Minimum Distance Belief Updating with General Information*

Adam Dominiak   Matthew Kovach   Gerelt Tserenjigmid

January 23, 2021
Current Version

Abstract: We study belief revision when information is given as a set of relevant probability distributions. This flexible setting encompasses (i) the standard notion of information as an event (a subset of the state space), (ii) qualitative information ("A is more likely than B"), (iii) interval information ("chance of A is between ten and twenty percent"), and more. In this setting, we behaviorally characterize a decision maker (DM) who selects a posterior belief from the provided information set that minimizes the subjective distance between her prior and the information. We call such a DM a Minimum Distance Subjective Expected Utility (MDSEU) maximizer. Next, we characterize the collection of MDSEU distance notions that coincide with Bayesian updating on standard events. We call this class of distances **Generalized Bayesian Divergence**, as they nest Kullback-Leibler Divergence. MDSEU provides a systematic way to extend Bayesian updating to general information and zero-probability events. Importantly, Bayesian updating is not unique. Thus, two Bayesian DM's with a common prior may disagree after common information, resulting in polarization and speculative trade. We discuss related models of non-Bayesian updating.

Keywords: Uncertainty, subjective expected utility, imprecise information, non-Bayesian updating, generalized Bayesian divergence.

JEL: D01, D81.

*Dominiak: Virginia Tech (dominiak@vt.edu); Kovach: Virginia Tech (mkovach@vt.edu); Tserenjigmid: Virginia Tech (gerelt@vt.edu). We are very grateful to David Freeman, Paolo Ghirardato, Faruk Gul, Edi Karni, and Yusufcan Masatlioglu for valuable comments and discussions, as well as the audience of WiERD 2020.

1 Introduction

How decision makers revise their beliefs after receiving information is a foundational problem in economics and game theory. While the benchmark model of Bayesian updating is broadly appealing for a variety of reasons, there is robust experimental evidence that people's beliefs systematically deviate from what Bayesian updating prescribes.[1] When information is qualitative or imprecise, there is a more fundamental issue at hand. To illustrate, consider the following statements:

(i) *The chance that the stock market will go down is at least 60%.*

(ii) *An urn with 100 balls contains between 30 and 50 green balls.*

(iii) *Candidate A is more likely to win than Candidate B.*

(iv) *Your heart disease risk is* 9% *if you do not smoke, and* 17% *if you do smoke.*[2]

Despite the fact that each of these is a natural statement in daily life, it is not clear how to apply Bayes' rule to such information.

These statements are examples of what we refer to as *general information*, and they fall outside the standard framework used to study belief updating. The standard framework to study updating represents information as an event, or a subset of some grand state space. For a set of states S, the "information" that a decision maker (henceforth, DM) receives is assumed to be given by the statement: "the event $E \subseteq S$ occurred." We interpret this as if the DM has learned that the true state is an element of E and that states outside of E are no longer possible. However, in many real-life circumstances, DMs may receive information in more general or nuanced forms. For instance, a doctor may ask a colleague for her opinion on the chances of a treatment's success. If this

[1]For instance, they may exhibit confirmation bias, the representativeness heuristic, under- or over-reaction, or a myriad of other biases (see Benjamin (2019) for a discussion).

[2]These estimates come from the Mayo Clinic Heart Disease Risk Calculator.

opinion is given as a specific probability estimate over the states, information in this example is a particular element of $\Delta(S)$.[3] Similarly, a political commentator may declare that he believes it is *more likely than not* that a bill passes through Congress, and so information here is a probability interval, a subset of $\Delta(S)$. Our setting allows us to model updating behavior under these richer forms of information.

To study these forms of information, we consider an extended Anscombe and Aumann (1963) framework in which preferences are conditional on general information. We define general information as a collection of possible probability measures over a given, payoff relevant state space (see Damiano (2006), Ahn (2008), Gajdos et al. (2008), Chambers and Hayashi (2010), Zhao (2020)). Importantly, this generalizes both the standard notion of an event and notions of qualitative information. Within this framework, we provide behavioral foundations for a "minimally rational" form of belief revision in response to this information. A DM who behaves in accordance with our behavioral postulates acts as if she selects a revised (or posterior) belief that minimizes the subjective distance between her prior and the provided information. We will refer to such a DM as a **Minimum Distance Subjective Expected Utility (MDSEU)** maximizer. Importantly, the class of MDSEU preferences allows for both Bayesian and non-Bayesian belief dynamics.

More formally, we assume that the DM forms a subjective belief μ, represented by a probability measure over the finite state space S, and uses it to evaluate acts (maps from states to consequences) via their expected utilities. This information is encoded in her initial preference $\succsim$. We then suppose that our DM receives information depicted by a set I of probability distributions over the states (i.e., I is a closed subset of $\Delta(S)$). She now has a preference over acts conditional on the received information $\succsim_I$.

As new information I emerges, a subjective expected utility (SEU) DM may revise

[3] $\Delta(S)$ denotes the set of all probability distributions over S.

her initial belief μ. In particular, she must decide whether to abandon μ and, if so, which new belief to form. Our DM solves this problem by selecting a new belief μ_I that is (i) consistent with I and (ii) closest to her initial belief. That is, her new belief μ_I is the element of I that is closest to μ among all of the probability measures in I (illustrated for two information sets in Figure 1). When her initial belief μ is consistent with I (i.e., $\mu \in I$), the closest probability measure is necessarily μ itself, and thus the DM retains her initial belief (i.e., $\mu = \mu_I$).

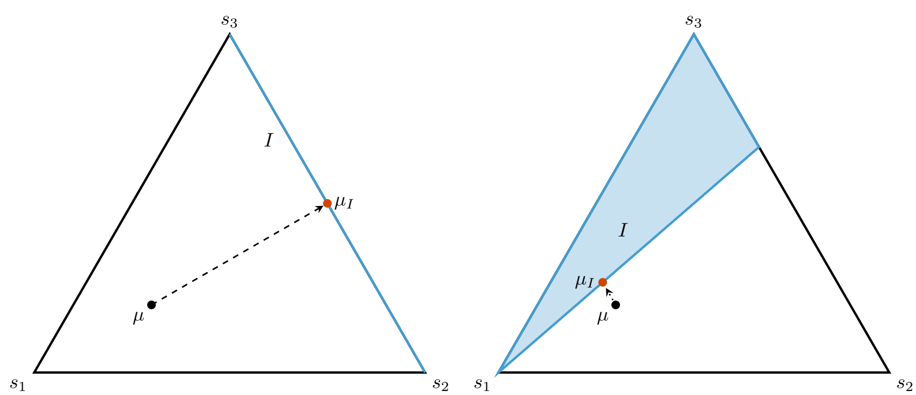

Figure 1: Information Sets and Minimum Distance Updating. The left panel illustrates "only s_2 and s_3 are possible," while the right panel illustrates "s_3 is at least twice as likely as s_2."

We provide a complete behavioral analysis of this form of belief revision. In addition to a few standard and technical postulates, we provide three novel axioms that are essential to our analysis. The axioms of **Compliance** and **Responsiveness** ensure that the DM accepts the information as truthful. Consequently, she selects a posterior belief that is consistent with I. Our other crucial axiom, **Informational Betweenness**, ensures that beliefs are consistent with minimization of some subjective distance notion.

One important feature of our framework is that it only requires a prior over the payoff-relevant states. This is in contrast to standard Bayesian models that assume the DM utilizes a joint prior over payoff states *and* possible signals. In many real-life settings, it is implausible to expect a decision maker to have *any* idea what types of information she will receive, and it is even less plausible to expect her to have constructed a unique belief about how this information correlates with payoff states. Thus, while there is a sense in which our model captures "bounded rationality," it allows for a well-defined notion of updating for statements that a Bayesian DM cannot handle.[4] Further, a model that only requires a simplified specification of beliefs is often more convenient to work with and a more plausible description of behavior.

While our notion of information is more general than the one typically used to study (Bayesian) updating, it allows for a natural analogue of the standard notion of an event. For any $E \subseteq S$, the information set $I^1_E = \{\pi \in \Delta(S) : \pi(E) = 1\}$ means "E has occurred" (the right-hand side of Figure 1 depicts $I^1_{\{s_2,s_3\}}$). Therefore, we can check for consistency of minimum distance updating with Bayesian updating on these "standard" events. We show that suitably adapting the classic axiom of Dynamic Consistency to our environment essentially characterizes distance functions that generate posteriors consistent with Bayesian updating on standard events. Our result shows that there is a family of such distance functions, which we refer to as **Generalized Bayesian Divergence** (Kullback-Leibler divergence is an example). Consequently, in our setting, there may be Bayesian disagreement. That is, "Bayesian" DMs may start with the same prior, receive the same information, and arrive at different posteriors. Therefore, our results show that the structure of information is crucial for the generation of polarization and disagreement.[5]

A well-known limitation of Bayesian updating (in standard settings) is that it is not defined for zero-probability events. Because our framework is much richer than the standard setting, MDSEU allows us to extend updating to zero-probability events,

[4]Indeed, as argued in Zhao (2020), even in the event of a prior over payoff states and signals (i.e., we expand to a product state space), we could still refer to general information in the expanded state space, which a Bayesian DM could not incorporate.

[5]Somewhat similarly, Baliga et al. (2013) show that polarization may arise under ambiguity (e.g., imprecise beliefs). We show that polarization may arise from imprecise information, and thus view our results as a partial complement.

thereby offering a complete theory of updating. We illustrate this explicitly by showing that MDSEU "nests" the Hypothesis Testing model (HT) of Ortoleva (2012). In the HT, an agent's behavior is in accord with SEU, yet she also has a second-order belief and thus has multiple beliefs in mind. She updates her prior according to Bayes' rule if she receives "expected" information. When information is "unexpected," she rejects her prior and uses her second-order belief to select a new belief according to a maximum likelihood rule. This suggests an interpretation of an essentially Bayesian agent who is nevertheless open to fundamentally shifting her worldview. The corresponding distance function in MDSEU is consistent with this interpretation: the agent uses a support-dependent distance that is "piece-wise Bayesian."

The remainder of this paper is structured as follows. In section 2, we introduce the formal framework and our notion of updating. We provide behavioral foundations of the MDSEU representation in section 3. We discuss Bayesian updating, disagreement, and zero-probability events in section 4. In section 5, we show how to recover a particular distance function from updating rules specified on classic events, and we apply this result to construct non-Bayesian distance notions. We close with a discussion of related work in section 6. The proofs are in the Appendix.

2 Model

2.1 Basic Setup

We study choice under uncertainty in the framework of Anscombe and Aumann (1963). A DM faces uncertainty described by a nonempty and finite set of states of nature $S = \{s_1, \ldots, s_n\}$. A nonempty subset E of S is called an event. Let X be a nonempty, finite set of outcomes and $\Delta(X)$ be the set of all lotteries over X, $\Delta(X) := \Big\{p : X \to [0, 1] \mid \sum_{x \in X} p(x) = 1\Big\}$.

We model the DM's preference over acts. An act is a mapping $f : S \to \Delta(X)$ that assigns a lottery to each state. Any act f that assigns the same lottery to all states ($f(s) = p$ for all $s \in S$) is called a constant act. Using a standard abuse of notation, we denote by $p \in F$ the corresponding constant act. Hence, we can identify the set of lotteries $\Delta(X)$ with the constant acts. The set of all acts is $F := \{f : S \to \Delta(X)\}$. A preference relation over F is denoted by $\succsim$. As usual, $\succ$ and $\sim$ are the asymmetric and symmetric parts of $\succsim$, respectively. We denote by $\Delta(S)$ the set of all probability distributions on S. For notational convenience, for each $\mu \in \Delta(S)$ and each $s_i \in S$, we will sometimes write μ_i in place of $\mu(s_i)$: the probability of state s_i according to μ. Finally, let $\|\cdot\|$ denote the Euclidean norm. For any set A and a function d on A, we write $\arg\min d(A) = \{x \in A \mid d(y) \geq d(x) \text{ for any } y \in A\}$ (whenever this is well-defined).

2.2 Information Sets

We consider an environment in which the DM receives new pieces of information about the uncertainty she faces (i.e., the states in S). Importantly, we explicitly develop a general information structure, defined below.

Definition 1. We call $I \subseteq \Delta(S)$ an **information set** if it is non-empty and closed. The collection of all information sets is denoted $\mathscr{I}$.

Initially, the DM faces this uncertainty about the states with "complete ignorance." We may think of this as the case where her initial information is the set of all probability distributions over S (i.e., $I = \Delta(S)$). Later, the DM receives "more precise" information in the form of $I \subset \Delta(S)$. In other words, the DM learns that some probability distributions are impossible. This is analogous to the standard setup, in which the DM is informed that certain states of nature are no longer possible (i.e., the DM is informed that $E \subset S$ has occurred). However, our notion of information is more general and nests the idea of an event. The information set containing all probability distributions concentrated on E, $I = \{\pi \in \Delta(S) \mid \pi(E) = 1\}$, is equivalent to learning that all states outside E are impossible. We will therefore refer to such information sets as an *info-event*. Our setting allows for information sets that capture richer statements, and we provide some examples below.

(i) For any $E \subset S$ and $\alpha \in [0,1]$, the information set "E occurs with probability α" is

$$I = \{\pi \in \Delta(S) \mid \pi(E) = \alpha\} = I_E^\alpha. \tag{1}$$

We will refer to such an information set as an *α-event*.[6] When $\alpha = 1$, this corresponds to an info-event.

(ii) For any probability distribution π, the information set "π is the true distribution" is

$$I = \{\pi\}. \tag{2}$$

(iii) For any $A, B \subseteq S$, the information set "A is at least δ-as likely as B" is

$$I = \{\pi \in \Delta(S) \mid \pi(A) \geq \delta\pi(B)\}. \tag{3}$$

Notice that for $\delta = 1$, this corresponds to the classic notion of qualitative information.

(iv) For $E \subset S$, and $0 < \alpha < \beta < 1$, the information set "the probability of E is

[6] For instance, in the 3-color Ellsberg experiment, subjects are informed that a ball will be drawn from an urn containing red, blue and green balls, and it is standard to assume that the probability of a red ball is $\frac{1}{3}$.

between α and β" is

$$I = \{\pi \in \Delta(S) \mid \alpha \leq \pi(E) \leq \beta\}. \tag{4}$$

2.3 Minimum Distance Belief Revision

The DM's behavior is depicted by a family $\{\succsim_I\}_{I \in \mathscr{I}}$ of preference relations, each defined over F. Before any information is revealed (i.e., $I = \Delta(S)$), we write $\succsim$ in place of $\succsim_{\Delta(S)}$, and we call $\succsim$ the initial preference. As a new piece of information $I \subset \Delta(S)$ emerges, the DM revises $\succsim$ given I. The new preference is denoted by $\succsim_I$ and governs the DM's conditional choice in light of I.

We assume that the DM's initial preference is of the SEU form. That is, the initial preference $\succsim$ admits a SEU representation with respect to an expected utility function $u : \Delta(X) \to \mathbb{R}$ and a (unique) probability distribution $\mu \in \Delta(S)$ such that for any $f, g \in F$,

$$f \succsim g \text{ if and only if } \sum_{s \in S} \mu(s) u\big(f(s)\big) \geq \sum_{s \in S} \mu(s) u\big(g(s)\big). \tag{5}$$

Hence, the DM's initial behavior is characterized by the pair (u, μ).

How does the DM incorporate I into her conditional choice? We assume that the DM updates her initial preference $\succsim$ by revising her initial belief μ while keeping her risk attitude unchanged. Let μ_I denote the DM's revised (updated) belief conditional upon I. How does the DM form her new belief μ_I when $\mu \notin I$, i.e., her old belief μ conflicts with the available information?

We impose two properties on the DM's belief revision. First, we assume that the DM reacts to and accepts the information, so that her new belief μ_I is consistent with I (i.e., $\mu_I \in I$). Second, we assume that she exhibits "inertia of initial beliefs." Therefore, she chooses the μ_I closest to her initial belief μ. That is, the DM forms her new belief μ_I as if she was minimizing the distance between her initial belief μ and all probability distributions consistent with I (see Figure 1). We call this updating procedure **minimum distance updating**, and μ_I the **minimum distance update** of μ. When her initial belief μ is consistent with I, the closest probability measure is μ, and thus the DM keeps it (i.e., $\mu = \mu_I$).

Putting these assumptions together, we consider a DM whose preference relation $\succsim_I$ admits a SEU representation with respect to the same expected utility function u and a new probability distribution $\mu_I \in I$ that is of minimal "distance" from the initial belief μ for each information $I \in \mathscr{I}$. We also require that the DM's notion of distance satisfy two intuitive distance properties, which we formally define below.

Definition 2 (Distance Function and Its Tie-Breaker)**.** A function $d : \Delta(S) \to \mathbb{R}$ is a **distance function** with respect to $\mu \in \Delta(S)$, denoted by d_μ, if,

(i) for any distinct $\pi, \pi' \in \Delta(S)$ with $d_\mu(\pi) = d_\mu(\pi')$, there is some $\alpha \in (0, 1)$ such that $d_\mu(\pi) > d_\mu(\alpha\pi + (1 - \alpha)\pi')$, and

(ii) $d_\mu(\mu) < d_\mu(\pi)$ for any $\pi \in \Delta \setminus \{\mu\}$.

Moreover, a function $\hat{d}_\mu : \Delta(S) \to \mathbb{R}$ is a **tie-breaker for** d_μ if (i) $\hat{d}_\mu$ is injective and (ii) for any $\pi, \pi' \in \Delta$, $d_\mu(\pi) > d_\mu(\pi')$ implies $\hat{d}_\mu(\pi) > \hat{d}_\mu(\pi')$.

Property (i) requires that if two beliefs are equidistant from μ, then there is a mixed belief that is strictly closer to μ. Property (ii) ensures that the current belief is unique, in that all different beliefs are in fact considered to be different. Most distance notions allow for two objects to be "equidistant" (i.e., allow indifference); hence we also introduce a notion of tie-breaking. As we will show, tie-breaking is only necessary when I is not convex. Note that any tie-breaker $\hat{d}_\mu$ is also a distance function with respect to μ.

Definition 3 (MDSEU). A family of preference relations $\{\succsim_I\}_{I\in\mathscr{I}}$ admits a **Minimum Distance Subjective Expected Utility** representation if there are a Bernoulli utility function $u : X \to \mathbb{R}$, a prior $\mu \in \Delta(S)$, a distance function $d_\mu : \Delta(S) \to \mathbb{R}$, and its tie-breaker $\hat{d}_\mu : \Delta(S) \to \mathbb{R}$ such that

(i) for each $I \in \mathcal{I}$, the preference relation $\succsim_I$ admits a SEU representation with (u, μ_I), meaning that for any $f, g \in F$,

$$f \succsim_I g \quad \text{if and only if} \quad \sum_{s\in S} \mu_I(s)u\big(f(s)\big) \geq \sum_{s\in S} \mu_I(s)u\big(g(s)\big), \tag{6}$$

where

$$\mu_I \equiv \arg\min \hat{d}_\mu\big(\arg\min d_\mu(I)\big),$$

(ii) for each convex $I \in \mathcal{I}$,

$$\mu_I = \arg\min d_\mu(I).$$

The family $\{\succsim_I\}_{I\in\mathscr{I}}$ of MDSEU preferences is characterized by $(u, d_\mu, \hat{d}_\mu)$. We will restrict our attention to continuous notions of distance. Since we do not require d_μ to be injective (e.g., we allow for ties), we also require the tie-breaker to satisfy a notion of continuity that is consistent with the distance notion. To that end, we provide the following definition.

Definition 4. A pair $(d_\mu, \hat{d}_\mu)$ is **upper semicontinuous** if for any two sequences $\{\pi^n\}, \{\overline{\pi}^n\}$ in $\Delta(S)$ with $\pi^n \to \pi$ and $\overline{\pi}^n \to \overline{\pi}$, if $(d_\mu(\pi^k), \hat{d}_\mu(\pi^k)) > (d_\mu(\overline{\pi}^k), \hat{d}_\mu(\overline{\pi}^k))$ for every $k \in \mathbb{N}$, then $(d_\mu(\pi), \hat{d}_\mu(\pi)) > (d_\mu(\overline{\pi}), \hat{d}_\mu(\overline{\pi}))$. A distance function d_μ is **locally nonsatiated** with respect to μ if for any $\pi \in \Delta(S) \setminus \{\mu\}$ and $\epsilon > 0$, there is a π' such that $\|\pi - \pi'\| < \epsilon$ and $d_\mu(\pi') < d_\mu(\pi)$.

2.4 Notions of Distance

In our model, the DM's notion of distance is subjective, and thus our framework allows for a wide variety distance notions. In this section, we discuss a few examples of distance functions. We begin with a well-known method of measuring the distance between probability measures from information theory.

Definition 5 (**Kullback-Leibler Divergence**)**.** Let d_μ be the Kullback-Leibler (KL) distance function:

$$d_\mu(\pi) = -\sum_{i=1}^{n} \mu_i \ln\left(\frac{\pi_i}{\mu_i}\right) \tag{7}$$

This distance function is particularly easy to interpret when information is in the form of an α-event: $I_E^\alpha = \{\pi \in \Delta(S) \mid \pi(E) = \alpha\}$. In this case,

$$\mu_I(s) = \alpha\,\frac{\mu(s)}{\mu(E)}\mathbb{1}\{s \in E\} + (1-\alpha)\,\frac{\mu(s)}{\mu(E^c)}\mathbb{1}\{s \in E^c\}. \tag{8}$$

That is, the DM shifts probability mass between events E and E^c, maintaining the relative probabilities between states within E and E^c. When $\alpha = 1$ (i.e., I_E^α represents a standard event), the KL distance function yields Bayesian updating: $\mu_I(s) = \frac{\mu(s)}{\mu(E)}\mathbb{1}\{s \in E\}$.

Note that KL is well-defined only for distributions that have the same support[7] and so we are abusing notation a bit. Thus, KL and Bayesian updating are both undefined for certain forms of information. When we refer to KL as an MDSEU, we really mean that d_μ is consistent with KL where it is well-defined. In section 4.2, we show how a MDSEU distance may be viewed as a "support-dependent extension" of KL, so that it is well defined for all distributions. By extending KL in this way, it can be used to study updating after zero-probability events.

[7] In infinite state spaces, for distributions that are mutually absolutely continuous.

Definition 6 (**Generalized Bayesian Divergence**)**.** For a twice differentiable function $\sigma : \mathbb{R}_+ \to \mathbb{R}$ such that $\sigma'' < 0$, $\lim_{x\to+\infty} \sigma(x) = +\infty$, and $\lim_{x\to+\infty} \frac{\sigma(x)}{x} = 0$, let d_μ be given by

(9)
$$d_\mu(\pi) = -\sum_{i=1}^{n} \mu_i \, \sigma(\frac{\pi_i}{\mu_i}).$$

Notice that this includes the KL distance function as a special case ($\sigma(x) = \ln(x)$). When information corresponds to the standard dynamic setup, i.e., $I = \{\pi \in \Delta(S) \mid \pi(E) = 1\}$, d_μ yields a Bayesian posterior: $\mu_I(s) = \frac{\mu(s)}{\mu(E)} \mathbb{1}\{s \in E\}$. However, for more general information sets, the precise form of σ will matter. Consequently, "Bayesian" DMs might disagree with each other when provided more general information sets I. This is studied in section 4.

Definition 7 (h-**Bayesian**)**.** Let $d_\mu(\pi) = \sum_{i=1}^{n} h_i(\mu_i) \, \sigma(\frac{\pi_i}{h_i(\mu_i)})$, where $h_i : \mathbb{R}_+ \to \mathbb{R}$ and σ satisfies the conditions from **Generalized Bayesian Divergence**.

The h-Bayesian distance notion captures a form of non-Bayesian updating where the agent is Bayesian with respect to biased beliefs. In this example, h specifies the belief distortion applied to the initial belief. For α-events,

$$\mu_{I_E^\alpha}(s) = \alpha \frac{h_s(\mu(s))}{\sum_{s' \in E} h_{s'}(\mu(s'))} \mathbb{1}\{s \in E\} + (1-\alpha) \frac{h_s(\mu(s))}{\sum_{s' \in E^c} h_{s'}(\mu(s'))} \mathbb{1}\{s \in E^c\}.$$

When $h_i(\mu_i) = \mu_i$, this reduces to Bayes' rule. When $h_i(\mu_i) = (\mu_i)^\rho$, this corresponds to a special case of Grether (1980). For $\rho < 1$, this captures under-reaction to information and base-rate neglect, while $\rho > 1$ captures over-reaction to information. It is straightforward to generalize h to capture more general belief distortions, including asymmetric reactions based on prior beliefs like confirmation bias (á la Rabin and Schrag (1999)) or over(under) reaction to small(large) probabilities (Kahneman and Tversky (1979)).

A final example that we wish to mention is the Euclidian distance.

Definition 8 (Euclidean). Let $d_\mu(\pi) = \sum_{i=1}^{n}(\pi_i - \mu_i)^2$.

For the Euclidian distance, when $I_E^1 = \{\pi \in \Delta(S) \mid \pi(E) = 1\}$, we have

$$\mu_{I_E^1}(s) = \mu_i + \frac{1 - \mu(E)}{|E|}.$$

Here, prior odds are "ignored" when updating beliefs: probability is allocated to the remaining states (i.e., those in E) uniformly.

3 Axiomatic Characterization

In this section, we present behavioral postulates that characterize the family of Minimum Distance SEU preferences. Our first axiom imposes the standard SEU conditions of Anscombe and Aumann (1963) on each (conditional) preference relation $\succsim_I$. Because these conditions are well-understood, we will not provide a formal discussion of the conditions.

Axiom 1 (**SEU Postulates**). For each $I \in \mathscr{I}$, the following conditions hold.

(*i*) **Weak Order:** $\succsim_I$ is complete and transitive.

(*ii*) **Archimedean:** For any $f, g, h \in F$, if $f \succ_I g$ and $g \succ_I h$, then there are $\alpha, \beta \in (0,1)$ such that $\alpha f + (1-\alpha)h \succ_I g$ and $g \succ_I \beta f + (1-\beta)h$.

(*iii*) **Monotonicity:** For any $f, g \in F$, if $f(s) \succsim_I g(s)$ for each $s \in S$, then $f \succsim_I g$.

(*iv*) **Nontriviality:** There are $f, g \in F$ such that $f \succ_I g$.

(*v*) **Independence:** For any $f, g, h \in F$ and $\alpha \in (0,1]$, $f \succsim_I g$ if and only if $\alpha f + (1-\alpha)h \succsim_I \alpha g + (1-\alpha)h$.

The next two axioms ensure that the DM forms a new belief that is consistent with the available information. To express our next axiom, **Compliance**, we need additional notation. For any $f \in F$ and $\pi \in \Delta(S)$, we denote by $\pi(f) \in \Delta(X)$ the lottery that yields the outcomes of f according to the probability distribution π; i.e., $\pi(f)(x) = \pi(\{s \in S \mid f(s) = x\})$ for each $x \in X$.

Axiom 2 (**Compliance**). For any $f \in F$ and $\pi \in \Delta(S)$, $f \sim_\pi \pi(f)$.

Compliance requires that the DM adheres to precise information when it is provided. That is, whenever the new piece of information is a singleton $I = \{\pi\}$ for some $\pi \in \Delta(S)$, the DM's belief conforms to π. One way to think of this axiom is to imagine a patient visiting her doctor and inquiring about a treatment. If the doctor provides extremely precise information about the treatment and the chances of success or failure, the patient accepts this information completely and adopts the doctor's information as her beliefs about the states. **Compliance** resembles consequentialism (see Ghirardato (2002)) in dynamic settings under uncertainty.

To formally state the next axiom, we first define a weak notion of equivalent information sets. Given two sets of information I and I', we say that they are *preference equivalent* if $\succsim_I = \succsim_{I'}$ (that is, $f \succsim_I g$ if and only if $f \succsim_{I'} g$ for all $f, g \in F$). In this case, we may also say that $\succsim_I$ and $\succsim_{I'}$ are equivalent. Our next axiom, **Responsiveness**, requires that the DM's preferences "respond" to the information. Consider two information sets I and I'. If these sets of information are preference equivalent (i.e., $\succsim_I$ and $\succsim_{I'}$ are equivalent), so that the DM responds to them in the same way, then these two pieces of information must have some "common information" (i.e., $I \cap I' \neq \emptyset$).

Axiom 3 (**Responsiveness**). For any $I, I' \in \mathscr{I}$, if $\succsim_I = \succsim_{I'}$, then $I \cap I' \neq \emptyset$.

Another way to understand this condition is to consider the contrapositive: mutually exclusive sets of information should never be preference equivalent.

Under Axioms 1, 2, and 3, a family of preference relations $\{\succsim_I\}_{I\in\mathscr{I}}$ admits a SEU representation, $\big((u_I, \mu_I)_{I\in\mathscr{I}}\big)$, with $\mu_I \in I$ for all I. We call such a family **Information-Dependent SEU** preferences. Notice that the initial preference $\succsim$ and new preferences $\succsim_I$ may be unrelated. In particular, (i) the DM's risk attitudes may vary across different pieces of information and (ii) the DM may form a new belief that is completely independent of the initial belief.

The goal of our next few axioms is to connect the conditional preferences with the initial one in a systematic way. The next axiom, **Invariant Risk Preference**, requires that the DM's preference over lotteries does not change when information is provided.

Axiom 4 (**Invariant Risk Preference**). For any $I \in \mathscr{I}$ and all lotteries $p, q \in \Delta(X)$, $p \succsim_I q$ if and only if $p \succsim q$.

This postulate ensures that for each $I \in \mathcal{I}$, the expected utility function u_I is a positive, affine transformation of the initial utility function u. Hence, we can normalize the utility functions, and the family of Information-Dependent SEU preferences $\{\succsim_I\}_{I\in\mathscr{I}}$ is characterized by $\big(u, (\mu_I)_{I\in\mathscr{I}}\big)$.

The next axiom, **Informational Betweenness**, is the most important behavioral condition for our model. Loosely, **Informational Betweenness** implies a form of "belief consistency" across various information sets. For an intuition behind **Informational Betweenness**, consider $E \subseteq S$ and let $I_1 = \{\pi \in \Delta(S) \mid \pi(E) \geq \frac{1}{2}\}$ be the information "E is more likely than E^c." This may alter the DM's preferences regarding bets on E. Suppose that the more refined information, $I_2 = \{\pi \in \Delta(S) \mid \frac{3}{4} \geq \pi(E) \geq \frac{1}{2}\}$, induces the same conditional preferences regarding bets on E. This suggests that the DM's willingness to bet on E is not dependent on the upper bound of I_2 (e.g., because it is determined by the lower bound placed on the probability of E), and so any information set of the form $I_3 = \{\pi \in \Delta(S) \mid \beta \geq \pi(E) \geq \frac{1}{2}\}$ should yield exactly the same willingness to bet on E as I_1 for any $\beta \in [\frac{3}{4}, 1]$. **Informational Betweenness** extends this idea to more general information sets.

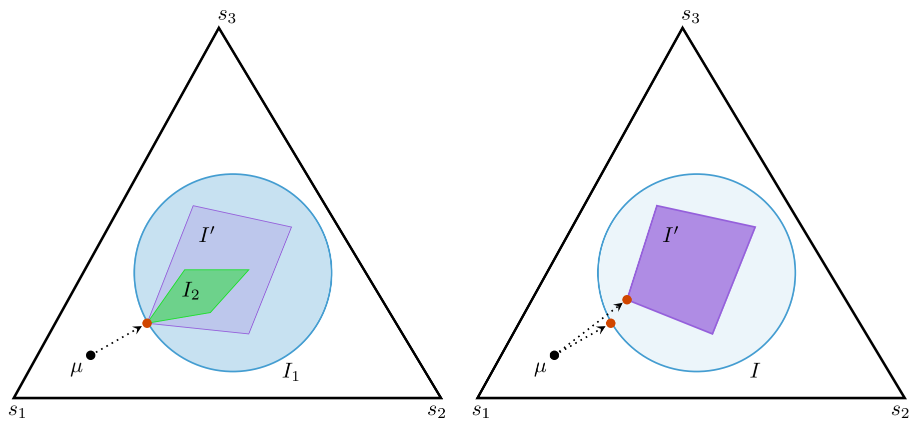

Figure 2: Illustrations of Informational Betweenness (left) and Extremeness (right).

Axiom 5 (**Informational Betweenness**). For all $I_1, I_2, I' \in \mathscr{I}$ such that $I_2 \subseteq I' \subseteq I_1$, if $\succsim_{I_1} = \succsim_{I_2}$, then $\succsim_{I_1} = \succsim_{I'}$.

In other words, **Informational Betweenness** requires that if the least precise and most precise information sets, I_1 and I_2, provide "the same information" (i.e., are behaviorally equivalent), then any intermediate information set, I_3, must also provide the same information as these two sets. This logic is illustrated in Figure 2.

At this point, it is worth remarking that **Informational Betweenness** captures most of the behavioral content of MDSEU. That is, consider the family of Information-Dependent SEU preferences $\{\succsim_I\}_{I \in \mathscr{I}}$ characterized by $\big(u, (\mu_I)_{I \in \mathscr{I}}\big)$. Under **Informational Betweenness**, it turns out that μ_I is the minimizer of some (complete and transitive) ordering, which, of course, depends on the prior μ. To ensure this order is consistent with a distance function, we require two technical conditions, **Extremeness** and **Continuity**.

For an intuition behind **Extremeness**, imagine that the DM exhibits a change in behavior after learning I; she finds I to be "informative" and changes her beliefs ($\succsim_I \neq \succsim$). Then, any "more informative" I' must similarly result in additional changes in behavior

($\succsim_{I'} \neq \succsim_I$). In other words, a DM who adjusts her beliefs after some information I must continue to move her beliefs when you present her with any strictly more informative information set I'. As we have seen in our discussion of **Informational Betweenness**, being a (strict) subset is not sufficient for the DM to perceive a set of information as more informative because the information sets may intersect at the chosen belief. Drawing on this insight, we suggest that interiority is the correct notion of "more informative." This logic is illustrated in Figure 2. Formally, for any I, let $\text{int}(I)$ be the interior of I.[8]

Axiom 6 (**Extremeness**). For any convex $I, I' \in \mathscr{I}$ with $I' \subseteq \text{int}(I)$, if $\succsim_I \neq \succsim$, then $\succsim_{I'} \neq \succsim_I$.

Our last postulate, **Continuity**, ensures that conditional preferences change in a continuous fashion with respect to the provided information.[9]

Axiom 7 (**Continuity**). For any two sequences $\{I_k\}, \{J_k\}$ in $\mathscr{I}$ such that $I_k \to I$ and $J_k \to J$, if $\succsim_{I_k} = \succsim_{J_k}$ for each k, then $\succsim_I = \succsim_J$.

Theorem 1. *A family of preference relations $\{\succsim_I\}_{I \in \mathscr{I}}$ satisfies Axioms 1 through 7 if and only if it admits a* **Minimum Distance SEU** *representation with respect to some locally nonsatiated distance function d_μ and its continuous tie-breaker $\hat{d}_\mu$ such that $(d_\mu, \hat{d}_\mu)$ is upper semicontinuous.*

By the uniqueness of subjective expected utility representations, u, μ, and μ_I are unique. We obtain two forms of ordinal uniqueness for the distance notions. First, the set of minimizers of the distance functions must be "consistent." This means that there are no probabilities with opposite, strict ranking in terms of distance: $d'_\mu(\pi') > d'_\mu(\pi)$ and $d'_\mu(\pi) > d_\mu(\pi')$ may not happen for any π, π'. Second, tie-breaking rules must be ordinally equivalent. This is summarized in the following proposition.

[8] More formally, $\text{int}(I) = \{\pi \in I \mid \exists \epsilon > 0 \text{ such that } B(\pi, \epsilon) \subseteq I\}$ where $B(\pi, \epsilon) = \{\pi' \in \Delta \mid \|\pi - \pi'\| \leq \epsilon\}$.

[9] We endow $\mathscr{I}$ with the Hausdorff topology. The Hausdorff distance between information sets I and I' is given by $h(I, I') \equiv \max\{\max_{\pi \in I} \min_{\pi' \in I'} \|\pi - \pi'\|, \max_{\pi' \in I'} \min_{\pi \in I} \|\pi - \pi'\|\}$.

Proposition 1. *Suppose the family of preference relations* $\{\succsim_I\}_{I\in\mathscr{I}}$ *admits MDSEU representations with* $(u, d_\mu, \hat{d}_\mu)$ *and* $(u', d'_{\mu'}, \hat{d}'_{\mu'})$*. Then (i)* $u = \alpha u' + \beta$ *for some* α, β *with* $\alpha > 0$*, (ii)* $\mu = \mu'$*, (iii)* $\arg\min \hat{d}_\mu(\arg\min d_\mu(I)) = \arg\min \hat{d}'_{\mu'}(\arg\min d'_{\mu'}(I))$ *for each* $I \in \mathscr{I}$*, and (iv)* $\hat{d}_\mu$ *and* $\hat{d}'_{\mu'}$ *are ordinally equivalent, i.e., for any* $\pi, \pi' \in \Delta$*,* $\hat{d}_\mu(\pi) > \hat{d}_\mu(\pi')$ *if and only if* $\hat{d}'_{\mu'}(\pi) > \hat{d}_{\mu'}(\pi')$*.*

4 Bayesian Updating with General Information

In section 2.4, we defined **Generalized Bayesian Divergence**, a distance notion that leads to posteriors consistent with Bayesian updating on standard events. In this section, we show that these are essentially the only distance notions that lead to a version of Bayesian updating defined for general information.

We say that $E \subseteq S$ is $\succsim$-*null* if $fEg \sim g$ for all $f, g \in F$, otherwise E is $\succsim$-nonnull. We extend this definition to conditional preferences in the natural way.

Recall our notion of an α-event; information sets of the form: $I_E^\alpha \equiv \{\pi \in \Delta \mid \pi(E) = \alpha\}$ for some $E \subseteq S$. In the standard setting, revealing that a nonnull event E occurred (i.e., $\alpha = 1$) induces a Bayesian DM to revise her prior μ by proportionally allocating all probability mass among states in E; i.e., $\pi_i = \frac{\mu_i}{\mu(E)}$. This principle of proportionality ought to apply to more general information sets, including all α-events I_E^α. That is, a Bayesian DM should allocate the given probability α among states in E in such a way that preserves relative probabilities. Thus, updated beliefs will still be proportional to her prior, $\pi_i = \frac{\mu_i}{\mu(E)}\alpha$. We call this updating procedure **Extended Bayesian Updating** (henceforth, EBU), which we formally define below.

Definition 9 (**Extended Bayesian Updating**)**.** Beliefs satisfy Extended Bayesian

Updating if for any $E \subseteq S$ with $\mu(E) > 0$, $s \in E$, and $\alpha \in (0, 1]$,

$$\mu_{I_E^\alpha}(s) = \frac{\mu(s)}{\mu(E)}\,\alpha.$$

EBU is illustrated in Figure 3. The dashed line can be thought of as the "Bayesian expansion path" of beliefs μ. That is, EBU ensures that for any α, posterior beliefs after the $I^\alpha_{\{s_2,s_3\}}$ information set must lie on this dashed line.

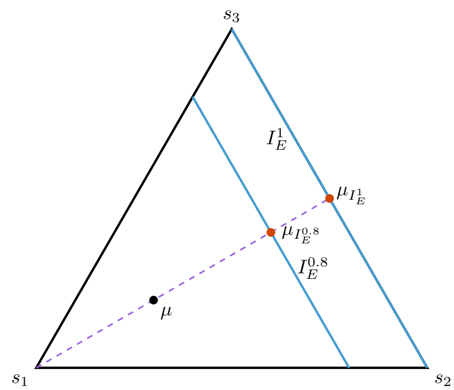

Figure 3: Bayesian expansion path of μ and $I^\alpha_{\{s_2,s_3\}}$, for $\alpha = 0.8$ and $\alpha = 1$.

For the standard notion of information as an event, it is well-known that dynamic consistency characterizes Bayesian updating (see Epstein and Breton (1993) and Ghirardato (2002)). However, as our setting allows for more general information, we need to extend dynamic consistency beyond standard events. The key axiom, called **Informational Dynamic Consistency** (IDC), requires that if a DM prefers act f to act g before she receives any information and the two acts coincide outside of E (i.e., $f(s) = g(s)$ for all $s \in E^c$), then she prefers f to g after the information set I_E^α is revealed, and vice versa. Obviously, IDC corresponds to the standard notion of dynamic consistency when $\alpha = 1$.

Axiom 8 (**Informational Dynamic Consistency**). For all acts $f, g, h \in \mathcal{F}$, any $\succsim$-

nonnull event $E \subseteq S$, and $\alpha \in (0,1]$,

$$(10) \qquad fEh \succsim gEh \quad \text{if and only if} \quad fEh \succsim_{I_E^\alpha} gEh.$$

It turns out that within the class of MDSEU preferences, **Informational Dynamic Consistency** fully characterizes **Extended Bayesian Updating**.

Proposition 2. *Let $\{\succsim_I\}_{I\in\mathscr{I}}$ be a family of MDSEU preferences. Then, $\{\succsim_I\}_{I\in\mathscr{I}}$ satisfy* **Informational Dynamic Consistency** *if and only if beliefs exhibit* ***Extended Bayesian Updating****.*

When an information set I_E^α is revealed, a DM following EBU allocates α among the states in E in a proportional manner and the remaining probability mass, $1-\alpha$, among the states in the complementary event E^c. This implies that all $\succsim$-null states remain $\succsim_{I_E^\alpha}$-null after receiving information I_E^α.

We now formally show that under some minor technical conditions on the distance function d_μ, **Informational Dynamic Consistency** essentially characterizes the family of **Generalized Bayesian Divergence** distance functions.

Proposition 3. *Suppose $d_\mu(\pi) = \sum_{i=1}^n d_i(\pi_i)$ with $d_i'' > 0$ for each i. If* **Informational Dynamic Consistency** *is satisfied, then there is a* ***Generalized Bayesian Divergence****, σ, such that $d_\mu(\pi) = \sum_{i=1}^n \mu_i\, \sigma(\frac{\pi_i}{\mu_i})$ when $\pi \in \Delta \cap [0,\frac{1}{2}]^n$.*

4.1 Bayesian Disagreement

While the previous results shed light on the general structure of Bayesian distance (i.e., it must be a **Generalized Bayesian Divergence**), they also show that there is no unique "Bayesian distance." Indeed, any σ satisfying the requirements in Definition 6 generates posteriors consistent with EBU on α-events. This lack of uniqueness, therefore,

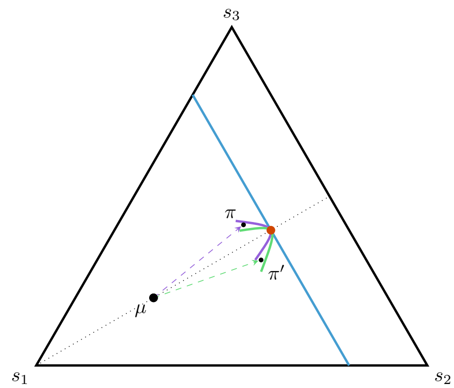

Figure 4: Bayesian disagreement between π and π'.

leaves open the possibility that on "irregular" information sets (e.g., those that do not correspond to α-events), Bayesian DM's may disagree.

To see why disagreement arises, consider any **Generalized Bayesian Divergence**. The properties on σ ensure that for any α-event, the DM always shifts weight proportionally across states. Put another way, dynamic consistency places restrictions on beliefs after information sets that are "event-like." Crucially, such information does not fundamentally challenge the DM's beliefs about the relative likelihood of states. However, information that is not "event-like" may necessitate the revision of relative likelihoods by the DM. When information sets preclude proportional shifts, the specifics of the distance function matter.

4.1.1 Polarization

To illustrate polarization, imagine two policy makers, Alice and Bob, trying to understand the risks of climate change. They begin with the same beliefs and consult a panel of scientists. The panel advises that there are two accepted models, π and π'. Thus, Alice and Bob have both been provided with the information set $I = \{\pi, \pi'\}$. As they are both MDSEU DMs, each will each adopt one of these models as his or her new belief. Alice and Bob will agree after (binary) information sets I if and only if their distance functions are ordinally equivalent. This is much more demanding than requiring that they are, individually, Bayesian agents. Indeed, Figure 4 illustrates two iso-distance curves. The blue curve implies that μ is closer to π' than π, while the red curve implies the opposite. Importantly, the curves intersect on the Bayesian expansion path of μ, and so they both generate Bayesian posteriors on α-events.

More concretely, suppose that there are three states, $S = \{s_1, s_2, s_3\}$, and that Alice and Bob agree on their prior, $\mu^A = \mu^B = (0.4, 0.3, 0.3)$. They differ, however, in how they judge information. Let Alice's distance be $d^A_\mu(\pi) = -\sum_i \mu_i \ln(\frac{\pi_i}{\mu_i})$ and Bob's distance be $d^B_\mu(\pi) = -\sum_i \mu_i (\frac{\pi_i}{\mu_i})^{\frac{1}{2}}$. Note that these are both examples of **Generalized Bayesian Divergence**, and so Alice and Bob satisfy IDC and will agree after every α-event. However, consider $\pi = (0.3, 0.375, 0.325)$ and $\pi' = (0.335, 0.405, 0.26)$. Since neither π nor π' lie on the Bayesian expansion path of μ (illustrated in Figure 4), IDC places no restriction on posterior beliefs. In this instance, Alice will adopt π, while Bob adopts π'.

4.1.2 Trade and Public Information

In economies with Bayesian traders who share a common prior over the states, neither public nor private information generates incentives to re-trade Pareto-efficient allocations (see Milgrom and Stokey, 1982; Morris, 1994). In particular, if the initial Pareto allocation has the full-insurance property and the traders receive public information about events that occurred, there will be no trade. This "no-trade" result extends to α-events when DMs distances are **Generalized Bayesian Divergence**. However, when publicly available information is more general, speculative trade is possible.

To illustrate, consider a pure-exchange economy under uncertainty with $S = \{s_1, s_2, s_3\}$. Two traders, Alice and Bob, share a common prior over S given by $\mu = (0.5, 0.3, 0.2)$. An allocation $f = (f^A, f^B)$ is a tuple of state-contingent consumption of one commodity (i.e., $f^i \in \mathbb{R}^3_+$ with $i \in \{A, B\}$). Both traders are MD-SEU maximizers with respect to the same (strictly) concave utility function, $u^A(x) = u^B(x) = \sqrt{x}$ for any $x \in \mathbb{R}_+$. However, they have different (Bayesian) distance functions $d^A_\mu(\pi) = -\sum_i \mu_i \ln(\frac{\pi_i}{\mu_i})$ and let $d^B_\mu(\pi) = -\sum_i \mu_i (\frac{\pi_i}{\mu_i})^{\frac{1}{2}}$. The initial allocation is the full-insurance allocation: $e^A = e^B = (5, 5, 5)$.

Suppose that the Bayesian traders learn publicly that probability of an event E is α. Since I^α_E crosses the Bayesian-expansion path, both traders choose the same posterior. Thus, the full-insurance allocation remains efficient after updating. Due to common posteriors in the presence of publicly available α-events, there will be no-trade among the Bayesian traders.

Now, suppose there are two research institutes that provide likelihood estimates for the economy. In their annual reports, both institutes publish different probability estimates: $\pi^1 = \{0.25, 0, 25, 0.5\}$ and $\pi^2 = \{0.2, 0.4, 0.4\}$. Will such information generate trade?

Alice updates the common prior by selecting π^1 while Bob chooses π^2 as his new belief.[10] Since both traders disagree on their posteriors, a Pareto improving exchange is possible. For instance, the feasible allocation $f^A = (5.15, 3.5, 6)$ and $f^B = (4.85, 6.5, 4)$ makes both traders strictly better off than the full-insurance allocation.[11] Notice that the trade leading to the Pareto-superior allocation (f^A, f^B) is not driven by risk sharing but by speculative motives. Put differently, both traders are willing to abandon the full-insurance allocation in order to bet against each other by purchasing assets that correspond to the transfers $(f^A - e^A)$ and $(f^B - e^B)$. Gilboa et al. (2014) call such trades "speculative" Pareto improvement bets.[12]

[10] We have that $d^A_\mu(\pi^1) = 0.29 < 0.23 = d^A_\mu(\pi^2)$ and $d^B_\mu(\pi^1) = -0.94 > -0.95 = d^B_\mu(\pi^2)$

[11] $0.25 \cdot \sqrt{5.15} + 0.25 \cdot \sqrt{3.5} + 0.5 \cdot \sqrt{6} = 2.26 > 2.236 = \sqrt{5}$ and $0.2 \cdot \sqrt{4.85} \ + 0.8(\cdot\sqrt{6.5} + \cdot\sqrt{4}) = 2.26 > 2.236$.

[12] Gilboa et al. (2014) distinguish between Pareto improvements due to betting and due to risk-sharing. An allocation f is called a *bet* if f Pareto dominates another allocation g with the full-insurance property.

4.2 Zero-probability Events and Hypothesis Testing

One of the well-known weaknesses of Bayesian updating is that it is not defined for zero-probability events. In contrast, our notion of belief updating is well-defined for zero-probability events. Thus, MDSEU provides a way to extend (non-)Bayesian updating to all events.

A recent addition to the literature on updating after zero-probability events is the Hypothesis Testing model (HT) of Ortoleva (2012).[13] Such an agent will update using Bayes' rule for expected events: events with probability above some threshold ϵ. When an event E is unexpected (i.e., under the agent's prior $\mu(E) \leq \epsilon$), the agent rejects her prior, updates a second-order prior over beliefs, and selects a new belief according to a maximum likelihood procedure. Formally, a HT representation is given by a triple, (μ, ρ, ϵ), consisting of a prior $\mu \in \Delta(S)$, a second order prior $\rho \in \text{int}\big(\Delta(\Delta(S))\big)$ and a threshold $\epsilon \in [0, 1)$. Then, for any $E \subseteq S$ and $s \in E$,

$$\mu^E(s) = \begin{cases} \frac{\mu(s)}{\mu(E)} & \text{when } \mu(E) > \epsilon, \\ \frac{\pi^\rho_E(s)}{\pi^\rho_E(E)} & \text{when } \mu(E) \leq \epsilon \end{cases}$$

where $\pi^\rho_E = \arg\max_{\pi \in \Delta} \rho(\pi)\pi(E)$.

As it turns out, HT may be viewed as special case of the MDSEU. We show this by explicitly constructing a distance function that is behaviorally equivalent to a given HT representation on standard events.

For an intuition behind the construction of this distance, note that HT involves "multiple" beliefs and is non-Bayesian only for certain events (e.g., unexpected events). For simplicity, consider the case of $\epsilon = 0$. Then, an event is expected if it is given positive

[13] There are other models that allow for conditioning on zero-probability events, such as the conditional probability systems introduced by Myerson (1986a,b) and the conditional lexicograhic probability systems axiomatized by Blume et al. (1991).

probability by the prior, and so an event is surprising if and only if it was considered "impossible" under the prior. Thus, the support of the event and prior have a non-empty intersection in the former case and are disjoint in the latter. Correspondingly, the distance function must distinguish between potential beliefs with over-lapping supports and non-overlapping supports. However, once this restriction is accommodated, the distance function is almost Bayesian.

For any $\pi \in \Delta(S)$, let $sp(\pi)$ denote the support of π.

Proposition 4. *Consider a Hypothesis Testing representation, (μ, ρ, ϵ). Further, suppose that for each $\pi \in sp(\rho)$, $\pi \neq \mu$, $sp(\pi) = S$ and that $\rho(\pi) \neq \rho(\pi')$ for any two distinct $\pi, \pi' \in \Delta$. Then let*

$$d_\mu(\tilde{\pi}) = \begin{cases} -\sum_{s \in sp(\tilde{\pi})} \mu(s) \log(\tilde{\pi}(s)) - M\left|\{sp(\tilde{\pi}) \cap sp(\mu)\}\right| \ \textit{if} \ \mu(sp(\tilde{\pi})) > \epsilon, \\ -\sum_{s \in sp(\tilde{\pi})} \pi^{\rho}_{sp(\pi)}(s) \ \log(\tilde{\pi}(s)) + M(|S| + 1 - |sp(\tilde{\pi})|) \ \textit{if} \ \mu(sp(\tilde{\pi})) \leq \epsilon. \end{cases}$$

If M is large enough, then for any $E \subseteq S$ and $s \in E$,

$$\mu_{I^1_E}(s) = \begin{cases} \frac{\mu(s)}{\mu(E)} & \textit{when} \ \mu(E) > \epsilon, \\ \frac{\pi^{\rho}_E(s)}{\pi^{\rho}_E(E)} & \textit{when} \ \mu(E) \leq \epsilon. \end{cases}$$

The details of this construction can be found in Appendix section A.5.

5 Recovering Distance Functions

In order to operationalize MDSEU, we show how to recover the underlying distance function for a particular model of updating from the standard framework. That is, consider a standard event (signal) structure on a set of states and some updating rule. We show how to construct a distance function in the MDSEU framework that coincides with this updating rule on information sets that are "equivalent" to the events (signals).

Proposition 5. *Suppose d_μ has an additive form: $d_\mu(\pi) = \sum_{i=1}^n f_i(\pi_i)$. Further, suppose there is a function $g : [0,1] \to [0,1]$ such that $g_i(\alpha) = \mu_{I^\alpha_{\{s_1,s_i\}}}(s_i)$ for each $\alpha \in [0,1]$ and that g_i^{-1} exists. Then for some function f,*

$$d_\mu(\pi) = f(\pi_1) + \sum_{t=2}^{n} \int_0^{\pi_i} f'(g_i^{-1}(t) - t)dt \text{ for any } \pi \in [0,1] \times \prod_{i=1}^{n}[0, g_i(1)].$$

Mathematically, the problem of recovering a distance function is related to the classic problem of recovering utility from demand. Therefore, the technical assumptions regarding separability are similar. Here, $g_i(\alpha)$ is a function that specifies the probability allocated to state i when the DM splits α probability between s_i and s_1. We illustrate how this can be used by explicitly deriving the h-Bayesian distance in the following section.

5.1 Constructing Non-Bayesian Distance Functions

Since we already know the general structure of Bayesian distance functions, the primary use of Proposition 5 is to enable the construction of non-Bayesian distance functions. This allows us to extend "standard" rules to general information. As an illustration, suppose that we want to find a distance function that yields, for each $A \subseteq S$, the following generalization of Bayes' rule:

$$\pi_i^A = \frac{h_i(\mu_i)}{\sum_{j \in A} h_j(\mu_j)}.$$

First, we can naturally extend this to α-events in our setting, so that

$$\mu_{I^\alpha_A}(s_i) = \alpha \frac{h_i(\mu_i)}{\sum_{j \in A} h_j(\mu_j)}.$$

Note that this corresponds to h-Bayesian updating. Then we define $g_i(\alpha) = \alpha \frac{h_i(\mu_i)}{h_1(\mu_1)+h_i(\mu_i)}$. Hence $g_i^{-1}(t) = t(1 + \frac{h_1(\mu_1)}{h_i(\mu_i)})$, and using Proposition 5 we determine

$$d_\mu(\pi) = f(\pi_1) + \sum_{t=2}^{n} \int_0^{\pi_i} f'(g_i^{-1}(t) - t)dt$$

$$= f(\pi_1) + \sum_{t=2}^{n} \int_0^{\pi_i} f'(t\,\frac{h_1(\mu_1)}{h_i(\mu_i)})dt = f(\pi_1) + \sum_{t=2}^{n} \frac{h_i(\mu_i)}{h_1(\mu_1)} f(\pi_i\,\frac{h_1(\mu_1)}{h_i(\mu_i)}).$$

Letting $d(t) = \frac{f(t\,h_1(\mu_1))}{h_1(\mu_1)}$, we derive $d_\mu(\pi) = \sum_{i=1}^{n} h_i(\mu_i) d(\frac{\pi_i}{h_i(\mu_i)})$, as in Definition 7.

5.2 Non-Bayesian Behavior

While the MDSEU provides a structured way to extend "Bayesian updating" to general information sets, it also allows for a variety of distance notions that generate familiar models of non-Bayesian updating. In this way, the MDSEU framework contextualizes non-Bayesian behavior as merely the result of different notions of distance. For instance, consider the following generalization of Bayes' rule. For $\rho > 0$,

$$\mu_{I_E^1}(s) = \frac{(\mu(s))^\rho}{\sum_{s' \in E}(\mu(s'))^\rho} \text{ when } s \in E.$$

This corresponds to a special case of h-Bayesian distance function, where $h_i(\mu_i) = (\mu_i)^\rho$. This notion of distance is a special case of the non-Bayesian model introduced by Grether (1980) and can capture over- and under-inference from signals.

As an illustration, consider a stylized learning experiment with two payoff-relevant states $\mathbb{P} = \{A, B\}$ and two signals $\Theta = \{a, b\}$; the state space is $S = \mathbb{P} \times \Theta$. Let $I_a = \{\pi : \pi(S \times \{a\}) = 1\}$ and $I_b = \{\pi : \pi(S \times \{b\}) = 1\}$ denote the general information sets "there was an a signal" and "there was a b signal," respectively. Let $\sigma(a|A) := \frac{\mu(A,a)}{\mu(A,a)+\mu(A,b)} = \beta_A > \frac{1}{2}$ denote the conditional probability of an a signal, while $\mu(A) = \mu(A,a) + \mu(A,b)$ is the unconditional probability of payoff state A. We extend this to b

and B in the natural way. Upon learning I_a,

$$\mu_{I_a}(A|a) = \frac{\mu(A,a)^\rho}{\mu(A,a)^\rho + \mu(B,a)^\rho} = \frac{\sigma(a|A)^\rho \mu(A)^\rho}{\sigma(a|A)^\rho \mu(A)^\rho + \sigma(a|B)^\rho \mu(B)^\rho}.$$

For $\rho = 1$, we have Bayesian updating. For $\rho < 1$, this captures aspects of both under-inference and base-rate neglect, which are both commonly observed. In the reverse case, $\rho > 1$, we capture over-inference and over-reliance on priors.[14] Therefore, the MDSEU framework allows us to extend non-Bayesian updating to general information.

6 Related Literature

There are a few papers that develop a theory of belief updating for general information structures. Perhaps the first to study general information was Damiano (2006), who studies the (non)existence of belief selection rules with various properties.[15] In a similar spirit to Damiano (2006), Chambers and Hayashi (2010) prove the non-existence of a selection rule satisfying a particular form of Bayesian consistency. In contrast, we focus on the behavioral implications of a particular selection rule: distance minimization.

As far as we are aware, the first paper to combine general information and distance minimization is Zhao (2020), who considers environments in which a DM receives a sequence of qualitative statements of the form "event A is more likely than event B." The DM has a probabilistic belief and updates it via the so-called Pseudo-Bayesian updating rule.[16] There are three key differences between Zhao (2020) and our paper.

[14] See Benjamin (2019) for an excellent discussion regarding both under- and over-inference.

[15] In particular, he shows that there is no selection rule that satisfies three properties: state-independence, location-consistency, and update-consistency.

[16] The pseudo-Bayesian updating rule is axiomatized by two axioms, Exchangeability and Stationarity, directly imposed on posteriors. Exchangeability requires that the order of information does not matter as long as the DM receives qualitative statements that "neither reinforce nor contradict each other." Stationarity requires that the DM's beliefs do not change when a qualitative statement is consistent with the prior.

First, he focuses on Kullback-Leibler divergence while we allow general distance. Thus, his model coincides with Bayesian updating on standard events, while we allow non-Bayesian updating. Second, he focuses on a specific form of general information, "A is more likely than B," while we do not restrict information. In particular, this means that we allow for zero-probability events. Third, his axioms are on beliefs, while ours are on preferences.

Other papers that feature general information include Gajdos et al. (2008), Dominiak and Lefort (2020) as well as Ok and Savaockin (2020). These papers study the embedding of general information under more general preferences than SEU. In Ok and Savaockin (2020), an agent ranks "info-acts" (μ, f), which consist of a probability distribution μ over S and a (Savage) act f. They characterize the notion of probabilistically amenable preferences: the agent adopts μ as her own belief, and evaluates act f via the lottery induced by μ over the outcomes of f, although lotteries may not be evaluated by expected utility.[17] In their setting, each act might be evaluated with respect to a different probability distribution. In our model, when an info-event $I = \{\mu\}$ is a singleton, **Compliance** ensures that a SEU maximizer adopts μ to evaluate all acts.

Gajdos et al. (2008) studies preferences defined over more general "info-acts" (P, f) where P is a set of probability distributions $P \subseteq \Delta(S)$. They characterize a representation in which an ambiguity-averse agent selects a set of priors $\varphi(P) \subseteq \Delta(S)$ to evalulate f via the Maxmin-criterion, and derive conditions for $\varphi(P)$ to be consistent with P (i.e., $\varphi(P) \subseteq P$). In situations where P specifies probabilities for some events (e.g., $P = I_E^\alpha$), Dominiak and Lefort (2020) show that depending on whether an ambiguity-averse agent "selects" her priors from the objective set $\Delta(S)$ or the exogenous set P will fundamentally affect her preference, being either consistent with Machina's (2012)

[17]Probabilistically amenable preferences are weaker than the probabilistically sophisticated preferences introduced by Machina and Schmeidler (1992) as they do not need to be complete, continuous and consistent with respect to first-order stochastic dominance.

intriguing behavior or not.

The literature on updating with standard information, i.e., events, is much larger. Within this environment, a few papers have studied minimum distance updating rules. Perea (2009) axiomatizes *imaging* rules, which are minimum distance rules utilizing Euclidean distance. Under imaging, for each $E \subseteq S$ a posterior π is selected that minimizes $d_\mu(\pi) = \|\phi(\mu) - \phi(\pi)\|$, where $\pi \in \Delta(E)$ and ϕ is an affine function. Our model includes this as a special case. More recently, Basu (2019) studies AGM (Alchourrón et al., 1985) belief revision in a standard environment. Within this setting, he establishes an equivalence between updating rules that are AGM-consistent, Bayesian, and weak path independent and lexicographic updating rules. He then turns to minimum distance updating rules and shows that every support-dependent lexicographic updating rule admits a minimum distance representation. In contrast, we focus on minimum distance updating rules with general information and allow for non-Bayesian updating.

As we have shown, our model can capture some forms of non-Bayesian updating, which has a large literature (see Benjamin (2019) for an excellent summary of experimental findings and behavioral models). Some axiomatic papers on non-Bayesian updating include Epstein (2006) and Epstein et al. (2008). Both papers utilize Gul and Pesendorfer (2001)'s theory of temptation to study a DM who may be tempted to use a posterior that is inconsistent with Bayesian updating. More recently, Kovach (2020) utilizes the conditional preference approach and characterizes *conservative updating*: posterior beliefs are a convex combination of the prior and the Bayesian posterior. This behavior violates consequentialism (i.e., **Compliance**), and therefore cannot be accommodated by our model.

A Proofs of Main Results

A.1 Proof of Theorem 1

Sufficiency. We first prove sufficiency. Consider a family of preferences $\{\succsim_I\}_{I \in \mathscr{I}}$ that satisfies **SEU Postulates**, **Compliance**, **Responsiveness**, **Invariant Risk Preference**, **Informational Betweenness**, **Extremeness**, and **Continuity**.

Notations. For any set A and a preference relation $\succsim$ on A, let $\min(A, \succsim) \equiv \{x \in A | y \succsim x \text{ for any } y \in A\}$. For any $\pi, \pi' \in \Delta(S)$, let $[\pi, \pi'] \equiv \{\alpha\pi + (1-\alpha)\pi' | \alpha \in [0,1]\}$.

Step 0. By **SEU Postulates**, for each $I \in \mathscr{I}$, the conditional preference relation $\succsim_I$ admits a (conditional) SEU representation with some (u_I, μ_I), an expected utility function $u_I : X \to \mathbb{R}$, and a probability distribution $\mu_I \in \Delta(S)$. As usual, u_I is unique up to a positive transformation, and μ_I is unique. Let $u = u_{\Delta(S)}$ and $\mu = \mu_{\Delta(S)}$.

Step 1. $(u_I = u)$. Take $I \in \mathscr{I}$ and consider $\succsim$ and $\succsim_I$. By Step 0, for all $p, q \in \Delta(X)$,

$$(11) \qquad p \succsim q \text{ if and only if } u(p) \geq u(q) \quad \text{and} \quad p \succsim_I q \text{ if and only if } u_I(p) \geq u_I(q).$$

By **Invariant Risk Preference**,

$$(12) \qquad u(p) \geq u(q) \text{ if and only if } u_I(p) \geq u_I(q).$$

Hence, u_I is a positive and affine transfromation of u; i.e., there are $a \in \mathbb{R}$ and $b > 0$ such that $u_I(p) = a + bu(p)$ for all $p \in \Delta(X)$. Without loss generality, we can set $u_I := u$ for each $I \in \mathscr{I}$.

Step 2. $(\mu_\pi = \pi)$. Take a probability distribution $\pi \in \Delta(S)$ and let $I = \{\pi\}$ be a singleton set. By **Compliance** and Step 0, for any $f \in F$,

$$f \sim_{\{\pi\}} \pi(f) \text{ if and only if } \sum_{s\in S} \mu_\pi(s)u(f(s)) = \sum_{s\in S} \pi(s)u(f(s)). \tag{13}$$

By the uniqueness of u and μ_π and Step 1, we have $\mu_\pi = \pi$.

Step 3. ($\mu_I \in I$). Take $I \in \mathscr{I}$. By Step 0, $\succsim_I$ is a SEU preference with respect some probability measure $\mu_I \in \Delta(S)$. Define $I' = \{\mu_I\}$. By Step 2, $\mu_{I'} = \mu_I$. Hence, $\succsim_I$ and $\succsim_{I'}$ are equivalent. Then, by **Responsiveness**, $I \cap I' \neq \emptyset$. Therefore, we have $\mu_I \in I$.

We define a mapping $C : \mathscr{I} \to \Delta(S)$ such that for any $I \in \mathscr{I}$, $C(I) = \mu_I$. Since $C(I) \in I$, mapping C is a choice function on $\mathscr{I}$.

Step 4. C satisfies Sen's α-property. That is, for any I, I' with $I' \subset I$ and $\pi \in I'$, if $\pi = C(I)$, then $\pi = C(I')$.

Let $I_1 = I$ and $I_2 = \{\pi\}$. Note that $I_2 \subseteq I' \subseteq I_1$ and $\pi = C(I)$ implies that $\succsim_{I_1} = \succsim_{I_2}$. By **Informational Betweenness**, $\succsim_{I_1} = \succsim_{I'}$, equivalently, $\pi = C(I')$.

Revealed Preference 1. Define $\succsim^*$ as follows: for any distinct $\pi, \pi' \in \Delta(S)$, $\pi' \succ^* \pi$ if $\pi = C(\{\pi, \pi'\})$ and $\pi \sim^* \pi$. Note that $\succsim^*$ is a strict preference relation.

Step 5. Since C is a choice function and satisfies Sen's α-property, $\succsim^*$ is complete and transitive preference relation. Moreover, $C(I) = \min(I, \succsim^*)$.

Let $\mu \equiv \min(\Delta(S), \succsim^*)$.

Step 6. Take any convex $I \in \mathscr{I}$ and $\pi \in \text{int}(I)$. Note that $\succsim_I \neq \succsim$ iff $\mu_I \neq \mu$. Since $\mu_I = \min(I, \succsim^*)$ and $\mu \equiv \min(\Delta(S), \succsim^*)$, then $\succsim_I \neq \succsim$ iff $\mu \notin I$. Therefore, by Extremeness, $\mu \notin I$ implies $\succsim_\pi \neq \succsim_I$; equivalently, $\mu_I \neq \pi$. Therefore, $\mu_I \notin \text{int}(I)$ when $\mu \notin I$. In other words, $\mu \notin I$ implies $\mu_I \in \partial(I) = I \setminus \text{int}(I)$, the boundary of I.

Step 7. $\succsim^*$ is continuous. That is, for any $\pi \in \Delta(S)$, $U(\pi) = \{\pi' \in \Delta(S) | \pi' \succsim^* \pi\}$ and $L(\pi) = \{\pi' \in \Delta(S) | \pi \succsim^* \pi'\}$ are closed. Therefore, there is a continuous function $\hat{d} : \Delta(S) \to \mathbb{R}$ that represents $\succsim^*$. Note that $\succsim^*$ is a strict preference relation, $\hat{d}$ is injective.

Take any $\pi \in \Delta(S)$. Take a sequence $\{\pi^k\}_{k\in\mathbb{N}}$ in $U(\pi)$ such that $\pi^k \to \pi^*$. Consider two cases. First, suppose $\pi^k = \pi$ for infinitely many k. Then $\pi^* = \pi$. Hence, $\pi^* \succsim^* \pi$; i.e., $\pi^* \in U(\pi)$. Second, suppose there is some N such that $\pi^k \succ^* \pi$ for any $k > N$. Then $C(\{\pi^k, \pi\}) = \pi$ for any $k > N$. Since $\{\pi^k, \pi\} \to \{\pi^*, \pi\}$ and $\pi \to \pi$, by Continuity we have $C(\{\pi^*, \pi\}) = \pi$; i.e., $\pi^* \succ \pi$.

Similarly, Take a sequence $\{\pi^k\}_{k\in\mathbb{N}}$ in $L(\pi)$ such that $\pi^k \to \pi^*$. Consider two cases. First, suppose $\pi^k = \pi$ for infinitely many k. Then $\pi^* = \pi$. Hence, $\pi \succsim^* \pi^*$; i.e., $\pi^* \in L(\pi)$. Second, suppose there is some N such that $\pi \succ^* \pi^k$ for any $k > N$. Then $C(\{\pi^k, \pi\}) = \pi^k$ for any $k > N$. Since $\{\pi^k, \pi\} \to \{\pi^*, \pi\}$ and $\pi^k \to \pi^*$, by Continuity we have $C(\{\pi^*, \pi\}) = \pi^*$; i.e., $\pi \succ \pi^*$.

Revealed Preference 2. Define $\succsim^{**}$ as follows: for any $\pi, \pi' \in \Delta(S)$, $\pi' \succsim^{**} \pi$ if $\pi = C(\{\pi, \pi'\})$ and $\pi \succ^{**} \pi'$ if $\pi = C([\pi, \pi'])$.

Step 8. $\succsim^{**}$ is acyclic; that is, there is no sequence $\pi^1, \ldots, \pi^m$ such that $\pi^i \succsim^{**} \pi^{i+1}$ for each $i \leq m-1$ and $\pi^m \succ^{**} \pi^1$.

Take any $\pi^1, \ldots, \pi^m$ such that $\pi^i \succsim^{**} \pi^{i+1}$ for each $i \leq m-1$. Note that $\pi^i \succsim^{**} \pi^{i+1}$ implies $\pi^i \succ^* \pi^{i+1}$. Since $\succ^*$ is complete and transitive, we have $\pi^1 \succ^* \pi^1$. Hence, $\neg \pi^m \succ^{**} \pi^1$.

Step 9. There is a function $d : \Delta(S) \to [0, 1]$ such that (i) $d(\pi) \leq d(\pi')$ if $C(\pi, \pi') = \pi$, and (ii) $d(\pi) < d(\pi')$ if $C([\pi, \pi']) = \pi$.

Step 9.1. Let us construct a countable subset Z of $\Delta(S)$ as follows.

Take any k. Since $\cup_{\pi\in\Delta(S)} B(\pi, \frac{1}{k}) = \Delta(S)$, and $\Delta(S)$ is compact, there is a finite sequence $\pi^{1,k},\ldots,\pi^{m_k,k}$ such that $\cup_{i=1}^{m_k} B(\pi^{i,k},\frac{1}{k}) = \Delta(S)$. Let $C(B(\pi^i,\frac{1}{k})) = z^{i,k}$. Note that by Extremeness and Step 6, $z^{i,k} \neq \pi^i$, consequently, $\pi^i \succ^{**} z^{i,k}$.

We now construct Z_k for each k recursively. First, when $k=1$, let $Z_1 = \{z^{1,1},\ldots,z^{m_1,1}\}$. Now suppose we have constructed Z_j, and we will construct Z_{j+1}. Let $Z_{j+1} = Z_j \cup \{z^{1,j+1},\ldots,z^{m_{j+1},j+1}\} \cup C_{j+1}$ where $C_{j+1} = \bigcup_{z\in Z_j}\{C(B(z,\frac{1}{j+1}))\}$. Let $Z = \bigcup_{k=1}^{\infty} Z_k$. Since each Z_k has finite elements, Z is countable.

Step 9.2. For any $x,y \in \Delta(S)$ with $x \succ^{**} y$, then there is $z \in Z$ such that $x \succ^{**} z$ and $z \succ^{*} y$.

Take any k. Since $\cup_{i=1}^{m_k} B(\pi^{i,k},\frac{1}{k}) = \Delta(S)$, $x \in B(\pi^{i,k},\frac{1}{k})$ for some i. Therefore, either $x \succ^{**} z^{i,k} = C(B(\pi^{i,k},\frac{1}{k}))$ or $x = z^{i,k}$. When $x \succ^{**} z^{i,k} = C(B(\pi^{i,k},\frac{1}{k}))$, let $t^k = z^{i,k}$. Note that $t^k \in Z$. When $x = z^{i,k}$, let $t^k = C(z^{i,k},\frac{1}{k+1})$. By **Extremeness**, $t^k \neq z^{i,k}$. Note that $x = z^{i,k} \succ^{**} t^k$ and $t^k \in Z$.

Therefore, we have constructed $t^k \in Z$ such that $x \succ^{**} t^k$ and $\|x - t^k\| \leq \frac{1}{k}$. If $C(\{t^k,y\}) = t^k$ for each k, then by Continuity, we have $C(\{x,y\}) = x$; i.e., we have $y \succ^{*} x$, a contradiction. Therefore, $C(\{t^k,y\}) = y$ for some k. Therefore, we have $x \succ^{**} t^k$ and $t^k \succ^{*} y$ for some $t^k \in Z$.

Step 9.3. Since Z is countable, there is a utility function $d : Z \to [0,1]$ such that for any $z,z' \in Z$,

$$d(z) > d(z') \text{ if } z \succ^{*} z'.$$

Step 9.4. We now extend d to $\Delta(S)$ as follows.

For any $x \in \Delta(S)\setminus Z$, let $d(x) = \sup\{d(z) | x \succ^{*} z \text{ for any } z \in Z\}$.

Step 9.5. We now show that $d : \Delta(S) \to [0,1]$ satisfies the following property: for any $x,x' \in \Delta(S)$,

$$d(x) > d(x') \text{ if } x \succ^{**} x' \text{ and } d(x) \geq d(x') \text{ if } x \succsim^{*} x'.$$

Case 1. Take any x, x' with $x \succsim^* x'$. If $x, x \in Z$, then by Step 2.3, the desired condition is satisfied. If $x' \in Z$ and $x \notin Z$, then $d(x) = \sup\{d(z)|x \succ^* z \text{ and } z \in Z\} \geq d(x')$ since $x \succ^* x'$. If $x \in Z$ and $x' \notin Z$, then $d(x) > d(z)$ for any $z \in Z$ with $x' \succ^* z$ since $x \succ^* x' \succ^* z$. Therefore, $d(x) \geq \sup\{d(z)|x' \succ^* z \text{ and } z \in Z\}$. Finally, suppose $x, y \in \Delta(S) \setminus Z$. For any $z \in Z$, by transitive of $\succsim^*$, $x' \succ^* z$ implies $x \succ^* z$. Therefore, $\{z|x' \succ^* z\} \subseteq \{z|x \succ^* z\}$. Hence, $d(x) = \sup\{d(z)|x \succ^* z \text{ for any } z \in Z\} \geq d(x') = \sup\{d(z)|x' \succ^* z \text{ for any } z \in Z\}$.

Case 2. Take any x, x' with $x \succ^{**} x'$. We shall show that $d(x) > d(x')$.

By Step 2.2, there is $z \in Z$ such that $x \succ^{**} z$ and $z \succ^* x'$. By applying Step 2.2 on x and z, we have $z' \in Z$ such that $x \succ^{**} z'$ and $z' \succ^* z$. Then by Case 1 and Step 2.3, we have $d(x) \geq d(z') > d(z) \geq d(x')$.

Step 9. Therefore, $\mu_I = \arg\min T\big(\arg\min d(I)\big)$.

Since $\mu_I = C(I)$, $\pi \succ^* \mu$ for any $\pi \in I \setminus \{\mu_I\}$. By the construction of d, we have $d(\mu_I) \leq d(\pi)$ for any $\pi \in I$. Hence, $\mu_I \in \arg\min d(I)$. Therefore, since $\hat{d}$ represents $\succ^*$ and $\mu_I \in \arg\min d(I)$, $\mu_I = \arg\min \hat{d}\big(\arg\min d(I)\big)$.

Step 10. d is a distance function and locally-nonsatiated with respect to μ. By the construction of d, $\hat{d}$ is a continuous tie-breaker function of d.

For any $\pi \in \Delta(S) \setminus \{\mu\}$, since $\mu = C(\Delta(S))$ and $[\pi, \mu] \subset \Delta(S)$, we have $\pi \succ^{**} \mu$. By the construction of d, we have $d(\pi) > d(\mu)$ for any $\pi \in \Delta(S) \setminus \{\mu\}$.

Take any distinct $\pi, \pi' \in \Delta(S)$ with $d(\pi) = d(\pi)'$. Let $I = [\pi, \pi'] = \{\alpha\pi + (1 - \alpha)\pi' | \alpha \in [0, 1]\}$. By the definitions of $\succsim^{**}$, if $\mu_I = \pi$, then we have $u(\pi') > u(\pi)$, which contradicts the assumption that $d(\pi) = d(\pi')$. Similarly, if $\mu_I = \pi'$, then we have $d(\pi) > d(\pi')$, which contradicts the assumption that $d(\pi) = d(\pi')$. Finally, suppose $\mu_I = \alpha\pi + (1 - \alpha)\pi'$ for some $\alpha \in (0, 1)$. By Step 4, we have $\mu_I = \mu_{[\mu_I, \pi]}$ since $[\mu_I, \pi] \subseteq I$. Therefore, $d(\pi) > d(\mu_I) = d(\alpha\pi + (1 - \alpha)\pi')$.

Now take any $\pi \neq \mu$ and $\epsilon > 0$. If $\mu \in B(\pi, \frac{\epsilon}{2})$, then μ is the desired alternative; i.e., $\|\mu - \pi\| < \epsilon$ and $d(\mu) < d(\pi)$. If $\mu \notin B(\pi, \frac{\epsilon}{2})$, then $\pi^* = C(B(\pi, \frac{\epsilon}{2}))$ is different from π and μ by Extremeness. Hence, we found π^* such that $d(\pi) > d(\pi^*)$ and $\|\pi - \pi^*\| < \epsilon$.

We now write d_μ and $\hat{d}_\mu$.

Step 11. For any convex information set I, $\mu_I = \arg\min d_\mu(I)$.

Since $\mu_I = \arg\min \hat{d}_\mu(\arg\min d_\mu(I))$ by Step 9, we have $\mu_I \in \arg\min d_\mu(I)$. Hence, it is enough to show that $\arg\min d_\mu(I)$ is singleton. By way of contradiction, suppose there are distinct $\pi, \pi' \in \arg\min d_\mu(I)$. Note that $\pi, \pi' \in \arg\min d_\mu(I)$ implies $d_\mu(\pi) = d_\mu(\pi')$. By the definition of distance preference relations, there is some $\alpha \in (0, 1)$ such that $d_\mu(\pi) > d_\mu(\alpha\pi + (1-\alpha)\pi')$. Since I is convex, we have $\alpha\pi + (1-\alpha)\pi' \in I$. Therefore, $d_\mu(\pi) > d_\mu(\alpha\pi + (1-\alpha)\pi')$ contradicts the assumption that $\pi \in \arg\min d_\mu(I)$.

Step 12. $(d_\mu, \hat{d}_\mu)$ is upper semicontinuous.

Take any sequences $\{\pi^k\}, \{\overline{\pi}^k\}$ with $\pi^k \to \pi$ and $\overline{\pi}^k \to \overline{\pi}$. Note that if $(d_\mu(\pi^k), \hat{d}_\mu(\pi^k)) > (d_\mu(\overline{\pi}^k, \mu), \hat{d}_\mu(\overline{\pi}^k))$ for every $k \in \mathbb{N}$, then $C(\{\pi^k, \overline{\pi}^k\}) = \overline{\pi}^k$ for each k. By Continuity, we have $C(\{\pi, \overline{\pi}\}) = \overline{\pi}$; i.e., $(d_\mu(\pi), \hat{d}_\mu(\pi)) > (d_\mu(\overline{\pi}), \hat{d}_\mu(\overline{\pi}))$.

Necessity. Suppose the family of preference relations $\{\succsim_I\}_{I \in \mathscr{I}}$ admits the MDSEU representation with $(u, d_\mu, \hat{d}_\mu)$ where d_μ is a locally nonsatiated distance function, $\hat{d}_\mu$ is a continuous tie-breaker of $\hat{d}_\mu$, and $(d_\mu, \hat{d}_\mu)$ is upper semicontinuous.

For each $I \in \mathscr{I}$, $\succsim_I$ satisfies **SEU Postulates** since it has a SEU representation. Moreover, **Invariant Risk Preference** is satisfied since all SEU representations for $\{\succsim_I\}_{I \in \mathscr{I}}$ use the same Bernoulli utility function u.

Compliance. Take any $f \in F$ and $\pi \in \Delta(S)(S)$. Since $\mu_{\{\pi\}} = \pi$, the expected utility of f is $\sum_{s \in S} \pi(s)\, u(f(s))$ and the utility of $\pi(f)$ is $u(\pi(f)) = \sum_{s \in S} \pi(s)\, u(f(s))$. Therefore, $f \sim_\pi \pi(f)$.

Responsiveness. Take any $I, I' \in \mathscr{I}$ with $\succsim_I = \succsim_{I'}$. By the uniqueness of μ_I and $\mu_{I'}$ in SEU, $\succsim_I = \succsim_{I'}$ implies that $\mu_I = \mu_{I'}$. Since $\mu_I \in I$ and $\mu_{I'} \in I'$, we have $\mu_I \in I \cap I'$. Hence, $I \cap I' \neq \emptyset$.

Informational Betweenness. Take any $I_1, I_2, I_3 \in \mathscr{I}$ such that $I_3 \subseteq I_2 \subseteq I_1$ and $\succsim_{I_1} = \succsim_{I_3}$. Note that $\succsim_{I_1} = \succsim_{I_3}$ implies that $\mu_{I_1} = \arg\min \hat{d}_\mu(\arg\min d_\mu(I_1)) = \arg\min \hat{d}_\mu(\arg\min d_\mu(I_3)) = \mu_{I_3}$. Hence, $\mu_{I_1} \in I_3$. Since $I_3 \subseteq I_2$, $\mu_{I_1} \in I_2$. Therefore, $\mu_{I_1} = \arg\min \hat{d}_\mu(\arg\min d_\mu(I_2))$ since $\mu_{I_1} = \arg\min \hat{d}_\mu(\arg\min d_\mu(I_1, \mu))$ and $\mu_{I_1} \in I_2 \subseteq I_1$. Hence, $\succsim_{I_1} = \succsim_{I_2}$.

Extremeness. Take any convex $I, I' \in \mathscr{I}$ with $I' \subseteq \text{int}(I)$ and $\succsim_I \neq \succsim$. Note that $\succsim_I \neq \succsim$ is equivalent to $\mu \notin I$. We need to show that $\mu_I \neq \mu_{I'}$, which is equivalent to $\mu_I \notin I'$ since $I' \subset I$. Hence, we shall show that $\mu_I \in I \setminus I'$. Since $I' \subseteq \text{int}(I)$, it it sufficient to show that $\mu_I \in \partial I$.

By way of contradiction, suppose $\mu_I \in \text{int}(I)$. Then there is $\epsilon > 0$ such that $B(\mu_I, \epsilon) \subset \text{int}(I)$. By local nonsatiation of d_μ and $\mu \notin I$, there is $\mu' \in B(\mu_I, \epsilon)$ such that $d_\mu(\pi') < d_\mu(\mu_I)$, which contradicts the fact that $\mu_I = \arg\min d_\mu(I)$.

Continuity. Take any two sequences $\{I_k\}, \{J_k\}$ in $\mathscr{I}$ such that $I_k \to I$, $J_k \to J$, and $\succsim_{I_k} = \succsim_{J_k}$ for each k. Note that $\succsim_{I_k} = \succsim_{J_k}$ is equivalent to $\mu_{I_k} = \mu_{J_k}$. Hence, it is sufficient to show that $\mu_{I_k} \to \mu_I$ and $\mu_{J_k} \to \mu_J$.

Take any $\pi \in I$. Then there is a sequence $\{\pi^k\}$ such that $\pi^k \in I_k$ and $\pi^k \to \pi$. Since $\mu_I = \arg\min \hat{d}_\mu(\arg\min d_\mu(I_k))$ and $\pi^k \in I_k$, we have $(d_\mu(\mu_{I_k}), \hat{d}_\mu(\mu_{I_k})) > (d_\mu(\pi^k), \hat{d}_\mu(\pi^k))$ for each k. By upper semicontinuity, $(d_\mu(\mu_I), \hat{d}_\mu(\mu_I)) > (d_\mu(\pi), \hat{d}_\mu(\pi))$. Hence π cannot be chosen over μ_I from I for any $\pi \in I \setminus \{\mu_I\}$. Hence, $\mu_{I_n} \to \mu_I$.

Finally, we show that our model is well-defined as long as $\hat{d}_\mu$ is a continuous tie-breaker of d_μ. First, we show that $\arg\min \hat{d}_\mu(\arg\min d_\mu(I))$ is nonempty. Since $\hat{d}_\mu$ is continuous and I is closed, by the Weierstrass theorem, $\arg\min \hat{d}_\mu(I)$ is not empty. Since $\hat{d}_\mu$ is injective, $\arg\min \hat{d}_\mu$ is singleton. Let $\mu_I = \arg\min \hat{d}_\mu(I)$. Since $\hat{d}_\mu$ is a tie-breaker function, we have $\mu_I = \arg\min \hat{d}_\mu(I) \subseteq \arg\min d_\mu(I)$. Therefore, $\arg\min \hat{d}_\mu(\arg\min d_\mu(I)) = \mu_I$.

Suppose now I is convex. By the above argument, we have $\mu_I \in \arg\min d_\mu(I)$. To show that $\mu_I \in \arg\min d_\mu(I)$, by way of contradiction, suppose there are distinct π, π' such that $\pi, \pi' \in \arg\min d_\mu(I)$. Since $d_\mu(\pi) = d_\mu(\pi')$ and d_μ is a distance function, there is an $\alpha \in (0,1)$ such that $d_\mu(\alpha\pi + (1-\alpha)\pi') < d_\mu(\pi)$, which contradicts with the fact that $\pi \in \arg\min d_\mu(I)$.

A.2 Proof of Proposition 1

The uniqueness of u and μ_I are standard. Since $\mu = \mu_{\Delta(S)}$, μ is also unique. We now show that $\hat{d}_\mu$ and $\hat{d}'_{\mu'}$ are ordinally equivalent. As we proved in the necessity part of Theorem 1, $\mu_I = \arg\min \hat{d}_\mu(I)$. Since μ_I is unique, we have $\arg\min \hat{d}_\mu(I) = \arg\min \hat{d}'_{\mu'}(I)$. Take any $\pi, \pi' \in \Delta(S)$. Note that $\hat{d}_\mu(\pi) > \hat{d}_\mu(\pi')$ is equivalent to $\pi' = \arg\min \hat{d}_\mu(I)$ where $I = \{\pi, \pi'\}$. Then $\hat{d}_\mu(\pi) > \hat{d}_\mu(\pi')$ is equivalent to $\pi' = \arg\min \hat{d}'_{\mu'}(I)$, which is equivalent to $\hat{d}'_{\mu'}(\pi) > \hat{d}_{\mu'}(\pi')$. Therefore, $\hat{d}_\mu(\pi) > \hat{d}_\mu(\pi')$ if and only if $\hat{d}'_{\mu'}(\pi) > \hat{d}_{\mu'}(\pi')$.

A.3 Proof of Proposition 2

Proof. Take $f, p, h \in \mathcal{F}$ such that $fEh \sim pEh$ and $fEh \sim_{I_E^\alpha} pEh$. By the MDSEU representation:

$$fEh \sim pEh \text{ if and only if } \sum_{s\in E} \mu(s)u(f(s)) = \mu(E)u(p) \text{ and}$$
$$fEh \sim_{I_E^\alpha} pEh \text{ if and only if } \sum_{s\in E} \mu_I(s)u(f(s)) = \mu_I(E)u(p).$$

By **Informational Dynamic Consistency**, we thus have

$$\frac{1}{\mu(E)}\sum_{s\in E}\mu(s)u(f(s)) = \frac{1}{\mu_I(E)}\sum_{s\in E}\mu_I(s)u(f(s))$$

which is equivalent to

$$\frac{\mu(s)}{\mu(E)} = \frac{\mu_I(s)}{\mu_I(E)} = \frac{\mu_I(s)}{\alpha}$$

for each $s \in E$. □

A.4 Proof of Proposition 3

Proof. Without loss of generality, suppose $\mu_1 = \min_i \mu_i$. Take any $\alpha \in [0,1]$ and i. Let $I = I_E^\alpha$ where $E = \{s_1, s_i\}$. Consider the minimum distance updating with I:

$$\min_{\pi} d_1(\pi_1) + d_i(\pi_i) \text{ s.t. } \pi_1 + \pi_i = \alpha.$$

The first order condition gives $d_i'(\pi_i) = d_1'(\pi_1)$. The second order condition is satisfied since $d_i'' > 0$. Since Generalized Dynamic Consistency is satisfied, we have $\pi_i = \alpha \frac{\mu_i}{\mu_i+\mu_1}$. Then $d_i'(\frac{\mu_i}{\mu_i+\mu_1}\alpha) = d_1'(\frac{\mu_1}{\mu_1+\mu_i}\alpha)$. Equivalently,

$$d_i'(\pi) = d_1'(\frac{\mu_1}{\mu_i}\pi) \text{ for any } \pi \in [0, \frac{\mu_i}{\mu_i+\mu_1}].$$

Since $\frac{\mu_i}{\mu_i+\mu_1} \geq \frac{1}{2}$, the above holds for any $\pi \leq \frac{1}{2}$. Note that

$$d_i(\pi) - d_i(0) = \int_0^\pi d_i'(\tilde{\pi})d\tilde{\pi} = \int_0^\pi d_1'(\frac{\mu_1}{\mu_i}\tilde{\pi})d\tilde{\pi} = \frac{\mu_i}{\mu_1}\int_0^{\frac{\mu_1}{\mu_i}\pi} d_1'(\bar{\pi})d\bar{\pi} = \frac{\mu_i}{\mu_1}\big(d_1(\frac{\mu_1}{\mu_i}\pi) - d_1(0)\big).$$

Therefore, $d_i(\pi) = \frac{\mu_i}{\mu_1} d_1(\frac{\mu_1}{\mu_i}\pi) + c_i$ where $c_i = d_i(0) - \frac{\mu_i}{\mu_1} d_1(0)$. Then the distance function is given by

$$d_\mu(\pi) = \sum_{i=1}^n d_i(\pi_i) = \sum_{i=1}^n \frac{\mu_i}{\mu_1} d_1(\frac{\mu_1}{\mu_i}\pi_i) + \sum_{i=1}^n c_i \text{ where } \pi_i \in [0, \frac{1}{2}].$$

Let $f(t) = \frac{d_1(\mu_1 t)}{\mu_1} + \frac{\sum c_i}{n}$. Then we have $d_\mu(\pi) = \sum_{i=1}^n \mu_i f(\frac{\pi_i}{\mu_i})$. □

A.5 Proof of Proposition 4

Consider a Hypothesis Testing representation (μ, ρ, ϵ) defined on the finite state space S.

Proof. We first establish some notation. Let $\pi_A^\rho = \arg\max_{\pi \in \Delta(S)} \rho(\pi)\pi(A)$ for any $\rho \in \Delta(\Delta(S))$ such that $\rho(\pi) \neq \rho(\pi')$ for any distinct $\pi, \pi' \in \Delta(S)$. We denote the support of π by $sp(\pi)$. Without loss of generality, suppose that $sp(\pi) = S$ for any $\pi \in sp(\rho) \setminus \{\mu\}$.

For any $B \subseteq S$, let

$$f(B) = -\sum_{s \in B} \mu(s) \log(\frac{\mu(s)}{\mu(B)}) \text{ when } \mu(B) > 0$$

and

$$g(B) = -\sum_{s \in B} \pi_B^\rho(s) \log(\frac{\pi_B^\rho(s)}{\pi_B^\rho(B)}).$$

Note that $0 \leq f(B), g(B) < +\infty$. Let

$$M = \max_{B, B' \subseteq S} \{|f(B) - f(B')|, |f(B) - g(B')|, |g(B) - g(B')|\} + 1.$$

Now, consider the following distance:

$$d_\mu(\pi) = \begin{cases} -\sum_{s\in sp(\pi)} \mu(s)\log(\pi(s)) - M\,|\{sp(\pi)\cap sp(\mu)\}| \text{ if } \mu(sp(\pi)) > \epsilon, \\ -\sum_{s\in sp(\pi)} \pi^\rho_{sp(\pi)}(s)\ \log(\pi(s)) + M(|S|+1-|sp(\pi)|) \text{ if } \mu(sp(\pi)) \le \epsilon. \end{cases}$$

Then for any $A \subseteq S$ and $s \in A$,

$$\mu_{I^1_A}(s) = \begin{cases} \frac{\mu(s)}{\mu(A)} & \text{when } \mu(A) > \epsilon, \\ \frac{\pi^\rho_A(s)}{\pi^\rho_A(A)} & \text{when } \mu(A) \le \epsilon. \end{cases}$$

To demonstrate this, we consider any $A \subseteq S$ and derive $\mu_{I^1_A}$. There are two cases.

Case 1. Suppose $\mu(A) \le \epsilon$.

For any $\pi \in I^1_A$, since $\pi(A) = 1$, we have $sp(\pi) \subseteq A$. Therefore, $\mu(sp(\pi)) \le \mu(A) \le \epsilon$. Hence, we have

$$d_\mu(\pi) = -\sum_{s\in sp(\pi)} \pi^\rho_{sp(\pi)}(s)\ \log(\pi(s)) + M(|S|+1-|sp(\pi)|) \text{ for any } \pi \in I^1_A.$$

Take any $B \subseteq A$. For any $\pi \in I^1_A$ with $sp(\pi) = B$,

$$d_\mu(\pi) = -\sum_{s\in B} \pi^\rho_B(s)\ \log(\pi(s)) + M(|S|+1-|B|).$$

Since $M(|S|+1-|B|)$ is fixed for given B, the above distance function leads to Bayesian posterior μ^B such that $\mu^B(s) = \frac{\pi^\rho_{sp(\pi)}(s)}{\pi^\rho_{sp(\pi)}(B)}$ for any $s \in B$. Hence, $d_\mu(\mu^B) = g(B) + M(|S|+1-|B|)$. Note that if $B \subset A$, then $d_\mu(\mu^B) > d_\mu(\mu^A)$ since $g(B) + M(|S|+1-|B|) > g(A) + M(|S|+1-|A|)$, which is equivalent to $M(|A|-|B|) \ge g(A) - g(B)$ and by the definition of M, we have $M(|A|-|B|) \ge M > |g(A)-g(B)| \ge g(A) - g(B)$. Therefore, μ^A minimizes $d_\mu(\pi)$ subject to $\pi \in I^1_A$. Hence, $\mu_{I^1_A}(s) = \mu^A(s) = \frac{\pi^\rho_A(s)}{\pi^\rho_A(A)}$.

Case 2. Suppose $\mu(A) > \epsilon$. Take any $\pi \in I_A^1$.

Case 2.1. $\mu(sp(\pi)) \leq \epsilon$.

By the argument for $\mu(A) \leq \epsilon$, $d_\mu(\mu^A) \leq d_\mu(\pi)$ for any $\pi \in I_A^1$ with $\mu(sp(\pi)) \leq \epsilon$. Moreover, $d_\mu(\mu^A) = g(A) + M(|S| + 1 - |A|) \geq g(A) + M$. Hence, $g(A) + M \leq d_\mu(\pi)$ for any $\pi \in I_A^1$ with $\mu(sp(\pi)) \leq \epsilon$. We now show that there is a $\pi \in I_A^1$ such that $d_\mu(\pi) < g(A) + M$.

Let π^A be a Bayesian posterior such that $\pi^A(s) = \frac{\mu(s)}{\mu(A)}$ for any $s \in A$. Then $sp(\pi^A) = sp(\mu) \cap A$. Hence, $\mu(sp(\pi^A)) = \mu(sp(\mu) \cap A) = \mu(A) > \epsilon$. Therefore,

$d_\mu(\pi^A) = -\sum_{s \in sp(\pi^A)} \mu(s) \log(\frac{\mu(s)}{\mu(A)}) - M\,|sp(\pi^A) \cap sp(\mu)|$. Moreover,

$$d_\mu(\pi^A) = -\sum_{s \in sp(\pi^A)} \mu(s) \log(\frac{\mu(s)}{\mu(A)}) - M|sp(\pi^A) \cap sp(\mu)| \leq -\sum_{s \in A} \mu(s) \log(\frac{\mu(s)}{\mu(A)}) = f(A)$$

since $sp(\pi^A) \subseteq A$ and $\mu(s) \log(\frac{\mu(s)}{\mu(A)}) \leq 0$ for each s. Hence, $d_\mu(\pi^A) \leq f(A) < g(A) + M$ by the definition of M.

Case 2. $\mu(sp(\pi)) > \epsilon$.

In ther case, we have $d_\mu(\pi) = -\sum_{s \in sp(\pi)} \mu(s) \log(\pi(s)) - M\,|\{sp(\pi) \cap sp(\mu)\}|$. Take any $B \subseteq A \cap sp(\mu)$. Take any $\pi \in I_A^1$ such that $sp(\pi) = B$. Since $|\{sp(\pi) \cap sp(\mu)\}| = |B|$,

$$d_\mu(\pi) = -\sum_{s \in B} \mu(s) \log(\pi(s)) - M\,|B|.$$

When B is fixed, the above leads to Bayesian posterior π^B such that $\pi^B(s) = \frac{\mu(s)}{\mu(B)}$ for any $s \in B$. In other words, π^B minimizes $d_\mu(\pi)$ subject to $sp(\pi) = B$. Hence, we obtain $d_\mu(\pi^B) = f(B) - M\,|B|$.

By the definition of M, if $B \subset A \cap sp(\mu)$, then

$$d_\mu(\pi^B) = f(B) - M\,|B| > d_\mu(\pi^{A \cap sp(\mu)}) = f(A \cap sp(\mu)) - M\,|A \cap sp(\mu)|.$$

Hence, $\pi^{A\cap sp(\mu)}$ minimizes $d_\mu(\pi)$ subject to $\mu \in I_A^1$. Finally, note that $\pi^{A\cap sp(\mu)} = \pi^A$ since $\mu(A \cap sp(\mu)) = \mu(A)$ and $\pi^A(s) = \frac{\mu(s)}{\mu(A)} = \frac{\mu(s)}{\mu(A\cap sp(\mu))} = \pi^{A\cap sp(\mu)}(s)$ for each $s \in A$. Therefore,

$$\mu_{I_A^1}(s) = \pi^A(s) = \frac{\mu(s)}{\mu(A)}.$$

□

A.6 Proof of Proposition 5

Proof. Take any $i \geq 2$ and $\alpha \in [0,1]$. Note that $\pi_i = g_i(\alpha)$ solves the following optimization problem:

$$\min_{\pi\in[0,\alpha]} f_i(\pi_i) + f_1(\alpha - \pi_i).$$

The first order condition gives $f_i'(g_i(\alpha)) = f_1'(\alpha - g_i(\alpha))$. Let $t = g_i(\alpha) \in [0, g_i(1)]$. Then $\alpha = g_i^{-1}(t)$. Then we have $f_i'(t) = f_1'(g_i^{-1}(t) - t)$. Hence, $f_i(\pi_i) = \int_0^{\pi_i} f_1'(g_i^{-1}(t) - t) + f_i(0)$. Finally,

$$d_\mu(\pi) = f_1(\pi_1) + \sum_{t=2}^{n} \int_0^{\pi_i} f_1'(g_i^{-1}(t) - t)dt + \sum_{i=2}^{n} f_i(0) \text{ for any } \pi \in [0,1] \times \prod_{i=2}^{n}[0, g_i(1)].$$

Let $f(\pi) = f_1(\pi) + \sum_{i=2}^{n} f_i(0)$. Then we have the desired result. □

B Preference Differences Lead to Distance Functions

Under minimum distance updating, the DM selects a posterior belief that is "as close as possible" to her prior belief. Another interpretation of minimum distance updating is that the DM is choosing a posterior belief that requires a minimal change in behavior, which is captured by the changes in her preference over acts.[18] Taking this view, any function that measures the difference between two preferences will lead to its own distance function d_μ. We consider the following two cases.

Ordinal Preferences: In many contexts, preferences only have ordinal meaning. Hence, a natural way to measure the difference between two preference relations is to simply count the number of times the preferences rank objects differently. For example, consider $d(\succsim_I, \succsim) = \#\{f, g \in F | f \succsim g \text{ and } g \succ_I f\}$. This function calculates the number of pairs (f, g) such that f is preferred to g according to the ex-ante preference relation, but g is preferred to f according to the ex-post preference relation under I.

Example 1. Suppose $\hat{F} = \{pEw | E \subseteq S \text{ and } p \in \Delta(X)\}$, where w is the worst outcome in X. Then we obtain $d_\mu(\pi) = \sum_{E,E' \subseteq S} d(E, E')$ where

$$d(E, E') = \begin{cases} \frac{1}{2f(E,E')} & \text{if } f(E, E') = \min\{\frac{\pi(E)}{\pi(E')}, \frac{\mu(E)}{\mu(E')}\} \geq 1 \\ 1 - \frac{f(E,E')}{2} & \text{if } f(E, E') < 1. \end{cases}$$

Cardinal Preference: In SEU, preferences have some cardinal meaning. Hence, we can calculate the difference between SEU preferences based on utility differences. For example, consider $d(\succsim_I, \succsim) = \int_{\mathcal{F}} \rho(\mathbb{E}u_\mu(f), \mathbb{E}u_\pi(f))df$ where ρ is a function that measures the difference between expected utilities.

Example 2. Let $F^* = \{p\{s\}w | s \in S \text{ and } p \in \Delta(X)\}$ and suppose ρ is homogenous of degree k. Then we will obtain a distance function $d_\mu(\pi) = \sum_{i=1}^n (\mu_i)^k f(\frac{\pi}{\mu_i})$. [19]

This distance function simultaneously generalizes Bayesian distance functions and Euclidean distance functions. In particular, when $k = 1$, we obtain a Bayesian distance

[18] Indeed, if we consider the initial preference a representation of behavior, then the idea of "behavioral inertia" suggests that a DM will only make minimal changes.

[19] Since $\rho(a, b) = (a)^k \rho(1, \frac{b}{a})$, we obtain $d_\mu(\pi) = \sum_{i=1}^n (\mu_i)^k f(\frac{\pi}{\mu_i})$ where $f(\frac{b}{a}) = c\,\rho(1, \frac{b}{a})$ and $c = \int_X (u(x))^\rho dx$.

function $d_\mu(\pi) = \sum_{i=1}^{n} \mu_i \, f(\frac{\pi}{\mu_i})$. When $k = 2$ and $f(t) = (1-t)^2$, we obtain the Euclidean distance function $d_\mu(\pi) = \sum_{i=1}^{n}(\mu_i - \pi_i)^2$.

References

Ahn, D. S. (2008): "Ambiguity Without a State Space," *Review of Economic Studies*, 75, 3–28.

Alchourrón, C. E., P. Gärdenfors, and D. Makinson (1985): "On the logic of theory change: Partial meet contraction and revision functions." *The Journal of Symbolic Logic*, 50.

Anscombe, F. and R. Aumann (1963): "A Definition of Subjective Probability," *Annals of Mathematical Statistics*, 34, 199–205.

Baliga, S., E. Hanany, and P. Klibanoff (2013): "Polarization and Ambiguity," *American Economic Review*, 103, 3071–3083.

Basu, P. (2019): "Bayesian updating rules and AGM belief revision," *Journal of Economic Theory*, 179, 455 – 475.

Benjamin, D. J. (2019): "Errors in probabilistic reasoning and judgment biases," in *Handbook of Behavioral Economics: Applications and Foundations*, ed. by B. D. Bernheim, S. DellaVigna, and D. Laibson, North-Holland, vol. 2, chap. 2, 69–186.

Blume, L., A. Brandenburger, and E. Dekel (1991): "Lexicographic Probabilities and Choice under Uncertainty," *Econometrica*, 59, 61–79.

Chambers, C. P. and T. Hayashi (2010): "Bayesian consistent belief selection," *Journal of Economic Theory*, 145, 432–439.

Damiano, E. (2006): "Choice under Limited Uncertainty," *Advances in Theoretical Economics*, 6.

Dominiak, A. and J.-P. Lefort (2020): "Ambiguity and Probabilistic Information," *Management Science*, forthcoming.

Epstein, L. G. (2006): "An Axiomatic Model of Non-Bayesian Updating," *Review of Economic Studies*.

Epstein, L. G. and M. L. Breton (1993): "Dynamically Consistent Beliefs Must Be Bayesian," *Journal of economic theory*, 61, 1–22.

Epstein, L. G., J. Noor, and A. Sandroni (2008): "Non-Bayesian Updating: A Theoretical Framework," *Theoretical Economics*.

Gajdos, T., T. Hayashi, J.-M. Tallon, and J.-C. Vergnaud (2008): "Attitude toward imprecise information," *Journal of Economic Theory*, 140, 27–65.

Ghirardato, P. (2002): "Revisiting Savage in a conditional world," *Economic Theory*, 20, 83–92.

Gilboa, I., L. Samuelson, and D. Schmeidler (2014): "NO-BETTING-PARETO DOMINANCE," *Econometrica*, 82, 1405–1442.

Grether, D. M. (1980): "Bayes Rule as a Descriptive Model: The Representativeness Heuristic," *Quarterly Journal of economics*.

Gul, F. and W. Pesendorfer (2001): "Temptation and Self-Control," *Econometrica*.

Kahneman, D. and A. Tversky (1979): "Prospect theory: An analysis of decision under risk," *Econometrica: Journal of the Econometric Society*, 263–291.

Kovach, M. (2020): "Conservative Updating," *mimeo*.

Machina, M. and D. Schmeidler (1992): "A More Robust Definition of Subjective Probability," *Econometrica*, 60, 745–780.

Milgrom, P. and N. Stokey (1982): "Information, trade and common knowledge," *Journal of Economic Theory*, 26, 17 – 27.

Morris, S. (1994): "Trade with Heterogeneous Prior Beliefs and Asymmetric Information," *Econometrica*, 62, 1327–1347.

Myerson, R. B. (1986a): "Axiomatic Foundations of Bayesian Decision Theory," Discussion Papers 671, Northwestern University, Center for Mathematical Studies in Economics and Management Science.

——— (1986b): "Multistage Games with Communication," *Econometrica*, 54, 323–358.

Ok, E. A. and A. Savaockin (2020): "Choice Under Uncertainty with Initial Information," Tech. rep., working Paper.

Ortoleva, P. (2012): "Modeling the Change of Paradigm: Non-Bayesian Reactions to Unexpected News," *American Economic Review*, 102, 2410–36.

Perea, A. (2009): "A Model of Minimul Probabilistic Belief Revision," *Theory and Decision*, 163–222.

Rabin, M. and J. L. Schrag (1999): "First Impressions Matter: A Model of Confirmatory Bias," *Quarterly Journal of Economics*, 114, 37–82.

Zhao, C. (2020): "Pseudo-Bayesian Updating," *working paper.*