+$, $\succsim_E^{*,n} = \succsim_E$, we conclude that $\succsim_E = \trianglerighteq^*$. □

The necessity of Theorem 1 is demonstrated by Lemmas 9-13. Proposition 4 can be implied by Lemmas 15 and 16. The sufficiency of Theorem 1 is shown by Lemmas 18 and 19. □

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 $\max_{p\in P} p(\{d_1\} \times \Theta)$. Given 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 $[1 - \max_{p\in P} p(d_2, \theta), \max_{p\in P} p(d_1, \theta)]$. Clearly, this set does not strictly contain the prior 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. We consider the non-trivial case in which D and Θ are not singleton sets. 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 the contraction posterior set is a singleton, then there is no dilation. If not, $\max_{p\in P} p(\{d\}\times\Theta) > \mu_P(d,\theta)$ implies that the ex-ante maximal probability of d is strictly higher than the ex-post one. Thus, the ex-post set of beliefs over D does not contain the ex-ante one. □

Proof of Proposition 7. The DM's evaluations of f and g are given by the minimal probabilities of states d_1 and d_2, respectively. First, consider scenario 1. The maximal probability of $(d_1,\hat{d}_1)$ is $\overline{\kappa}H$, and that of $(d_2,\hat{d}_1)$ is $(1-\underline{\kappa})(1-L)$. 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 $\overline{\kappa}H+(1-\underline{\kappa})(1-L)$ is less than one or not. If $\overline{\kappa}H+(1-\underline{\kappa})(1-L)\le 1$, then we have $v_f^a=\frac{\overline{\kappa}H}{\overline{\kappa}H+(1-\underline{\kappa})(1-L)}$ and $v_g^a=\frac{(1-\underline{\kappa})(1-L)}{\overline{\kappa}H+(1-\underline{\kappa})(1-L)}$. If $\overline{\kappa}H+(1-\underline{\kappa})(1-L)>1$, then we have $v_f^a=1-(1-\underline{\kappa})(1-L)$ and $v_g^a=1-\overline{\kappa}H$. Next, consider scenario 2. With signal $\hat{d}_1$, the maximal probability of $(d_1,\hat{d}_1)$ is $\overline{\kappa}(H+L)/2$, and that of $(d_2,\hat{d}_1)$ is $(1-\underline{\kappa})(2-H-L)/2$. Since $\overline{\kappa}(H+L)/2+(1-\underline{\kappa})(2-H-L)/2\le 1$, the realization of $\hat{d}_1$ resolves the ambiguity. Thus, we have $v_f^u=\frac{\overline{\kappa}(H+L)}{\overline{\kappa}(H+L)+(1-\underline{\kappa})(2-H-L)}$ and $v_g^u=\frac{(1-\underline{\kappa})(2-H-L)}{\overline{\kappa}(H+L)+(1-\underline{\kappa})(2-H-L)}$. Since $H+L>1$, we have $\frac{H+L}{2-H-L}>\frac{H}{1-L}$. Thus, when $\overline{\kappa}H+(1-\underline{\kappa})(1-L)\le 1$, $v_f^a=\frac{\overline{\kappa}H}{\overline{\kappa}H+(1-\underline{\kappa})(1-L)}<\frac{\overline{\kappa}(H+L)}{\overline{\kappa}(H+L)+(1-\underline{\kappa})(2-H-L)}=v_f^u$. When $\overline{\kappa}H+(1-\underline{\kappa})(1-L)>1$, we have $v_f^a=1-(1-\underline{\kappa})(1-L)\le\frac{\overline{\kappa}H}{\overline{\kappa}H+(1-\underline{\kappa})(1-L)}<\frac{\overline{\kappa}(H+L)}{\overline{\kappa}(H+L)+(1-\underline{\kappa})(2-H-L)}=v_f^u$, where the first inequality holds since $\frac{v_f^a}{1-v_f^a}=\frac{1-(1-\underline{\kappa})(1-L)}{(1-\underline{\kappa})(1-L)}\le\frac{\overline{\kappa}H}{(1-\underline{\kappa})(1-L)}$. The comparison between v_g^a and v_g^u follows from similar calculations. □

Proof of Theorem 2. Consider any $\tau: D\to A_v^*$ and $\lambda\in(0,1)$. For every $d\in D$, let $m(d)=\max_{r\in R} r(d)$. By Assumption 1, there exists $r_d\in\Delta^o(S)$ such that $\tau(d)$ is one of the v-optimal actions under r_d, i.e., for all $a\in A$, $\sum_{\hat{d}\in D} r_d(\hat{d})v(\tau(d),\hat{d})\ge\sum_{\hat{d}\in D} r_d(\hat{d})v(a,\hat{d})$. Let $\Theta=\{\theta_{d,t}\}_{(d,t)\in D\times N}$, where N is an index set and $|N|=n\in\mathbb{N}_+$ with n to be specified later. Define an information structure $\mathcal{I}=\left(\Theta,\{\chi_{d,t}\}_{(d,t)\in D\times N}\right)$ as follows: For every $(d^*,t^*)\in D\times N$, let χ_{d^*,t^*} be such that

(i) for all $d,\hat{d}\in D$, $\chi_{d^*,t^*}(\theta_{d^*,t^*}|d)<1-\lambda$ and

$$\frac{m(d)\chi_{d^*,t^*}(\theta_{d^*,t^*}|d)}{m(\hat{d})\chi_{d^*,t^*}(\theta_{d^*,t^*}|\hat{d})}=\frac{r_{d^*}(d)}{r_{d^*}(\hat{d})},$$

(ii) for all $\hat{d}\in D$ and $(d,t)\in D\times N$ with $(d,t)\neq(d^*,t^*)$,

$$\begin{aligned}
&\chi_{d^*,t^*}(\theta_{d,t}|\hat{d})=0,\ \text{if } d\neq\hat{d};\\
&\chi_{d^*,t^*}(\theta_{d,t}|\hat{d})=\frac{1-\chi_{d^*,t^*}(\theta_{d^*,t^*}|\hat{d})}{n-1},\ \text{if } d=\hat{d}=d^*,\\
&\chi_{d^*,t^*}(\theta_{d,t}|\hat{d})=\frac{1-\chi_{d^*,t^*}(\theta_{d^*,t^*}|\hat{d})}{n},\ \text{if } d=\hat{d}\neq d^*.
\end{aligned}$$

Given (i) and (ii), we can additionally require $\chi_{d^*,t^*}(\cdot|\cdot)$ to satisfy

$$\text{(iii)} \sum_{d\in D} \chi_{d^*,t^*}(\theta_{d^*,t^*}|d) = \frac{1-\lambda}{2}.$$

If conditions (i)-(iii) hold for all $(d^*, t^*) \in D \times N$ and n becomes large enough, then for all $d, d', \hat{d} \in D$ and $t, t' \in N$ with $(d, t) \neq (d', t')$, we have $\chi_{d,t}(\theta_{d,t}|\hat{d}) > \chi_{d',t'}(\theta_{d,t}|\hat{d})$. Therefore, when signal $\theta_{d,t}$ is realized, by condition (iii), the contraction posterior set of the DM is a singleton and determined by the normalization of $(m(\hat{d})\chi_{d,t}(\theta_{d,t}|\hat{d}))_{\hat{d}\in D}$, and by condition (i), the contraction posterior set is $\{r_d\}$. It then follows that the function $\zeta\colon \Theta \to A_v^*$ is $\mathcal{I}$-implementable, where ζ satisfies that for all $\theta_{d,t} \in \Theta$, $\zeta(\theta_{d,t}) = \tau(d)$. The theorem is then shown by the fact that for every $d, \hat{d} \in D$ and $t \in N$,

$$\sum_{(d',t')\in D\times N\colon \zeta(\theta_{d',t'})\neq\tau(\hat{d})} \chi_{d,t}(\theta_{d',t'}|\hat{d}) \leq \sum_{(d',t')\in D\times N: d'\neq\hat{d}} \chi_{d,t}(\theta_{d',t'}|\hat{d}) \leq \chi_{d,t}(\theta_{d,t}|\hat{d}) < 1-\lambda.$$

□

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.

Brooks, B., and S. Du (2021): "Optimal Auction Design with Common Values: An Informationally Robust Approach," *Econometrica*, 89(3), 1313–1360.

Carroll, G. (2019): "Robustness in Mechanism Design and Contracting," *Annual Review of Economics*, 11, 139–166.

Chandrasekher, M., M. Frick, R. Iijima, and Y. Le Yaouanq (2022): "Dual-Self Representations of Ambiguity Preferences," *Econometrica*, 90(3), 1029–1061.

Chen, J. Y. (2023): "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.

Chung, K.-S., and J. C. Ely (2007): “Foundations of Dominant-Strategy Mechanisms,” *The Review of Economic Studies*, 74(2), 447–476.

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.

Di Tillio, A., N. Kos, and M. Messner (2017): “The Design of Ambiguous Mechanisms,” *The Review of Economic Studies*, 84(1), 237–276.

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 (2024): “Hard-to-Interpret Signals,” *Journal of the European Economic Association*, 22(1), 393–427.

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.

Ke, S., B. Wu, and C. Zhao (2023): "Learning from a Black Box," *Working Paper*.

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 (2022): "Reacting to Ambiguous Messages: An Experimental Analysis," *Games and Economic Behavior*, 136, 360–378.

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

Klibanoff, P., M. Marinacci, and S. Mukerji (2005): "A Smooth Model of Decision Making under Ambiguity," *Econometrica*, 73(6), 1849–1892.

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

Kovach, M. (2023): "Ambiguity and partial Bayesian updating," *Economic Theory*, pp. 1–26.

Liang, Y. (forthcoming): "Learning from Unknown Information Sources," *Management Science.*

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.

Pearce, D. G. (1984): "Rationalizable Strategic Behavior and the Problem of Perfection," *Econometrica*, pp. 1029–1050.

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

Sadowski, P., and T. Sarver (2021): "An Evolutionary Perspective on Updating Risk and Ambiguity Preferences," Discussion paper, Technical report, Duke University.

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 (2023): "Ambiguous Information and Dilation: An experiment," *Journal of Economic Theory*, 208, 105610.

Tang, R., and M. Zhang (2021): "Maxmin Implementation," *Journal of Economic Theory*, 194, 105250.

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

Wolitzky, A. (2016): "Mechanism Design with Maxmin Agents: Theory and an Application to Bilateral Trade," *Theoretical Economics*, 11(3), 971–1004.

Zhao, C. (2022): "Pseudo-Bayesian Updating," *Theoretical Economics*, 17(1), 253–289.