Inertial Updating*

Adam Dominiak Matthew Kovach Gerelt Tserenjigmid

July 13, 2022

Abstract: We introduce a model of inertial belief updating in a decision maker (DM), upon learning an event $E \subseteq S$, selects a posterior belief that minimizes the subjective distance between her prior and potential posteriors that assign probability one to event E. By varying the subjective distance between probability distributions, this model provides a unifying framework that nests three separate belief updating rules: (i) Bayesian updating, (ii) non-Bayesian updating rules such as the $\alpha-\beta$ rule (Grether, 1980), and (iii) updating rules for zero-probability events such as conditional probability systems of Myerson (1986a,b). We also show that our model is behaviorally equivalent to the Hypothesis Testing model (HT) of Ortoleva (2012).

Keywords: Inertial belief updating, subjective expected utility, Bayesian updating, non-Bayesian updating, generalized Bayesian divergence.

JEL: D01, D81.

*Dominiak: Virginia Tech (dominiak@vt.edu); Kovach: Virginia Tech (mkovach@vt.edu); Tserenjigmid: UC Santa Cruz (gtserenj @ucsc.edu). We are very grateful to David Freeman, Paolo Ghirardato, Faruk Gul, Edi Karni, Shaowei Ke, Yusufcan Masatlioglu, Burkhard Schipper, and Chen Zhao for valuable comments and discussions, as well as the seminar participants at University of Michigan; CUHK-HKU-HKUST Joint Theory Seminar; University of California, Santa Cruz; University of California, Riverside; Royal Holloway (University of London); Dauphine University; ETH Zürich; Western University, along with the audiences of BRIC (2022), the Conference on Unawareness and Unintended Consequences (2021), RUD (University of Minnesota, 2021), 20th SAET Conference (Seoul); D-TEA (PSE); International Conference on Social Choice and Voting Theory (Virginia Tech) and the Workshop in Experiments, Revealed Preferences, and Decisions (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, is has two major issues. First, it is *incomplete*; a well-known limitation of Bayesian updating is that it is not defined for zero-probability events.[1] Second, it is descriptively limited; there is robust experimental evidence that people's beliefs systematically deviate from what Bayesian updating prescribes.[2] We resolve these limitations of Bayesian updating by introducing the Inertial Expected Utility (**IEU**) representation: a complete theory of belief updating that unifies Bayesian and non-Bayesian updating rules.

The **IEU** addresses these two issues by recasting the problem of belief updating as a problem of belief selection subject to simple constraints. For each event E, our DM selects a new belief μ_E that is (i) consistent with E and (ii) closest to her initial belief. That is, her new belief μ_E is the element of $\Delta(E)$ that is "closest" to μ among all of the probability measures over E. Because our DM is influenced by her prior, beliefs may exhibit "inertia," we refer to this behavior as Inertial Expected Utility. We provide a complete behavioral analysis of **IEU** and demonstrate that it provides a unifying framework for understanding many different forms of belief updating in the literature.

The **IEU** representation depends on three axioms. The first two postulates are standard: **SEU Postulates** imposes a subjective expected utility representation for each conditional preference, and **Consequentialism** ensures that for any event E, the DM only considers states within E possible. The third axiom, **Dynamic Coherence**, was introduced by Ortoleva (2012) to characterize the Hypothesis Testing model (HT).[3] To interpret this axiom, say that an event A is *revealed implied by* event B if every state that the DM believes is possible after learning B is also an element of A. That is, A^c is revealed to be null after B. **Dynamic Coherence** requires that this "revealed preference" over events is acyclic. Surprisingly, the **IEU** and HT require precisely the same axioms and thus are behaviorally equivalent despite their stark difference in appearance.[4]

Because our notion of updating is based on optimization, **IEU** allows us to naturally extend updating to zero-probability events, thereby offering a complete theory of updating. The **IEU**,

[1] This is an important issue in sequential games, as particular off-path beliefs are used to support certain equilibria. Accordingly, complete theories of belief updating, such as the Conditional Probability System introduced by Myerson (1986a,b), have been proposed.

[2] 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).

[3] 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.

[4] The proofs however are significantly different and deriving the **IEU** representation is nontrivial.

however, is not the first to address updating for zero-probabilty events. The most prominent such theory is Myerson's Conditional Probability System (CPS) (Myerson, 1986a,b), which was motivated by the Sequential Equilibria of Kreps and Wilson (1982). We provide a simpler behavioral foundation for CPS and clarify its relation to HT. We show that CPS is a special of **IEU** and therefore it is also a special case of HT. We also explicitly construct the subjective distance functions used in both CPS and HT.

The **IEU** may also accommodate non-Bayesian updating, a feature that sets it apart from the CPS. In particular, we introduce a form of distorted belief updating that we call ***h*-Bayesian** and provide an axiomatic characterization of this form of updating. This novel updating rule has a non-trivial connection to the well-known $\alpha - \beta$ rule from Grether (1980). Further, this rule allows for history-dependent updating and therefore it can capture a wide array of context effects.

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 **IEU** representation in section 3. In section 4 we discuss the connection to CPS and HT. We then show in section 5 how to extend the **IEU** to settings with signal structures and provide an explicit distance function that generates $\alpha - \beta$ rule from Grether (1980). We conclude with a discussion of related literature in section 6.

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 $S = \{s_1, \dots, s_n\}$. A nonempty subset E of S is called an event, and Σ is a algebra of events. 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 $\mathcal{F} := \{f : S \to \Delta(X)\}$. A preference relation over $\mathcal{F}$ is denoted by $\succsim$. As usual, $\succ$ and $\sim$ are the asymmetric and symmetric parts of $\succsim$, respectively.

The DM's behavior is depicted by a family $\{\succsim_E\}_{E \in \Sigma}$ of preference relations, each defined over $\mathcal{F}$. We write $\succsim$ in place of $\succsim_S$, and we call $\succsim$ the initial preference. Say that E is $\succsim$-null (or simply null) if $fEg \sim g$ for any $f, g \in \mathcal{F}$. Otherwise, E is non-null. Similarly, we say E is $\succsim_A$-null if $fEg \sim_A g$ for any $f, g \in \mathcal{F}$.

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. For any μ and event E such that $\mu(E) > 0$, let $\text{BU}(\mu, E)$ denote the Bayes' updating of μ conditional on E.

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 Inertial Updating

As a new piece of information $E \in \Sigma$ emerges, the DM revises $\succsim$ given E. The new preference is denoted by $\succsim_E$ and governs the DM's conditional choice in light of E.

Definition 1 (Distance Function)**.** A function $d : \Delta(S) \to \mathbb{R}$ is a **distance function** with respect to $\mu \in \Delta(S)$, denoted by d_μ, if $d_\mu(\mu) < d_\mu(\pi)$ for any $\pi \in \Delta \setminus \{\mu\}$.

This property on distance function ensures that the current belief is unique, in that all different beliefs are in fact considered to be different.

Definition 2 (IEU). A family of preference relations $\{\succsim_E\}_{E \in \Sigma}$ admits an **Inertial Expected Utility** representation if there are a Bernoulli utility function $u : X \to \mathbb{R}$, a prior $\mu \in \Delta(S)$, a distance function $d_\mu : \Delta(S) \to \mathbb{R}$ such that for each $E \in \Sigma$, the preference relation $\succsim_E$ admits a SEU representation with (u, μ_E), meaning that for any $f, g \in \mathcal{F}$,

$$f \succsim_E g \quad \text{if and only if} \quad \sum_{s \in E} \mu_E(s) u\big(f(s)\big) \geq \sum_{s \in E} \mu_E(s) u\big(g(s)\big), \tag{1}$$

where

$$\mu_E \equiv \arg\min_{\pi \in \Delta(E)} d_\mu(\pi). \tag{2}$$

Since $\Delta(E)$ is convex, $\arg\min_{\pi \in \Delta(E)} d_\mu(\pi)$ will be unique when d_μ is strictly quasi-convex. In fact, the following much weaker condition will be suffice: for any $\pi, \pi' \in \Delta(S)$ with $\pi \neq \pi'$, if $d_\mu(\pi) = d_\mu(\pi')$, then there is $\alpha \in (0, 1)$ such that $d_\mu(\alpha\pi + (1-\alpha)\pi') < d_\mu(\pi)$.

2.3 Notions of Distance

Our DM's notion of distance is subjective. Thus, our framework allows for a wide variety distance notions and consequently a wide variety of updating behaviors. In this section, we discuss a few examples of distance functions.

Definition 3 (**Bayesian Divergence**)**.** For a strictly increasing and strictly concave function $\sigma : \mathbb{R}_+ \to \mathbb{R}$, let d_μ be given by

$$d_\mu(\pi) = -\sum_{i=1}^{n} \mu_i \, \sigma\left(\frac{\pi_i}{\mu_i}\right). \tag{3}$$

Proposition 1. *For any non-null $E \in \Sigma$,*

$$\mu_E = \underset{\pi \in \Delta(E)}{\arg\min} - \sum_{i=1}^{n} \mu_i \, \sigma\left(\frac{\pi_i}{\mu_i}\right) = BU(\mu, E)$$

Moreover, Equation 3 "includes" the KL divergence as a special case ($\sigma(x) = \ln(x)$). However, since $\ln(0) = -\infty$, the KL divergence is not well-defined when $\mathrm{sp}(\mu) \subsetneq \mathrm{sp}(\pi)$. Therefore, we focus our attention to σ that is well defined on $\mathbb{R}_+$. For example, $\sigma(x) = \ln(\alpha\, x + \beta)$ where $\alpha, \beta > 0$ is a well-defined, strictly increasing, and strictly concave function.

Using the intuition from **Bayesian Divergence**, we can introduce a "perturbed" version of this distance notion to capture non-Bayesian beliefs.

Definition 4 (**h-Bayesian**)**.** Let $d_\mu(\pi) = \sum_{i=1}^{n} h(\mu_i)\, \sigma(\frac{\pi_i}{h(\mu_i)})$, where $h : \mathbb{R}_+ \to \mathbb{R}$ and σ satisfies the conditions from **Bayesian Divergence**. Then $\mu_E = \mathrm{BU}(h(\mu), E)$ for any non-null $E \in \Sigma$.

The h-Bayesian distance notion captures a form of non-Bayesian updating where the agent is Bayesian with respect to biased beliefs. When $h(\mu_i) = (\mu_i)^\alpha$, this corresponds to a special case of Grether's $\alpha - \beta$ rule (Grether, 1980) where $\alpha = \beta$. For $\alpha < 1$, this captures under-reaction to information and base-rate neglect, while $\alpha > 1$ 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)). Since we only require h to be independent of the current information, the h-Bayesian distance can also capture features of history or reference dependence. In section 5, we show that our model nests the general version of Grether's $\alpha - \beta$ rule.

We can also use a support-dependent Bayesian divergence to obtain updating rules for zero-probability events. For any $\pi \in \Delta(S)$, let $sp(\pi)$ denote the support of π.

Definition 5 (**Support-Dependent Bayesian Divergence**)**.** Let

$$d_\mu(\pi) = \begin{cases} -\sum_{s_i \in sp(\pi)} \mu_i \, \sigma\left(\frac{\pi_i}{\mu_i}\right) - M\, |\{sp(\pi) \cap sp(\mu)\}| & \text{if } \mu(sp(\pi)) > 0, \\ -\sum_{s_i \in sp(\pi)} \mu_i^* \, \sigma\left(\frac{\pi_i}{\mu_i^*}\right) & \text{otherwise.} \end{cases}$$

Suppose μ^* has a full-support and $M > \sigma(1) - \sigma(0)$. Then

$$\mu_E = \begin{cases} \mathrm{BU}(\mu, E) & \text{if } \mu(E) > 0, \\ \mathrm{BU}(\mu^*, E) & \text{otherwise.} \end{cases}$$

The form belief updating rule above was used in Galperti (2019) and is a special case of Myerson (1986a,b) and Ortoleva (2012).

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

Definition 6 (Euclidean distance). Let $d_\mu(\pi) = ||\mu - \pi||$. Then

$$\mu_E(s) = \mu(s) + \frac{1 - \mu(E)}{|E|} \text{ for any } E \in \Sigma \text{ and } s \in E.$$

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

Axiom 1 (**SEU Postulates**). For each $E \in \Sigma$, the following conditions hold.

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

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

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

(*iv*) **Nontriviality:** There are $f, g \in \mathcal{F}$ such that $f \succ_E g$.

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

(*vi*) **Invariant Risk Preference:** For all lotteries $p, q \in \Delta(X)$, $p \succsim_E q$ if and only if $p \succsim q$.

The next axiom is standard and ensures that the DM forms a new belief that is consistent with the available information.

Axiom 2 (**Consequentialism**). For any $E \in \Sigma$ and all $f, g \in F$,

$$f(s) = g(s) \text{ for all } s \in E \implies f \sim_E g.$$

The next axiom, **Dynamic Coherence**, was introduced in Ortoleva (2012), and a more careful discussion may be found there. In our setting, we say that an event A is *revealed implied by* event B if every state that the DM believes is possible after learning B is also an element of A. **Dynamic Coherence** requires that this "revealed preference" over events is acyclic.

Axiom 3 (**Dynamic Coherence**). For any $A_1, \ldots, A_n \subseteq S$, if $S \setminus A_i$ is $\succsim_{A_{i+1}}$-null for each $i \leq n-1$ and $S \setminus A_n$ is $\succsim_{A_1}$-null, then $\succsim_{A_1} = \succsim_{A_n}$.

If A_i^c is A_{i+1} null, then A_i is revealed implied by A_{i+1}. Since **Dynamic Coherence** implies this relation is acyclic, the revealed preference satisfies SARP. Using the result of Matzkin (1991), an extension of Afriat (1967) to general budget sets, SARP is a necessary and sufficient condition for having a subjective distance function for belief selection.

Theorem 1. *The following are equivalent.*

(i) A family of preference relations $\{\succsim_E\}_{E \in \Sigma}$ *admits an* **IEU** *representation.*

(ii) It satisfies **SEU Postulates**, **Consequentialism**, *and* **Dynamic Coherence**.

(iii) It admits an **IEU** *representation with respect to a continuous, strictly convex distance function.*

For a simple intuition behind our result, note that under **SEU Postulates** and **Consequentialism**, our DM has a conditional belief μ_E with support contained in E. Consequently, we may view each event E as generating a "budget set," $\Delta(E)$, from which the DM may choose her conditional belief. The conditional belief, μ_E, is therefore "revealed preferred" to any other belief in the budget set. **Dynamic Coherence** ensures that this revealed preference satisfies SARP, allowing for the construction of a distance measure that generates these beliefs.

3.1 Bayesian Updating

Our main theorem does not require **Dynamic Consistency**, and in fact our axioms are independent of this classic postulate. Similar to results from Ghirardato (2002) and Epstein and Breton (1993), imposing **Dynamic Consistency** in our setting ensures that conditional beliefs are consistent with Bayesian updating whenever possible. Recall that $f_E h$ denotes that conditional act that returns $f(s)$ for $s \in E$ and $h(s)$ otherwise.

Axiom 4 (**Dynamic Consistency**). For all non-null events $E \in \Sigma$ and $f, g, h \in \mathcal{F}$,

$$f_E h \succsim g_E h \text{ if and only if } f \succsim_E g.$$

Proposition 2. *A family of preference relations $\{\succsim_E\}_{E\in\Sigma}$ satisfies* **SEU Postulates**, **Consequentialism**, **Dynamic Coherence**, *and* **Dynamic Consistency** *if and only if it admits an* **IEU** *representation and $\mu_E = BU(\mu, E)$ for each non-null E.*

Since **Dynamic Consistency** has been discussed extensively, (both Ghirardato (2002) and Epstein and Breton (1993) include excellent discussions), we will not discuss this result further. Instead, we simply wish to remark that **Dynamic Consistency** makes no restrictions on beliefs evolve after ex-ante null events, which is a major drawback of the standard model. One of the primary motivations of **IEU** is to provide a coherent common framework for handling belief revision after null events, which we discuss in section 4.

3.2 Non-Bayesian Updating: Characterizing h-Bayesian Updating

One of the key insights provided by **IEU** is that distance minimization can be viewed as a unifying framework that accommodates many updating behaviors. In this section, we expand upon this insight by characterizing two special cases of **h-Bayesian** updating.

The **h-Bayesian** distance involves distortion of the prior before conditioning. Both of the special cases that we characterize involve disciplining the ways in which these likelihoods may be distorted. We begin by introducing our first postulate, which is a mild monotonicity condition ensuring that the DM preserves the "more likely than" judgments implied by her prior.

Axiom 5 (**Monotonicity**). For any $E \in \Sigma$, $s, s' \in E$, and $x, y \in X$,

$$x\{s\}y \succsim x\{s'\}y \text{ if and only if } x\{s\}y \succsim_E x\{s'\}y.$$

To understand **Monotonicity**, consider a DM placing bets on the outcome of a draw from an urn containing red, blue, and yellow balls. Suppose $S = \{r, b, y\}$, $\mu = (16/20, 3/20, 1/20)$, and $E = \{b, y\}$. Under **Dynamic Consistency**, relative likelihoods are exactly preserved and so a Bayesian DM continues to believe that b is three times as likely as y upon learning E. Without **Dynamic Consistency**, the **IEU** would place no restrictions on the conditional relative likelihoods of b and y. Since our DM believed that E was quite unlikely ex-ante, it is plausible that she is now less confident in her judgment about the relative odds of b and y. Consequently, she may desire to further modify her belief so that y is becomes more likely. For example, she may now think that b is only twice as likely as y, resulting in the posterior $\mu_E = (2/3, 1/3)$. Notice that b is still more likely that y; she does not entirely disregard her previous judgments. This restriction is precisely the content of **Monotonicity**.

Proposition 3. *A family of preference relations $\{\succsim_E\}_{E\in\Sigma}$ satisfies* **SEU Postulates**, **Consequentialism**, **Dynamic Coherence**, *and* **Monotonicity** *if and only if it admits an* **IEU** *rep-*

resentation and there is a strictly increasing $h_E : [0,1] \to [0,1]$ *such that* $\mu_E = BU(h_E(\mu), E)$ *for any non-null* E.

Since h_E is increasing, the comparative likelihood judgments from her prior are maintained by her posterior. Notice however that h_E is event dependent, and so the way in which beliefs are adjusted may change across events. The following axiom strengthens **Monotonicity** to ensure that her distortions are consistent across events.

Axiom 6 (**Independence of Irrelevant Information**). For any $E_1, E_2 \in \Sigma$, $s, s' \in E_1 \cap E_2$, and $x, y, z \in X$,

$$x\{s\}z \succsim_{E_1} y\{s'\}z \text{ implies } x\{s\}z \succsim_{E_2} y\{s'\}z.$$

We can now characterize monotone h-Bayesian updating.

Proposition 4. *A family of preference relations* $\{\succsim_E\}_{E\in\Sigma}$ *satisfies* **SEU Postulates**, **Consequentialism**, **Dynamic Coherence**, *and* **Independence of Irrelevant Information** *if and only if it admits an* **IEU** *representation and there is a strictly increasing* $h : [0,1] \to [0,1]$ *such that* $\mu_E = BU(h(\mu), E)$ *for any non-null* E.

4 Updating After Zero-probability Events

One of the well-known weaknesses of Bayesian updating is that it is not defined for zero-probability events. This is particularly problematic in game theoretic settings, where beliefs are induced by the equilibrium strategies and any action off the equilibrium path is a zero-probability event. In contrast, our notion of belief updating is well-defined for zero-probability events. Thus, **IEU** provides a way to extend (non-)Bayesian updating to all events.

4.1 Conditional Probability System

Perhaps the most well-known method for handling choice conditional on (ex-ante) null-events is the conditional probability system (CPS) introduced by Myerson (1986a,b). The development of CPS is closely related the developments of Perfect Bayesian Equilibrium and its refinements. PBE requires that agents's beliefs are Bayes-consistent with the prior whenever possible. However, PBE does not make any restrictions when Bayes' rule is not applicable. Hence, PBE may allow for some unreasonable beliefs for actions off the equilibrium path. The Sequential Equilibria of Kreps and Wilson (1982) refines the PBE by requiring that any belief in sequential equilibria should be a limit of full-support beliefs after applying Bayes rule accordingly. Checking whether conditional beliefs can be supported by full-support beliefs is not easy task and Myerson (1986a,b) shows that this limit requirement of sequential equilibria is equivalent to the following simple condition.

Definition 7. A **conditional probability system** is a collection of conditional probability functions $p(\cdot|E)$, one for each event $E \in \Sigma$, such that for all $G \subseteq F \subseteq E$, $F \neq \varnothing$.

$$p(G|E) = p(G|F)p(F|E), \tag{4}$$

When $p(F|E) \neq 0$, Equation 4 reduces to Bayes' rule. However, when $p(F|E) = 0$, it implies that $p(G|E) = 0$ as well, and so it places no restriction directly on $p(G|F)$.

A major distinction between CPS and **IEU** is that CPS requires Bayesian updating whenever possible, whereas a major goal of the current paper is to provide a unifying framework that allows for Bayesian and non-Bayesian updating. Further, the CPS framework places no restrictions on how beliefs change conditional on null events, whereas **IEU** disciplines belief revision even for null events.

Axiom 7 (**Conditional Consistency**). For all $E \in \Sigma$, $\succsim_E$-feasible $A \subset E$, and $f, g, h \in \mathcal{F}$

$$f_A h \succsim_E g_A h \text{ if and only if } f \succsim_A g.$$

Conditional Consistency implies **Dynamic Consistency** but also has bite on events that are (ex-ante) $\succsim_S$-null. In essence, **Conditional Consistency** extends the logic of **Dynamic Consistency** to all conditional preferences E and nested events that are $\succsim_E$-feasible.

Theorem 2. *A family of preference relations* $\{\succsim_E\}_{E\in\Sigma}$ *satisfies* **SEU Postulates**, **Consequentialism**, *and* **Conditional Consistency** *if and only if it admits a* **CPS** *representation.*

Proposition 5. *Suppose a family of preferences* $\{\succsim_E\}_{E\in\Sigma}$ *admits a CPS representation. Then there are* $\mu^0, \ldots, \mu^K \in \Delta(S)$ *such that* $sp(\mu^0), \ldots, sp(\mu^K)$ *is a partition of* S *and for any* $E \in \Sigma$,

$$\mu_E = BU(\mu^{k^*}, E) \text{ where } k^* = \min\{k \mid sp(\mu_k) \cap E \neq \varnothing\}.$$

Moreover, there are $M_0, \ldots, M_K > 0$ *such that* $\{\succsim_E\}_{E\in\Sigma}$ *has an IEU representation with respect to the following distance function:*

$$d_\mu(\pi) = -\sum_{s_i \in sp(\pi)} \mu_i^{k^*} \sigma\big(\frac{\pi_i}{\mu_i^{k^*}}\big) - M_{k^*} \left|\{sp(\pi) \cap sp(\mu_{k^*})\}\right|,$$

where $k^* = \min\{k \mid sp(\mu_k) \cap sp(\pi) \neq \varnothing\}$.

Proposition 5 shows that the CPS representation is generated by a support-dependent bayesian distance. Further, posterior selection under the CPS representation is "ordered" and the DM selects the first μ^k that is consistent with the event E.

4.2 Hypothesis Testing

A more recent addition to the literature on updating after zero-probability events is the Hypothesis Testing model (HT) of Ortoleva (2012). 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 $\rho \in \Delta(\Delta(S))$, and a threshold $\epsilon \in [0, 1)$ with the requirement that $\mu = \arg\max_{\pi \in \Delta(S)} \rho(\pi)$. Then, for any $E \in \Sigma$,

$$\mu_E = \begin{cases} \text{BU}(\mu, E) & \text{if } \mu(E) > \epsilon, \\ \text{BU}(\pi^\rho_E, E) & \text{otherwise.} \end{cases}$$

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

It turns out that HT is behaviorally equivalent to **IEU**.

Corollary 1. A family of preference relations $\{\succsim_E\}_{E \in \Sigma}$ admits an HT representation if and only if is admits an **IEU**.

This corollary immediately follows from our Theorem 1 and Theorem 1 of Ortoleva (2012). However, it is important to note that our proofs are quite different. Further, we also explicitly construct a distance function that generates any HT representation.

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 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.

By Proposition 2 of Ortoleva (2012), we can assume that $\rho(\pi) \neq \rho(\pi')$ for any two distinct $\pi, \pi' \in \Delta$ without loss of generality.

Proposition 6. *For any HT representation, (μ, ρ, ϵ), let*

$$d_\mu(\pi) = \begin{cases} -\sum_{s \in sp(\pi)} \mu(s)\sigma\left(\frac{\pi(s)}{\mu(s)}\right) - M\left|\{sp(\pi) \cap sp(\mu)\}\right| & \textit{if } \mu(sp(\pi)) > \epsilon, \\ -\sum_{s \in sp(\pi)} \pi^\rho_{sp(\pi)}(s)\,\sigma\left(\frac{\pi(s)}{\pi^\rho_{sp(\pi)}(s)}\right) + M(|S| + 1 - |sp(\pi)|) & \textit{otherwise.} \end{cases}$$

Suppose $M > \sigma(1) - \sigma(0)$. *Then for any* $E \in \Sigma$,

$$\mu_E = \begin{cases} BU(\mu, E) & \text{if } \mu(E) > \epsilon, \\ BU(\pi_E^\rho, E) & \text{otherwise.} \end{cases}$$

The details of this construction can be found in Appendix A.2.

4.3 Relating HT and CPS

In light of Corollary 1 and Proposition 5, it follows that CPS is a special case of HT.

Theorem 3. *For each family of* $\{\succsim_E\}_{E\in\Sigma}$ *CPS preferences, there exists an equivalent family of HT preferences with* $\varepsilon = 0$.

The construction of an HT representation for a given CPS is instructive however, as it illustrates a key behavioral distinction between the two models. Indeed, it sheds light on why this implication cannot be reversed. Of course, this is not surprising for HT with $\epsilon > 0$. However, even when $\epsilon = 0$, HT preferences may be inconsistent with CPS preferences. The reason for this is due to the way in which the selection of new beliefs occurs in HT. The following example illustrates this distinction.

Example 1. Consider rolling a die, which may land on a face, a corner, or an edge. Suppose that we enumerate the faces (from 1 to 6, as usual) as well as the eight corners (from 1 to 8) and the twelve edges (from 1 to 12). Hence, the state space is

$$S = \{ \underbrace{s_1^1, \dots, s_6^6}_{:=F}, \underbrace{s_7^1, \dots, s_{14}^8}_{:=C}, \underbrace{s_{15}^1 \ \dots s_{26}^{12}}_{:=E} \} \tag{5}$$

Then let $A_1 = C \cup E$, $A_2 = E$, and suppose π_S is uniform over F, π_{A_1} is uniform over C and π_{A_2} is uniform over E. Let $\rho(\pi_S) = 1-2\gamma$, $\rho(\pi_{A_1}) = \gamma+\alpha$, $\rho(\pi_{A_2}) = \gamma-\alpha$, for $0 < \alpha < \gamma < \frac{1}{14}$ and let $(u, \rho, 0)$ be the corresponding Hypothesis Testing Representation.

Consider the event $B = \{s_{14}^8, s_{15}^1, s_{16}^2, s_{17}^3\}$ and note that $B \cap C \neq \emptyset \neq B \cap E$. Intuitively, B "crosses" two distinct levels of ex-ante null events, and becomes non-null after either A_1 or A_2: $\pi_{A_1}(B) > 0$ and $\pi_{A_2}(B) > 0$. For sufficiently small α each of the following hold.

1. $\succsim_S$ is represented by (u, π).

2. $\succsim_{A_1}$ is represented by (u, π_{A_1}). This follows because $\pi_{A_1}(A_1) = \pi_{A_2}(A_1) = 1$ and for $\alpha > 0$,

$$\pi_{A_1}(A_1)\rho(\pi_{A_1}) = \gamma + \alpha > \gamma - \alpha = \pi_{A_2}(A_1)\rho(\pi_{A_2}).$$

3. $\succsim_B$ is represented by (u, π_B), where $\pi_B = BU(\pi_{A_2}, B)$ because

$$\pi_{A_1}(B)\rho(\pi_{A_1}) = \left(\frac{1}{8}\right)(\gamma + \alpha) < \left(\frac{3}{12}\right)(\gamma - \alpha) = \pi_{A_2}(B)\rho(\pi_{A_2}).$$

for small enough α.

The family of conditional beliefs generated by this HT does not constitute a CPS. Consider events A_1, $B = \{s^8_{14}, s^1_{15}, s^2_{16}, s^3_{17}\}$ and $B' = \{s^8_{14}\}$. Hence, $B' \subseteq B \subseteq A_1$. However, we have

$$\underbrace{\pi_{A_1}(B')}_{:=p(B'|A_1)} = 1 \neq 0 \cdot \frac{1}{8} = \underbrace{\pi_B(B')}_{:=p(B'|B)} \cdot \underbrace{\pi_{A_1}(B)}_{:=p(B|A_1)},$$

violating the condition in Definition 7.

The relationships between the three models is illustrated in Figure 1.

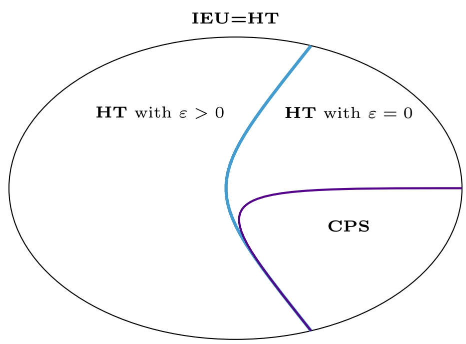

Figure 1: Illustration of Model Relationships

4.4 Other approaches

Finally, another approach to dealing with null events is the Lexicographic Probability System (LPS) of Blume et al. (1991). While LPS also involves a collection of probability distributions, LPS utilizes the entire collection of distributions in the evaluation process via a lexicographic ordering. Consequently, a DM described by LPS will violate Archimedean Continuity, (see Axiom 1(ii)) of the ex-ante preference. Further, LPS replaces (Savage) null-events with "infinitely more likely than," so that null-events are effectively precluded.

5 Incorporating a Signal Structure

We close by illustrating that our framework can incorporate standard signal structures utilized in experimental settings and games by considering a richer state space (e.g., S has a product structure).

Let K be the payoff relevant state space and M be the set of all signals. For each $k \in K$ and $m \in M$, let $P(k)$ be the (unconditional) probability that the payoff relevant state k occurs and $P(m|k)$ be the (conditional) probability that the DM receives the signal m when the state is k. Indeed, learning a signal will be equivalent to learning an event in an expanded state space, $S = K \times M$. Specifically, receiving signal m is equivalent to learning the event $\{(k,m)\}_{k \in K}$ in S.

Let μ be the prior on S, so that $\mu_{km} = P(m|k)\,P(k)$ for each (k,m). In the case of Bayesian updating, the connection between our framework and the signal structure is straightforward. Note that the KL divergence generates Bayesian updating in the signal structure framework:

$$P(k|m) = \frac{\mu_{km}}{\sum_{k' \in K} \mu_{k'm}} = \frac{P(m|k)\,P(k)}{\sum_{k' \in K} P(m|k')\,P(k')}.$$

A similar connection is possible for non-Bayesian updating rules. For example, consider now the following distance function:

$$d_\mu(\pi) = \sum_{(k,m) \in K \times M} \left(\sum_{m \in M} \mu_{k'm} \right)^{d-c} \mu_{km}^c \log\left(\frac{\pi_{km}}{\mu_{km}}\right).$$

This distance generates

$$P(k|m) = \frac{\left(\sum_{m \in M} \mu_{k'm}\right)^{d-c} \mu_{km}^c}{\sum_{k' \in K} \left(\sum_{m \in M} \mu_{km}\right)^{d-c} \mu_{km}^c} = \frac{(P(m|k))^c\,(P(k))^d}{\sum_{k' \in K} (P(m|k'))^c\,(P(k))^d},$$

the non-Bayesian updating rule proposed by Grether (1980).

6 Related Literature

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. 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 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.

A Proofs of Main Results

A.1 Proof of Theorem 1

Proof of Theorem 1. Note that (iii) trivially implies (i). Let us first show that (i) implies (ii). Suppose $\{\succsim_E\}$ admits an IEU representation with respect to (μ, u, d_μ). The IEU representation indeed satisfies SEU postulates. We now prove the necessity of Consequentialism and Dynamic Coherence.

Consequentialism. Take any $E \in \Sigma$ and $f, g \in F$ such that $f(s) = g(s)$ for all $s \in E$. Since $\mu_E(E) = 1$ and $f(s) = g(s)$ for all $s \in E$, we have

$$\sum_{s\in S} \mu_E(s) f(s) = \sum_{s\in E} \mu_E(s) f(s) = \sum_{s\in S} \mu_E(s) g(s) = \sum_{s\in E} \mu_E(s) g(s);$$

i.e., $f \sim_E g$.

Dynamic Coherence. Take any $A_1, \ldots, A_n \subseteq S$ such that $S \setminus A_i$ is $\succsim_{A_{i+1}}$-null for each $i \leq n-1$ and $S \setminus A_n$ is $\succsim_{A_1}$-null. Equivalently, $\mu_{A_{i+1}}(A_i) = 1$ for each $i \leq n-1$ and $\mu_{A_1}(A_n) = 1$. Since $\mu_{A_{i+1}} \in \Delta(A_i)$ and $\mu_{A_i} = \arg\min_{\pi \in \Delta(A)} d_\mu(\pi)$, $d_\mu(\mu_{A_i}) \geq d_\mu(\mu_{A_{i+1}})$. Similarly, we have $d_\mu(\mu_{A_n}) \geq d_\mu(\mu_{A_1})$. Therefore, we have

$$d_\mu(\mu_{A_1}) \geq d_\mu(\mu_{A_2}) \geq \ldots \geq d_\mu(\mu_{A_n}) \geq d_\mu(\mu_{A_1});$$

i.e., $d_\mu(\mu_{A_1}) = d_\mu(\mu_{A_n})$. Hence, $\mu_{A_1} = \mu_{A_n}$; i.e., $\succsim_{A_1} = \succsim_{A_n}$.

Let us now show that (ii) implies (iii). Suppose $\{\succsim_E\}_{E\in\Sigma}$ satisfies **SEU Postulates**, **Consequentialism**, and **Dynamic Coherence**. Since $\succsim$ satisfies SEU postulates, there is (μ, u) such that $\succsim$ has a SEU representation with (μ, u). Since $\succsim_E$ satisfies SEU postulates, there is (μ_E, u_E) such that $\succsim_E$ has a SEU representation with (μ_E, u_E). By Invariant Risk Preference, $u_E(p) \geq u_E(q)$ and $u(p) \geq u(q)$ for any $p, q \in \Delta(X)$. Without loss of generality, let us assume that $u_E = u$. Hence, $\succsim_E$ has a SEU representation with (μ_E, u).

Let us know discuss implications of **Consequentialism**. For any $E \in \Sigma$ and all $f, g \in F$ and $p, q \in \Delta(X)$ such that $p \succ q$ and $f(s) = g(s) = p$ for all $s \in E$ and $f(s) = p$ and $g(s) = q$

for any $s \in E^c$. By **Consequentialism**, we have $f \sim_E g$; equivalently,

$$\sum_{s\in S} \mu_E(s) f(s) = u(p) = \sum_{s\in E} \mu_E(s) g(s) = \mu_E(E) u(p) + (1 - \mu_E(E)) u(q).$$

In other words, we have $\mu_E(E) = 1$; i.e., $\mu_E \in \Delta(E)$.

Afriat's theorem for general budget sets. To obtain the Inertial EU representation, we use an extension of Afriat's theorem (Afriat (1967)) for general budget sets due to Matzkin (1991). To state Afriat's theorem for general budget sets, some notations are necessary. Let Z be a convex, bounded subset of $\mathbb{R}^n_+$. Let $\mathscr{D} = (\mathbf{x}^t, B^t)_{t\in T}$ be a data set where $\mathbf{x}^t \in B^t$ is the observed consumption bundle that is chosen from the budget set $B^t \subset Z$ at observation $t \in T$. We assume that for each $t \in T$, $(\mathbf{x}^t, B^t)$ is co-convex subset of Z; i.e., (i) $Z \setminus B^t$ is open and convex; (ii) for any $\mathbf{e} \geq 0$ and $\mathbf{x} \in Z \setminus B^t$, $\mathbf{x} + \mathbf{e} \in Z$ implies $\mathbf{x} + \mathbf{e} \in Z \setminus B^t$; (iii) for any $\mathbf{e} > 0$, $\mathbf{x}^t + \mathbf{e} \in Z$ implies $\mathbf{x}^t + \mathbf{e} \in Z \setminus B^t$.

Let us now define the following revealed preference relation on $\{x^t\}_{t\in T}$. We say $\mathbf{x}^t$ is revealed preferred to $\mathbf{x}^s$, denoted by $\mathbf{x}^t \succsim_R \mathbf{x}^s$ if $\mathbf{x}^s \in B^t$. We say $\mathbf{x}^t$ is strictly revealed preferred to $\mathbf{x}^s$, denoted by $\mathbf{x}^t \succ_R \mathbf{x}^s$ if $\mathbf{x}^s \in B^t$ and $\mathbf{x}^t \neq \mathbf{x}^s$. Finally, we say the data set $\mathscr{D} = (\mathbf{x}^t, B^t)_{t\in T}$ satisfies the Strong Axiom of Revealed Preferences (SARP) if $\succsim_R$ is acyclic; i.e., there is no sequence $\mathbf{x}^{t_1}, \mathbf{x}^{t_2}, \ldots, \mathbf{x}^{t_L}$ such that $\mathbf{x}^{t_i} \succsim_R \mathbf{x}^{t_{i+1}}$ for each $i \leq n-1$ and $\mathbf{x}^{t_L} \succ_R \mathbf{x}^{t_1}$.

Theorem 1 of Matzkin (1991). The data set $\mathscr{D} = (\mathbf{x}^t, g^t)_{t\in T}$ satisfies SARP if and only if there is a strictly increasing, continuous, strictly concave utility function $u : Z \to \mathbb{R}$ such that for any $t \in T$,

$$u(\mathbf{x}^t) > u(\mathbf{x}) \text{ for any } \mathbf{x} \in B^t \setminus \{\mathbf{x}^t\}.$$

To apply the above theorem, let us arbitrarily label the set of all events: $\Sigma = \{E_t\}_{t\in T}$. Then let $Z = \Delta(S)$ and $\mathbf{x}^t = \mu_{E_t}$ and $B^t = \Delta(E_t)$ for each $t \in T$. Let $\mathscr{D} = (\mathbf{x}^t, B^t)_{t\in T}$.

Note that Z is a convex, bounded subset of $\mathbb{R}^n_+$. Let us show that $(\mathbf{x}^t, B^t)$ is co-convex subset of Z. First, $Z \setminus B^t$ is open and convex in Z. Second, for any $\mathbf{x} \in Z$ and $\mathbf{e} \geq 0$, $\mathbf{x} + \mathbf{e} \in Z$ implies $\mathbf{e} = 0$. Hence, (ii) and (iii) of co-convexity are trivially satisfied.

Let us now show that **Dynamic Coherence** implies that $\mathscr{D} = (\mathbf{x}^t, B^t)_{t\in T}$ satisfies SARP. Take any sequence $\mathbf{x}^{t_1}, \mathbf{x}^{t_2}, \ldots, \mathbf{x}^{t_L}$ such that $\mathbf{x}^{t_i} \succsim_R \mathbf{x}^{t_{i+1}}$ for each $i \leq L-1$ and $x^{t_L} \succsim_R \mathbf{x}^{t_1}$. To prove SARP, we shall show that $\mathbf{x}^{t_L} = \mathbf{x}^{t_1}$. By definition of the revealed preference relation $\succsim_R$, $\mathbf{x}^{t_i} \succsim_R \mathbf{x}^{t_{i+1}}$ is equivalent to $\mathbf{x}^{t_{i+1}} \in \Delta(E^{t_i})$. In other words, $\mu_{E_{t_{i+1}}} \in \Delta(E_{t_i})$ for each $i \leq L-1$. Similarly, $\mu_{E_{t_1}} \in \Delta(E_{t_L})$.

Note that $\mu_{E_{t_{i+1}}} \in \Delta(E_{t_i})$ implies $\mu_{E_{t_{i+1}}}(E_{t_i}) = 1$; equivalently, $\mu_{E_{t_{i+1}}}(S \setminus E_{t_i}) = 0$. In other words, $S \setminus E_{t_i}$ is $\succsim_{E_{t_{i+1}}}$-null for each $i \leq L-1$. Similarly, $S \setminus E_{t_L}$ is $\succsim_{E_{t_1}}$-null. By

Dynamic Coherence, $\succsim_{E_{t_1}}=\succsim_{E_{t_L}}$; equivalently, $\mu_{E_{t_1}} = \mu_{E_{t_L}}$. In other words, $\mathbf{x}^{t_1} = \mathbf{x}^{t_L}$

Since $\mathscr{D} = (\mathbf{x}^t, B^t)_{t\in T}$ satisfies SARP, by Theorem 1 of Matzkin (1991), there is a strictly increasing, continuous, strictly concave utility function $u : Z \to \mathbb{R}$ such that for any $t \in T$,

$$u(\mathbf{x}^t) > u(\mathbf{x}) \text{ for any } \mathbf{x} \in B^t \setminus \{\mathbf{x}^t\}.$$

Let $d_\mu = -u$. Then since $B^t = \Delta(E_t)$ and $\mathbf{x}^t = \mu_{E_t}$,

$$\mu_{E_t} = \arg\min_{\pi\in\Delta(E_t)} d_\mu(\pi).$$

Finally, note that d_μ is continuous and strictly convex.

□

A.2 Proof of Proposition 6

Consider a Hypothesis Testing representation (μ, ρ, ϵ).

Proof. We first establish some notation. Let $\pi^\rho_A = \arg\max_{\pi\in\Delta(S)} \rho(\pi)\pi(A)$ for any $\rho \in \Delta(\Delta(S))$. We denote the support of π by $sp(\pi)$. For any $B \subseteq S$, let

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

and

$$g(B) = -\sum_{s\in B} \pi^\rho_B(s)\,\sigma\Big(\frac{\pi^\rho_B(s)}{\pi^\rho_B(B)}\Big).$$

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

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

Recall the distance function:

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

Take any $A \subseteq S$ and $s \in A$.

Case 1. We shall show that $\mu_A = BU(\pi^\rho_A, A)$ when $\mu(A) \le \epsilon$.

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

Hence, we have

$$d_\mu(\pi) = -\sum_{s\in sp(\pi)} \pi^\rho_{sp(\pi)}(s)\,\sigma(\frac{\pi(s)}{\pi^\rho_{sp(\pi)}(s)}) + M(|S|+1-|sp(\pi)|) \text{ for any } \pi \in \Delta(A).$$

Take any $B \subseteq A$. For any $\pi \in \Delta(A)$ 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 $M(|A| - |B|) \geq g(A) - g(B)$ and by the definition of M, we have $M(|A| - |B|) \geq M > |g(A) - g(B)| \geq g(A) - g(B)$. Therefore, μ^A minimizes $d_\mu(\pi)$ subject to $\pi \in \Delta(A)$. Hence, $\mu_{\Delta(A)}(s) = \mu^A(s) = \frac{\pi^\rho_A(s)}{\pi^\rho_A(A)}$.

Case 2. We shall prove that $\mu_A = BU(\mu, A)$ when $\mu(A) > \epsilon$. Take any $\pi \in \Delta(A)$.

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 $\pi \in \Delta(A)$ 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 $\pi \in \Delta(A)$ with $\mu(sp(\pi)) \leq \epsilon$. We now show that there is a $\pi \in \Delta(A)$ 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 $\pi \in \Delta(A)$ 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 $\mu \in \Delta(A)$. 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,

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

□

B Consequentialism and Weighted IEU

In this section, we relax generalize our main result by relaxing **Consequentialism**. Following out analogy to revealed preference theory, **Consequentialism** ensures that E is equivalent to the budget set $\Delta(E)$. By dropping **Consequentialism**, we allow for the DM to perceive a subjective budget set from which she may choose. We do however impose two natural conditions on her behavior.

Definition 8 (WIEU). A family of preference relations $\{\succsim_E\}_{E\in\Sigma}$ admits an **Weighted Inertial 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 a weight $\delta \in [0, 1]$ such that for each $E \in \Sigma$, the preference relation $\succsim_E$ admits a SEU representation with (u, μ_E), where

$$\mu_E \equiv \delta\,\mu + (1-\delta) \operatorname*{arg\,min}_{\pi\in\Delta(E)} d_\mu(\pi). \tag{6}$$

Definition 9 (Conditional Null Event)**.** We say E is a conditional $\succsim_A$-null event if there is $\alpha \in (0, 1]$ such that for any $f \in \mathcal{F}$ and $p, q \in \Delta(X)$,

$$f\,E\,w \sim p \text{ implies } f\,E\,w \sim_A \alpha\,p + (1-\alpha)w.$$

$$w\,E\,q \sim p \text{ implies } w\,E\,q \sim_A \alpha\,p + (1-\alpha)q.$$

Axiom 8 (**Conditional Dynamic Coherence**). For any $A_1, \ldots, A_n \subseteq S$, if $S \setminus A_i$ is conditional $\succsim_{A_{i+1}}$-null for each $i \le n-1$ and $S \setminus A_n$ is conditional $\succsim_{A_1}$-null, then $\succsim_{A_1} = \succsim_{A_n}$.

Axiom 9 (**Relative Tradeoff Consistency**). For any $E_1, E_2, A_1, A_2 \in \Sigma$ such that E_i is conditional $\succsim_{A_i}$-null, $p, q, r \in \Delta$, and $\alpha \in [0, 1]$,

$$\text{If } w\,E_1\,q \sim p \text{ and } w\,E_1\,q \sim_{A_1} \alpha\,p + (1-\alpha)q, \text{ then}$$

$$w\, E_2\, q \sim r \text{ implies } w\, E_2\, q \sim_{A_2} \alpha\, r + (1-\alpha) q.$$

Second, we require that her subjective belief in E weakly increases after she is told that E has occurred. While **Consequentialism** demands that the DM is convinced of E, our novel axiom as **Partial Trust** only demands that she puts more stock in E.

Axiom 10 (**Partial Trust**). For any $E \subseteq S$ and $p, q, r \in \Delta(X)$ with $p \succ q$,

$$p_E q \succsim r \text{ implies } p_E q \succsim_E r.$$

These two conditions ensure a structurally similar representation to Theorem 1. The main distinction is that the the budget set is a function of the DM's "trust" in the information. In particular, for each event E our DM assign it a trust value $t(E)$. The subjective budget set can be viewed as a convex combination of the prior and $\Delta(E)$, where $t(E)$ is the weight placed on $\Delta(E)$. Note that whenever $t(E) = 1$, our DM satisfies consequentialism at E.

Conjecture 1. *Suppose $\succsim$ has a full-support. The following are equivalent.*

(i) A family of preference relations $\{\succsim_E\}_{E\in\Sigma}$ admits a **Weighted Inertial EU** *representation.*

(ii) It satisfies **SEU Postulates**, **Conditional Dynamic Coherence**, **Partial Trust**, *and* ***Relative Tradeoff Consistency***.

(iii) It admits an **Weighted IEU** *representation with respect to a continuous, strictly convex distance function.*

References

Afriat, S. N. (1967): "The construction of utility functions from expenditure data," *International economic review*, 8, 67–77. 7, 16

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. 14

Anscombe, F. and R. Aumann (1963): "A Definition of Subjective Probability," *Annals of Mathematical Statistics*, 34, 199–205. 3, 6

Basu, P. (2019): "Bayesian updating rules and AGM belief revision," *Journal of Economic Theory*, 179, 455 – 475. 14

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. 2, 14

Blume, L., A. Brandenburger, and E. Dekel (1991): "Lexicographic Probabilities and Choice under Uncertainty," *Econometrica*, 59, 61–79. 13

Epstein, L. G. (2006): "An Axiomatic Model of Non-Bayesian Updating," *Review of Economic Studies*. 14

Epstein, L. G. and M. L. Breton (1993): "Dynamically Consistent Beliefs Must Be Bayesian," *Journal of economic theory*, 61, 1–22. 7, 8

Epstein, L. G., J. Noor, and A. Sandroni (2008): "Non-Bayesian Updating: A Theoretical Framework," *Theoretical Economics*. 15

Galperti, S. (2019): "Persuasion: The art of changing worldviews," *American Economic Review*, 109, 996–1031. 6

Ghirardato, P. (2002): "Revisiting Savage in a conditional world," *Economic Theory*, 20, 83–92. 7, 8

Grether, D. M. (1980): "Bayes Rule as a Descriptive Model: The Representativeness Heuristic," *Quarterly Journal of economics*. 1, 3, 5, 14

Gul, F. and W. Pesendorfer (2001): "Temptation and Self-Control," *Econometrica*. 15

Kahneman, D. and A. Tversky (1979): "Prospect theory: An analysis of decision under risk," *Econometrica: Journal of the Econometric Society*, 263–291. 5

Kovach, M. (2020): "Conservative Updating," *mimeo*. 15

Kreps, D. M. and R. Wilson (1982): "Reputation and imperfect information," *Journal of Economic Theory*, 27, 253–279. 3, 9

Matzkin, R. L. (1991): "Axioms of revealed preference for nonlinear choice sets," *Econometrica: Journal of the Econometric Society*, 1779–1786. 7, 16

Myerson, R. B. (1986a): "Axiomatic Foundations of Bayesian Decision Theory," Discussion Papers 671, Northwestern University, Center for Mathematical Studies in Economics and Management Science. 1, 2, 3, 6, 9

——— (1986b): "Multistage Games with Communication," *Econometrica*, 54, 323–358. 1, 2, 3, 6, 9

Ortoleva, P. (2012): "Modeling the Change of Paradigm: Non-Bayesian Reactions to Unexpected News," *American Economic Review*, 102, 2410–36. 1, 2, 6, 7, 11

Perea, A. (2009): "A Model of Minimul Probabilistic Belief Revision," *Theory and Decision*, 163–222. 14

Rabin, M. and J. L. Schrag (1999): "First Impressions Matter: A Model of Confirmatory Bias," *Quarterly Journal of Economics*, 114, 37–82. 5