Belief Updating without Complete Trust*

Jawwad Noor and Yuzhao Yang

September 4, 2026

Abstract

We represent a non-Bayesian agent as one who does not completely trust the information they receive. The behavioral expression of complete trust lies in a homogeneity property of Bayesian updating: posterior beliefs do not change if a signal is made arbitrarily rare by scaling down its likelihood vector. We show that simply dropping this property and retaining all other Bayesian behavioral properties yields a unique representation where the agent is still Bayesian but has subjective uncertainty over the information structure generating the signal. The representation result is proved using the Fundamental Theorem of Projective Geometry. We analyze how various updating biases may be rationalized by a lack of trust.

Keywords: Belief updating, information, belief biases, conservatism, confirmatory bias, projective geometry

1 Introduction

The information presented to an agent and the agent's reading of the information may not always coincide. While there may well be psychological reasons for this – perhaps the agent did not put in all the effort required to comprehend the information, or maybe the framing of the information impacted their perception of it – even fully rational agents will exhibit a gap if they wonder whether the provider of information possesses hidden motivations and incentives that cause them to lie. In a world where media slant, misinformation and fake news are often used to manipulate beliefs in both economic and political environments,[1]

*Noor is at the Department of Economics, Boston University, 270 Bay State Road, Boston MA 02215. Email: jnoor@bu.edu. Yang is at the Department of Economics, University of Michigan, 4330 North Quad, 105 S State Street, Ann Arbor, MI 48109. The usual disclaimer applies.

[1]See Gentzkow and Shapiro (2006) and Allcott and Gentzkow (2017) for economic models of media bias and fake news. See Liu and Moss (2024) for the impact of corporate misinformation and fake news in financial markets. See Noor and Payro (2026) for a model of how noisiness in the environment may be used as a smoke screen.

and where even academic research faces a replicability crisis,[2] it may even be irrational to simply take information at face value.

In this paper we explore under what conditions an agent's nonstandard behavior – in the sense of non-Bayesian updating with respect to some information structure σ – can be rationalized by Bayesian updating with respect to some subjective information structure π that is misspecified relative to σ. A literature (reviewed below) explores this question by asking whether some rationalizing π exists, a criterion that places few restrictions on π itself. Relatedly, Bohren and Hauser (2025) observe that "the line between [non-Bayesian updating] and model misspecification is ... fuzzy." We deviate from the existing literature by demanding structure on the agent's π: a plausible story for why the agent might subjectively believe that the information is π rather than σ. Making such a demand has methodological value: there "should" be discipline on the analyst's ability to explain away observed violations of incumbent models. Moreover, the hypothesis that the agent is rational despite their non-Bayesian behavior may well have testable implications on additional data, which should be made explicit. The story we tell is one where the agent harbors uncertainty about whether σ is indeed the true information structure, possibly arising from lack of trust, and the paper is devoted to uncovering the behavioral content of such a story.

In a standard setting of choice under uncertainty with some state space Ω, our primitive consists of how the agent ranks acts before and after receiving a signal $s \in S$ from some information structure (henceforth *experiment*) σ that varies in some primitive set of experiments Σ. Our first main result provides an axiomatization of Bayesian updating, where each preference admits a Subjective Expected Utility representation and where posterior beliefs are Bayesian with respect to σ:

$$p_\sigma(\omega|s) = \frac{\sigma(s|\omega)p(\omega)}{\sum_{\omega' \in \Omega} \sigma(s|\omega')p(\omega')}, \qquad \omega \in \Omega.$$

We observe that the behavioral content of "complete trust in σ" lies in a *Homogeneity* condition, which states that posterior beliefs conditional on a signal s depend only on relative likelihoods $\frac{\sigma(s|\omega)}{\sigma(s|\omega')}$. Intuitively, when there is uncertainty about σ, absolute likelihoods may well matter because rare signals from σ may prompt reservations about σ. For instance, if one receives an email from a work acquaintance asking for money, the sheer improbability of such a request might trigger the suspicion that the email is a scam.

We explore how the agent may be represented when Homogeneity is simply dropped while maintaining all the other axioms. Our second main result shows that posterior beliefs given σ are represented by

$$p_\sigma(\omega|s) = \frac{[\tau\sigma(s|\omega) + (1-\tau)\rho(s|\omega)]\, p(\omega)}{\sum_{\omega' \in \Omega} [\tau\sigma(s|\omega') + (1-\tau)\rho(s|\omega')]\, p(\omega')}, \qquad \omega \in \Omega.$$

[2]In psychology, the Open Science Collaboration (2015) was unable to replicate 64% of the 100 studies from three major psychology journals they investigated. In economics, Camerer et al. (2016) were unable to replicate 39% out of 18 experimental studies published in two major economics journals.

That is, beliefs must be Bayesian with respect to the subjective experiment $\tau\sigma + (1-\tau)\rho$, which deviates from σ depending on the degree of trust τ the agent has in it. Thus, a violation of Homogeneity on its own does not necessarily indicate a break with the Bayesian paradigm. A lack of trust is therefore behaviorally distinct from existing non-Bayesian updating rules (reviewed below), which retain Homogeneity and weaken other normative properties of Bayesian updating. Our results are proved using the Fundamental Theorem of Projective Geometry.[3]

We conclude this Introduction below with related literature. Section 2 provides the foundations of the Bayesian model and pinpoints Homogeneity as the behavioral expression of complete trust. Section 3 proves a representation theorem for the axioms of the Bayesian model excluding Homogeneity. Section 4 provides the technical insight behind the proof of the representation theorem. Section 5 studies special cases of the model, relating both to weakenings of Homogeneity and also experimental evidence of updating biases. Section 6 shows that the model endogenously gives rise to a form of confirmation bias, and it provides formal characterizations while also relating to the evidence. All proofs are relegated to appendices.

Related Literature. This paper sits at the intersection between the literature on belief updating with model misspecification and the literature on non-Bayesian updating; our characterization of Homogeneity speaks to the distinction between the two, as we discuss in Sections 2 and 3.

The classic reference for the intersection of the literatures is Ghirardato (2002), who works with dynamic Subjective Expected Utility preferences over acts and shows that, under regularity conditions and a partitional information structure, Bayesian updating is equivalent to a dynamic consistency condition (see Appendix A for a brief review). Adapted to our setting his state space is $\Omega \times S$, so the experiment is subjective and recovered from preference; ours is objectively given, which allows us to characterize when the Bayesian agent's subjective model coincides with (or systematically deviates from) it. See Section 2.2 for more comparison. Taking priors and posteriors as observable (that is, taking a belief-theoretic rather than a decision-theoretic approach), Shmaya and Yariv (2016) show that a belief sequence is consistent with Bayes' rule whenever the prior lies in the relative interior of the convex hull of the posteriors. Bohren and Hauser (2023) additionally take the agent's forecasts of posterior distributions as observable, and characterize when beliefs and forecasts are jointly consistent with Bayesian updating under a possibly misspecified model. Taking the empirical distribution of posteriors as observable, Molavi (2025) shows that beliefs are consistent with Bayesian updating if and only if the mean posterior is absolutely continuous with respect to the prior. Fudenberg and Lanzani (2026) characterize when one-step-ahead forecasts of signal realizations are consistent with Bayesian learning about a stable but unknown i.i.d. signal-generating pro-

[3]To the best of our knowledge, the Fundamental Theorem has not so far found applications in economics.

cess. This literature asks whether some rationalizing π exists. We ask instead what behavior looks like when π reflects the agent's uncertainty about whether the presented experiment is the one generating her signals.

Next, we relate our work with the non-Bayesian updating literature. Epstein, Noor, and Sandroni (2008) provide a general framework by relating non-Bayesian updating to preferences over commitment. Ortoleva (2012) proposes a hypothesis-testing model in which the agent applies Bayes' rule after anticipated events but reconsiders her prior upon sufficiently unlikely observations. The intuition that rare signals prompt reconsideration is shared with our model, though the response differs: his agent revises her prior, whereas ours remains Bayesian with respect to an enlarged model space. Cripps (2018), Chambers, Masatlioglu, and Raymond (2024), Jakobsen (2026), and Whitmeyer (2026) characterize generalizations of Bayes' rule satisfying distinct normative criteria. Each retains the Homogeneity axiom that we drop, so the non-Bayesian agent they describe is behaviorally distinct from a Bayesian one who is uncertain about the experiment she faces – see Section 2.2 for a more detailed discussion. Kovach (2021), Yang (2025), and Yang and Zhao (2026) study various systematic deviations from Bayesian updating, including conservatism, overprecision, and representativeness heuristics. Noor and Payro (2026) formalize the law of small numbers as a belief in mean reversion. These papers take the departure itself as the object of axiomatization, whereas ours derives it from uncertainty about the source of the signal. Zhao (2022), Ke, Wu, and Zhao (2024), Ma and Zhao (2024), and Dominiak, Kovach, and Tserenjigmid (2025) axiomatize updating rules for qualitative statements of relative likelihood and for sets of probability distributions. These take a different form of information as the primitive, rather than a different rule for updating on an experiment.

Finally, we relate our work with the small literature that studies uncertainty about information. Gentzkow and Shapiro (2006) study media bias in a model where agents are uncertain about the news source, which could either be of high or low quality, and they use Bayesian updating exactly as in our paper. They note that some form of confirmation bias arises endogenously in such a model – we flesh out their intuition formally in Section 6. Cheng and Hsiaw (2022) similarly adopt Bayesian updating to evaluate the uncertainty (which they call "distrust") about quality of experts, although their agents subsequently update beliefs about the state in a non-Bayesian fashion. While our paper works with posteriors after a single signal, Hu (2026) works with posteriors after various series of signals of different lengths and axiomatically characterizes Bayesian updating with complete certainty (which he calls "confidence") in the information structure. In an experiment, he shows that 95% of subjects behave as though they are not completely confident.

2 Bayesian Updating and Trust

We begin by characterizing Bayesian updating and identifying the behavioral content of complete trust.

2.1 Primitives

We begin with a characterization of Bayesian updating. Consider a finite state space Ω with $|\Omega| \geq 4$ and a finite signal space S. An experiment is a mapping $\sigma : \Omega \to \Delta(S)$. For each $s \in S$, let $\sigma_s := \big(\sigma(s \mid \omega)\big)_{\omega \in \Omega} \in [0,1]^{\Omega}$ denote the likelihood vector of signal s under experiment σ. We will often write it as a $|\Omega| \times |S|$ matrix with states corresponding to rows and signals corresponding to columns. For experiments σ, σ' and $\alpha \in [0,1]$, the mixture $\alpha\sigma + (1-\alpha)\sigma'$ is the experiment with likelihood vector $\alpha\sigma_s + (1-\alpha)\sigma'_s$ at each s.

Consider a nonempty set of experiments Σ. This is the set of experiments the agent may be presented with. The only assumption we impose on Σ is that, for each $s \in S$, the induced set of likelihood vectors

$$\Sigma_s := \{\sigma_s \mid \sigma \in \Sigma\} \subset \mathbb{R}^{\Omega}$$

is open in $\mathbb{R}^{\Omega}$ and contains 0 in its closure.[4] The openness requirement is weak: Σ_s may be an arbitrarily small neighborhood of a single likelihood vector. Thus our representation is identified only from local variation in σ, rather than through behavior across all experiments.

Let X be a convex subset of a normed vector space,[5] and denote the space of acts by $\mathcal{F} = \{f : \Omega \to X\}$. Our primitive is the family of binary relations over $\mathcal{F}$ given by $\{\succsim_0, \succsim_{\sigma,s}\}_{(\sigma,s)\in\Sigma\times S}$, respectively capturing how the agent ranks acts prior to receiving any information, and after being told that signal s has been generated by experiment σ. We maintain throughout that

Axiom 0 (SEU)**.** *For some non-constant affine $u : X \to \mathbb{R}$ and full-support distribution $p \in \Delta(\Omega)$, $\succsim_0$ admits a representation*

$$U(f) = \int_{\Omega} u \circ f \, dp, \qquad f \in \mathcal{F},$$

and for each $(\sigma, s) \in \Sigma \times S$, $\succsim_{\sigma,s}$ admits a representation

$$U_{\sigma,s}(f) = \int_{\Omega} u \circ f \, dp_{\sigma}(\cdot \mid s), \qquad f \in \mathcal{F},$$

for some distribution $p_{\sigma}(\cdot \mid s) \in \Delta\Omega$.

Therefore, there is a prior $p(\cdot)$ underlying $\succsim_0$ and a posterior $p_{\sigma}(\cdot \mid s)$ conditional on (s, σ) underlying $\succsim_{\sigma,s}$. Define the *Bayesian posterior* conditional on

[4]In many environments it may be natural that Σ is much smaller than the set of all possible experiments. For instance, the meaning of a signal may be fixed by language or convention: an enthusiastic letter of recommendation is diagnostic of the event that the candidate is strong, even though writers may differ in how strongly enthusiasm predicts strength. See (6) below. Another example is "approximately" partitional information: Σ may consist of ε-perturbations of deterministic experiments: those σ for which each $\sigma(s \mid \omega)$ lies within ε of 0 or of 1.

[5]Our axioms do not make use of the vector space structure per se. The vector space structure would be used explicitly if Axiom 0 is replaced with the Anscombe-Aumann axioms. The proof of our results relies on $u(X)$ being a nontrivial interval, which is guaranteed by the vector space structure and the assumption in Axiom 0 that u is non-constant and affine.

experiment $\sigma : \Omega \to \Delta(S)$ and signal $s \in S$ by

$$p^*_\sigma(\omega \mid s) = \frac{\sigma(s \mid \omega)p(\omega)}{\sum_{\omega' \in \Omega} \sigma(s \mid \omega')p(\omega')}, \qquad \omega \in \Omega.$$

2.2 Foundations

In the following axioms, we omit the quantifiers "for all $\sigma, \sigma', \sigma'' \in \Sigma$," "for all $s \in S$," "for all $f, g, h \in \mathcal{F}$," and "for all $x, y \in X$." Note well that since Σ is not assumed to be convex, for any $\sigma, \sigma' \in \Sigma$ it needs to be hypothesized that $\alpha\sigma + (1-\alpha)\sigma' \in \Sigma$. Our main axiom is

Axiom 1 (Informational STP)**.** *If $\sigma'' = \alpha\sigma + (1-\alpha)\sigma'$ for $\alpha \in (0,1)$, then*

$$f \succsim_{\sigma,s} g \text{ and } f \succsim_{\sigma',s} g \implies f \succsim_{\sigma'',s} g.$$

Imagine that the agent receives signal s but does not know yet which experiment they are facing: they know only that with probability α the signal s is generated by σ, and with the remaining probability that it is generated by σ'. Maintaining the "background assumption" that the agent reduces lotteries, this lottery corresponds to the facing experiment $\alpha\sigma + (1-\alpha)\sigma'$. The axiom states that if the agent would prefer f over g regardless of whether the experiment is known to be σ or σ', then they must prefer f over g when facing $\alpha\sigma + (1-\alpha)\sigma'$, that is, before finding out which experiment it really is. This normatively compelling axiom has a strong flavor of dynamic consistency in a risk setting: preferences before and after resolution of $\alpha\sigma + (1-\alpha)\sigma'$ are consistent with each other. It does not, however, correspond to Dynamic Consistency (Definition 3 in Appendix A) in the information setting, since the statement is not of the form that "if the agent would ex ante prefer to choose f over g conditional on signal s, then they respect this ex post after observing s".

The remaining axioms are more obvious properties of Bayesian updating. As usual, for any acts f, h and event $E \subset \Omega$,the act fEh is defined as one that pays $f(\omega)$ for all $\omega \in E$ and $h(\omega)$ for all $\omega \notin E$.

Axiom 2 (Independence of Irrelevant Details)**.** *If $\sigma(s \mid \omega) = \sigma'(s \mid \omega)$ for all $\omega \in E \subset \Omega$, then*

$$fEh \succsim_{\sigma,s} gEh \iff fEh \succsim_{\sigma',s} gEh.$$

When $E = \Omega$ the axiom says that $\succsim_{\sigma,s}$ depends only on the likelihood vector σ_s. When $E \subsetneq \Omega$, the evaluation of fEh versus gEh depends only on the likelihoods assigned to states in E. This implies that, for the agent's posterior belief conditional on a signal s, its restriction to E does not depend on the likelihood of s given states outside E. The axiom is thus a counterpart of Luce's independence of irrelevant alternatives, with the states outside E in the role of the irrelevant alternatives. Conditions of this form are known to underlie the characterization of geometric opinion pooling (Genest, Weerahandi, and Zidek, 1984) and of the power-weighted distortions of beliefs and experiments in Grether (1980) (Chambers, Masatlioglu, and Raymond, 2024; Chan, 2026).

Axiom 3 (Homogeneity)**.** *If $\sigma_s = \alpha\sigma'_s$ for $\alpha > 0$, then $\succsim_{\sigma,s}=\succsim_{\sigma',s}$.*

This is the well-known property that scaling the likelihood vector does not change the posterior, and thus has no bearing on $\succsim_{\sigma,s}$.

Axiom 4 (Consistency with Prior)**.** *For $\alpha > 0$, if $\sigma_s = \sigma'_s + \alpha 1_\Omega$, then*

$$f \succ_0 g \text{ and } f \succsim_{\sigma',s} g \implies f \succ_{\sigma,s} g.$$

Recall that σ is uninformative at signal s if it is a constant vector: the ratios of entries satisfy $\frac{\sigma(s|\omega)}{\sigma(s|\omega')} = 1$ for all ω, ω'. Note also that, for an arbitrary σ, adding a constant to its likelihood vector σ_s makes the signal s less informative, since it brings all the ratios closer to 1:

$$\left| \frac{\sigma(s \mid \omega)}{\sigma(s \mid \omega')} - 1 \right| \geq \left| \frac{\sigma(s \mid \omega) + c}{\sigma(s \mid \omega') + c} - 1 \right|.$$

Therefore, the axiom begins with σ_s, σ'_s where σ'_s is more informative than σ_s. The axiom states that if prior to any signal the agent preferred $f \succ_0 g$ and if they still exhibited $f \succsim_{\sigma',s} g$ with an informative likelihood vector σ'_s, then they would exhibit $f \succ_{\sigma,s} g$ with a less informative likelihood vector σ_s. Intuitively, less informative signals push posteriors towards the prior. This expresses that ex post beliefs maintain a connection with prior beliefs. This is unlike, for instance, Epstein, Noor, and Sandroni (2008) where the agent retroactively changes their prior.[6]

For any act f and state ω, consider the constant act that yields consequence $f(\omega)$ at every state, and denote it by f_ω. Our final axiom is a richness condition:

Axiom 5 (Richness)**.** *For all $s \in S$ and $f \in \mathcal{F}$, if $f_\omega \not\sim_0 f_\psi$ for some $\omega, \psi \in \Omega$, then there exist $\sigma, \sigma' \in \Sigma$ and $x \in X$ such that*

$$f \succ_{\sigma,s} x \text{ and } f \prec_{\sigma',s} x.$$

Richness states that, fixing any signal s and any f with distinct best and worst outcomes, there always exists a consequence x (which, given Axiom 0, must be ranked strictly between the best and worst outcomes of f) and a pair of experiments σ, σ' such that f is better than x at one experiment and worse at another. This rules out, for instance, the case where the agent does not update beliefs no matter what experiment is offered to them.

These axioms are the foundations for Bayesian updating in our domain.

Theorem 1. *$\{\succsim_0, \succsim_{\sigma,s}\}_{(\sigma,s)\in(\Sigma\times S)}$ satisfies Axioms 1 - 5 if and only if for all $(\sigma, s) \in \Sigma \times S$,*

$$p_\sigma(\cdot \mid s) = p^*_\sigma(\cdot \mid s). \tag{1}$$

[6]While appearing related, the axiom is logically independent of Informational STP (Axiom 1). Axiom 1, and indeed Axioms 1–3 together, impose no restriction relating $\succsim_{\sigma,s}$ to the ex ante ranking $\succsim_0$, and so cannot deliver Axiom 4. Conversely, Axiom 4 concerns a single pair of likelihood vectors differing by a constant, and says nothing about mixtures, so it cannot deliver Axiom 1.

See Section 4 for an outline of the proof. Our primitive consists of preferences over acts on Ω, rather than over the richer domain $\Omega \times S$ as in Ghirardato's (2002) approach adapted to our setting (see Appendix A). This has two advantages. First, it demands less of the analyst, who need not observe choices over acts contingent on both states and signals. Second, the two approaches characterize different things. When acts are defined on $\Omega \times S$, the experiment is subjective and recovered from preference, so what is characterized is Bayesian updating with respect to a purely subjective joint distribution over $\Omega \times S$. With acts defined on Ω alone and an experiment σ objectively given, Theorem 1 tells us that the agent's posterior agrees with the Bayesian posterior computed from that experiment. We thereby characterize not just Bayesian updating, but Bayesian updating with a correctly specified experiment.

Remark. Theorem 1 organizes existing models of non-Bayesian updating by the normative property each of them gives up. Table 1 records which of Axioms 1–4 hold under several prominent rules.

Table 1: Axioms satisfied by alternative updating rules

	STP (1)	IID (2)	Homogeneity (3)	CwP (4)
Bayesian updating	✓	✓	✓	✓
Grether (1980)	×	✓	✓	✓
Epstein, Noor, and Sandroni (2008)	✓	×	✓	✓
Affine distortion (Whitmeyer, 2026)	✓	×	✓	×
Other rules[a]	×	×	✓	×

Note: STP, IID, and CwP stand for Informational STP (Axiom 1), Independence of Irrelevant Details (Axiom 2), and Consistency with Prior (Axiom 4), respectively. × indicates that the axiom may fail.
[a] Cripps (2018); Dominiak et al. (2025); Jakobsen (2026).

Notably, the one axiom that none of these rules relaxes is Homogeneity. We turn next to the significance of this.

2.3 Homogeneity as Complete Trust

The Bayesian model embodies complete trust that signals will indeed be drawn from the presented experiment σ. We argue that this is behaviorally embodied in Homogeneity. Consider the following thought experiment. If one gets a text message from a family member asking for money, one may immediately get concerned about what situation they might be in. But if one gets a text message from an acquaintance, one's immediate thought might instead be that this is a scam text. Intuitively, seeing unlikely signals makes us suspect the veracity of the information source. To connect with Homogeneity, imagine that texts from a family member and an acquaintance are respectively described by the experiments

$$\begin{array}{lcc} & ask & dont \\ Need & \multicolumn{2}{c}{\left[\begin{array}{cc} 2/3 & 1/3 \\ 1/3 & 2/3 \end{array}\right]} \\ No\ Need & & \end{array} \quad \text{and} \quad \begin{array}{lcc} & ask & dont \\ Need & \multicolumn{2}{c}{\left[\begin{array}{cc} 0.0002 & 0.9998 \\ 0.0001 & 0.9999 \end{array}\right]} \\ No\ Need & & \end{array}.$$

A scam text can pose as either of these, and we can imagine it to draw signals from the experiment ρ given by

$$\begin{array}{lcc} & ask & dont \\ Need & \multicolumn{2}{c}{\left[\begin{array}{cc} 1 & 0 \\ 1 & 0 \end{array}\right]} \\ No\ Need & & \end{array}.$$

We see that the likelihood vector corresponding to the "ask" signal for the acquaintance scales down the corresponding likelihood vector for the family member, making it ex ante very unlikely whatever one's prior is. But conditional on receiving "ask", one's posterior beliefs about the texter's need is not the same, since the weight one places on the text being a scam is different, precisely because the "ask" signal is much less likely in one case than the other.[7]

3 Bayesian Updating with Limited Trust

Having determined that the behavioral meaning of complete trust is Homogeneity, in this section we ask what updating rule is delivered by the remaining axioms of the Bayesian model.

3.1 Foundations

Begin with some definitions. Say that $\{\succsim_0, \succsim_{\sigma,s}\}_{(\sigma,s)\in(\Sigma\times S)}$ satisfies *idempotence* if, for some $\sigma^* \in \Sigma$,

$$\succsim_{\sigma^*,s} = \succsim_0 \text{ for all } s \in S. \tag{2}$$

That is, there exists an experiment σ^* in Σ which the agent treats as uninformative, with all the posterior beliefs being the same as the prior. In the context of the Bayesian model any uninformative experiment would serve as σ^*, but Idempotence is silent on what σ^* is in the current context.

Define a *quasi-experiment* as a non-zero mapping $\rho : \Omega \to \mathbb{R}^S_+$, with $\rho(s \mid \omega) \geq 0$ for all $s \in S$, and $\sum_{s\in S} \rho(s \mid \omega) \leq 1$ for all $\omega \in \Omega$. This weakens the notion of an experiment by dropping the requirement that $\sum_{s\in S} \rho(s \mid \omega) = 1$. A quasi-experiment ρ can be interpreted as an experiment with an additional signal s^* outside S, occurring with probability $\rho(s^* \mid \omega) = 1 - \sum_{s\in S} \rho(s \mid \omega)$ at each $\omega \in \Omega$. Under this reading, the agent's perceived signal space is larger

[7]This intuition exists in Gentzkow and Shapiro (2006), though it is not expressed in terms of behavior. Ortoleva (2012) and Ba (2026) also feature such an intuition but in the context of a non-Bayesian model where observing an event deemed too unlikely under the prior prompts the agent to revise their prior.

than the analyst's, and the presented experiment σ is understood to produce the additional signal s^* with probability zero.

We find that dropping Homogeneity from the Bayesian model yields:

Theorem 2. *$\{\succsim_0, \succsim_{\sigma,s}\}_{(\sigma,s)\in(\Sigma\times S)}$ satisfies Axioms 1, 2, 4, and 5 if and only if there exists a quasi-experiment ρ and $\tau \in (0,1]$ such that for all $(\sigma, s) \in \Sigma \times S$,*

$$p_\sigma(\cdot \mid s) = p^*_{\tau\sigma+(1-\tau)\rho}(\cdot \mid s). \tag{3}$$

Moreover, there exists a unique maximal trust $\tau^* \in (0,1]$ and a quasi-experiment ρ^* that satisfies (3). If $\tau^* < 1$, then ρ^* is also unique.

Theorem 3. *Suppose $\{\succsim_0, \succsim_{\sigma,s}\}_{(\sigma,s)\in(\Sigma\times S)}$ satisfies idempotence. Then Axioms 1, 2, 4, and 5 are satisfied if and only if there exists an experiment ρ and $\tau \in (0,1]$ such that for all $(\sigma, s) \in \Sigma \times S$,*

$$p_\sigma(\cdot \mid s) = p^*_{\tau\sigma+(1-\tau)\rho}(\cdot \mid s). \tag{4}$$

Moreover, τ is unique, and ρ is unique if $\tau < 1$.

Thus, an agent who satisfies all the Bayesian axioms except Homogeneity can be represented as if she updates using Bayes' rule, but with respect to the mixture $\tau\sigma+(1-\tau)\rho$ rather than the presented experiment σ. The mixture is the reduction of a compound experiment in which each signal $s \in S$ is generated according to σ with probability τ and according to ρ otherwise, where the latter is in general a quasi-experiment but under Idempotence this is a unique experiment. Accordingly, we read τ as the agent's *trust* in σ, and ρ as her alternate theory of signal generation.

While the agent has prior trust τ over σ versus ρ, it should be appreciated that the agent *updates* this prior trust after observing a signal. For any experiment $e = \sigma, \rho$, denote the *ex ante* probability of signal s by $e_p(s) := \sum_{\omega'\in\Omega} e(s \mid \omega')p(\omega')$. Grouping the σ- and ρ-terms in the posterior and normalizing each by its own ex ante signal probability, the model can be rewritten as

$$p_\sigma(\omega \mid s) = \underbrace{\frac{\tau\sigma_p(s)}{\tau\sigma_p(s) + (1-\tau)\rho_p(s)}}_{\tau(\sigma|s)} p^*_\sigma(\omega \mid s) + \underbrace{\frac{(1-\tau)\rho_p(s)}{\tau\sigma_p(s) + (1-\tau)\rho_p(s)}}_{1-\tau(\sigma|s)} p^*_\rho(\omega \mid s),$$

where the belief $p^*_\sigma(\cdot \mid s)$ (*resp.* $p^*_\rho(\cdot \mid s)$) is the Bayesian posterior with respect to σ (*resp.* ρ). In the expression above, the agent's posterior is a weighted average of the two Bayesian posteriors, with the weight on experiment σ denoted as $\tau(\sigma \mid s)$. We interpret this term as the agent's *updated trust* in σ after observing s: it is the prior trust τ revised in light of the signal. This updated trust depends on the ex ante probability of s: if $\tau \in (0,1)$, then for every $s \in S$,

$$\tau(\sigma \mid s) \geq \tau \iff \sigma_p(s) \geq \rho_p(s),$$

that is, trust in σ rises at signal s if and only if s is more likely under σ than under ρ according to the agent's prior. This is the source of the Homogeneity violation in the model, and it formalizes the intuition of Section 2.3.

Together with Theorem 1, our results validate the intuition that Homogeneity is the behavioral content of complete trust, therefore yielding a testable distinction between a Bayesian agent with limited trust and a non-Bayesian agent. Indeed, the results reflect favorably on the fact noted in Section 2.2 that prominent models of non-Bayesian updating in the literature typically retain Homogeneity: models that violate Homogeneity might be harder to rationalize as being non-Bayesian if they can be rationalized by a Bayesian model with limited trust.

Besides presenting the exhaustive testable implications of a lack of trust, the theorem also admits a normative reading. Suppose the agent harbors uncertainty about the experiment presented to her—and accordingly violates Homogeneity—but adheres to the remaining axioms as normative criteria for updating. By the theorem, she must then update by Bayes' rule with respect to an enlarged model space, adjoining a subjective ρ to the presented σ. Thus, uncertainty about the information structure calls for a revision of the agent's model space, not of the Bayesian paradigm.

3.2 Interpretation of ρ

The representation (4) can be viewed as the reduced form of a richer environment perceived by the agent. We present three examples.

Example 1: Aggregation of Alternative Theories In the representation (4), ρ was introduced as a single theory of signal generation that is entertained as an alternate to σ. The agent may in fact entertain several such theories. While the analyst presents her with σ, she may have been presented with other experiments $\{\rho_i\}_{i\in I}$ by sources the analyst does not observe, or may have formed them on her own. For instance, a reader may be told that a news outlet reports according to σ, while other commentators claim that it in fact reports according to $\{\rho_i\}_{i\in I}$, each reflecting a different assessment of the outlet's editorial slant. Another example is a letter of recommendation. A letter could be taken at face value, implying a particular mapping σ between the qualities of a candidate and the words used to describe them. But one could also suspect that something is being relayed between the lines, suggesting different possible mappings $\{\rho_i\}_{i\in I}$ between the qualities of a candidate and the words used to describe them.

In such situations, uncertain which theory is correct, the agent may assign trust τ to σ and trust τ_i to each ρ_i, with $\tau + \sum_i \tau_i = 1$, and update her beliefs using the reduced experiment $\tau\sigma + (1-\tau)\rho$, where $\rho := \sum_{i\in I} \frac{\tau_i}{1-\tau}\,\rho_i$. Her behavior is thus observationally equivalent to the single-theory case, with ρ being the trust-weighted average of the alternatives. This resembles opinion pooling (DeGroot, 1974), with two differences: the agent pools experiments rather than priors, and the components ρ_i are not observed by the analyst.

Example 2: Distortions of σ The agent's alternative theory need not be a fixed experiment unrelated to σ. She may instead suspect that signals are generated by a distorted version $\rho(\sigma)$ of the experiment she is told, assigning probability π to the process being σ and $1-\pi$ to its being $\rho(\sigma)$. Her subjective experiment is $\pi\sigma + (1-\pi)\rho(\sigma)$. Whenever the distortion takes the form

$$\rho(\sigma) = \lambda\sigma + (1-\lambda)\rho, \qquad \lambda \in (0,1), \tag{5}$$

for some experiment ρ, the subjective experiment $\pi\sigma+(1-\pi)\rho(\sigma)$ reduces to (4) with trust $\tau = \pi+(1-\pi)\lambda$. Here ρ is the fixed point of $\rho(\cdot)$, the one experiment the distortion leaves unchanged. Note that the experiment $\rho(\sigma)$ that the agent contemplates varies with σ, whereas the ρ recovered from her behavior does not.

A leading example of $\rho(\cdot)$ of the form (5) is Blackwell garbling. Suppose $S = \{s_1, s_2\}$, and consider an agent who fears that signals are garbled by some undisclosed flaw in the process: with probability k a realization of s_1 is reported as s_2, and with probability k' a realization of s_2 is reported as s_1, where $k+k' < 1$. The garbled experiment is

$$\rho(\sigma) = (1-k-k')\,\sigma + (k+k')\,\rho,$$

where ρ is the uninformative experiment reporting s_1 with probability $k'/(k+k')$, and is left unchanged by the garbling. Her trust in σ exceeds π, the probability she assigns to the process being undistorted, since garbling preserves part of σ.

Example 3: Strategic Sender The agent may lack trust in her information because she distrusts the sender. She may consider the possibility that the sender has personal motivations that cause her not to be truthful. To illustrate, consider a 2×2 setting with $\Omega = \{\omega_1, \omega_2\}$ and $S = \{s_1, s_2\}$, and restrict attention to the interior experiments in which s_1 is diagnostic of ω_1:[8]

$$\Sigma^1 := \left\{ \begin{bmatrix} \alpha & 1-\alpha \\ \beta & 1-\beta \end{bmatrix} : 0 < \beta < \alpha < 1 \right\}. \tag{6}$$

Suppose that with probability τ the agent faces a *truthful* sender, who runs the experiment $\sigma \in \Sigma^1$, and with probability $1-\tau$ a *strategic* sender, who announces σ to mimic the truthful type but reports signals drawn from some ρ of her choosing.[9] The strategic sender targets only the receivers who take signals at face value (that is, trust completely) and seeks to raise their belief in ω_1. Since s_1 is diagnostic of ω_1 under every $\sigma \in \Sigma^1$, she reports s_1 regardless of the state, so that

$$\rho = \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix} \qquad \text{for every } \sigma \in \Sigma^1.$$

[8] Our representation theorems assume $|\Omega| \geq 4$ only to establish the sufficiency of our axioms. This restriction may consequently be ignored in applications, but the two-state setting here can also be read without loss of generality as a binary partition $\{E, E^c\}$ of such an Ω, with ω_1 standing for E and ω_2 for E^c, and with attention restricted to experiments whose likelihoods are constant on each cell.

[9] More generally one can assume that the strategic sender adheres with probability λ to her commitment to draw a signal from σ, but with probability $1-\lambda$ deviates and instead draws from some optimally chosen ρ. This generates a distortion of the form (5).

A Bayesian agent in the population who recognizes that the sender may be strategic will then interpret signals according to our model.

If the strategic sender maximizes a generic objective over posterior distributions, the optimal ρ may depend on σ. It is nevertheless locally constant under either of two conditions. First, adjustment frictions may make small changes in σ insufficient to induce a change in ρ. Second, if ρ is chosen from a finite feasible set, then any strictly optimal choice remains optimal in a neighborhood of σ. Thus our model applies to any open Σ over which the optimal ρ is constant. Finite feasible sets arise naturally. For example, a student may choose among courses on the same topic taught by professors of differing strictness. In a job interview, he may report only the topic, suggesting the typical grading standard σ, while having chosen a course with a lenient grading standard ρ. Farina and Herman (2026) study communication in environments of this kind.[10]

3.3 The Role of Σ

Representation (4) holds ρ fixed as the stated experiment σ varies over Σ. This is a restriction on the agent's behavior across Σ, and its plausibility can depend on Σ. To illustrate this, recall that in the strategic sender example above, we saw that $\rho = \left[\begin{smallmatrix} 1 & 0 \\ 1 & 0 \end{smallmatrix}\right]$ is invariant over Σ^1. This ceases to be so once the space of experiments is enlarged beyond Σ^1. Let $\Sigma := \left\{ \begin{bmatrix} \alpha & 1-\alpha \\ \beta & 1-\beta \end{bmatrix} : \alpha, \beta \in (0,1),\ \alpha \neq \beta \right\}$, which also contains experiments in which s_1 is diagnostic of ω_2, namely those with $\alpha < \beta$. Facing such a σ, the strategic sender reports s_2 instead, so that her signals are drawn from $\rho' = \left[\begin{smallmatrix} 0 & 1 \\ 0 & 1 \end{smallmatrix}\right]$. No single ρ describes her reports across all of Σ. The invariance of ρ is thus a property of the set of experiments, and one the analyst secures by restricting its range.

Representation (4) also holds τ fixed as σ varies over Σ. One can imagine a deeper inference process where the agent infers the trustworthiness of the sender from the σ that they offer. Here, "being told σ" is itself a signal in an inference problem where the states are whether or not the sender is truthful, and the signals are "stated experiments". Trust would then be a function $\tau(\sigma)$, varying with the announcement in the way any posterior varies with its signal. The representation therefore requires a Σ over which the sender's credibility does not vary.

4 Projective Geometry of Bayesian Updating

The preceding sections characterize Bayesian updating with complete and limited trust. Although the two representations differ in how the agent interprets

[10]In Farina and Herman (2026) the sender observes the state before choosing, so she may select a different experiment at each ω. Write $\rho(\cdot \mid \omega)$ for the signal distribution she induces at ω, namely the row corresponding to ω in the experiment she selects there. Collecting these rows gives a single experiment ρ, which generates exactly the signals the receiver observes. Our model therefore applies.

the experiment, they share a common geometric property that we define below as *Preservation of Collinearity.* As we show below, it identifies a parameterized family of updating rules that contains complete-trust updating, limited-trust updating, and more general forms of non-Bayesian updating that go beyond our model.

To formulate this property, fix a signal realization $s \in S$ and define

$$\phi_s : \Sigma_s \to \Delta(\Omega),$$

where $\phi_s(\sigma_s)$ is the posterior belief represented by $\succsim_{\sigma,s}$. Thus, ϕ_s maps each likelihood vector

$$\sigma_s = \big(\sigma(s \mid \omega)\big)_{\omega\in\Omega} \in \Sigma_s$$

into the posterior induced upon observing s. Under complete-trust Bayesian updating, $\phi_s(\sigma_s) = \left(\frac{p(\omega)\sigma(s|\omega)}{\sum_{\omega'\in\Omega} p(\omega')\sigma(s|\omega')}\right)_{\omega\in\Omega}$.

Informational STP implies that ϕ_s satisfies *Preservation of Collinearity.* A recent version of the *Fundamental Theorem of Projective Geometry* due to Sancho de Salas (2026) then restricts ϕ_s to a fractional-linear form. The remainder of the section derives this geometric implication, illustrates it for complete- and limited-trust updating, and uses it to obtain the general representation underlying Theorems 1–3.

4.1 Geometric Content of Axiom 1

Our strategy is to pin down the updating rule ϕ_s through the geometric structure it preserves, using tools from projective geometry. The first step is therefore to recast Informational STP (Axiom 1) – a condition stated in terms of preferences – as a geometric property of the posterior ϕ_s. This property is *Preservation of Collinearity*: whenever three likelihood vectors in Σ_s are collinear (lie on a common line in $\mathbb{R}^{|\Omega|}$), their images under ϕ_s are collinear as well. Formally, ϕ_s *preserves collinearity* if for all $\sigma_s, \sigma'_s, \alpha\sigma_s + (1-\alpha)\sigma'_s \in \Sigma_s$, there is $\gamma \in [0, 1]$ such that

$$\phi_s\big(\alpha\sigma_s + (1-\alpha)\sigma'_s\big) = \gamma\,\phi_s(\sigma_s) + (1-\gamma)\,\phi_s(\sigma'_s).$$

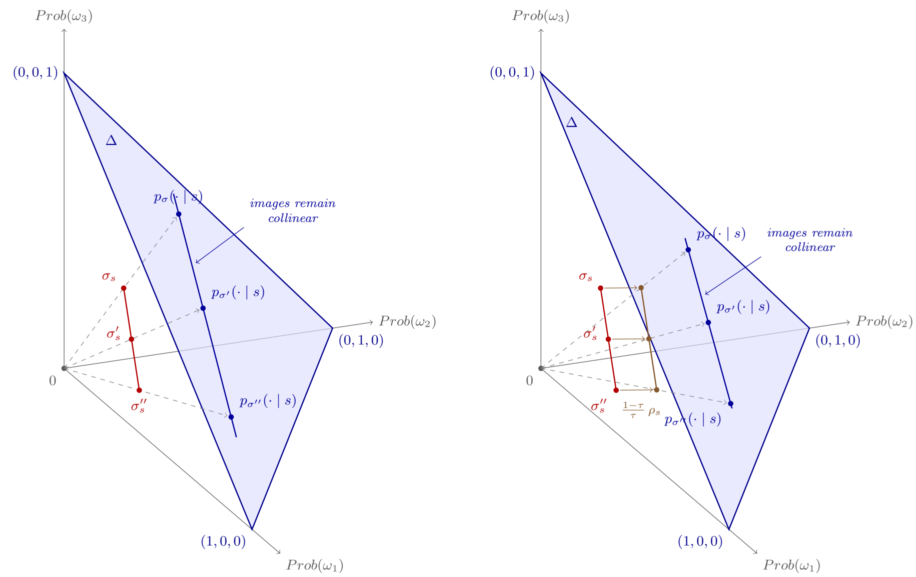
To see why the axiom reduces to this property, recall that it requires the following: if the experiment σ'' is a convex combination of σ and σ', then

$$f \succsim_{\sigma,s} g \text{ and } f \succsim_{\sigma',s} g \implies f \succsim_{\sigma'',s} g. \tag{7}$$

The three conditional preferences $\succsim_{\sigma,s}$, $\succsim_{\sigma',s}$, and $\succsim_{\sigma'',s}$ admit SEU representations with a common utility index and subjective beliefs $\phi_s(\sigma_s)$, $\phi_s(\sigma'_s)$, and $\phi_s(\sigma''_s)$, respectively. A standard separating-hyperplane argument then shows that (7) implies that $\phi_s(\sigma''_s)$ is a convex combination of $\phi_s(\sigma_s)$ and $\phi_s(\sigma'_s)$. Since σ''_s is itself a convex combination of σ_s and σ'_s, and since any three collinear vectors have one lying between the other two, this says exactly that ϕ_s carries collinear likelihood vectors to collinear posteriors—that is, ϕ_s preserves collinearity.

4.2 Necessity of Axiom 1

We begin by observing that both complete- and limited-trust Bayesian updating preserve collinearity. Beyond delivering the necessity of Axiom 1, this observation isolates a structural feature of these updating rules that will do the real work in the axiomatic characterization.



(a) Complete Trust

(b) Limited trust

Figure 1: Preservation of Collinearity

Figure 1 illustrates how Bayesian updating preserves collinearity. To fix ideas, take $\Omega = \{\omega_1, \omega_2, \omega_3\}$ with a uniform prior. For each $\sigma_s \in \Sigma_s$, if the agent fully trusts that σ is the true signal structure, then Bayesian updating yields the posterior

$$\phi_s(\sigma_s) = \frac{\sigma_s}{\|\sigma_s\|_1},$$

which admits a simple geometric reading (Figure 1a): draw the ray from the origin through σ_s and take its intersection with the simplex (the blue plane); that intersection is $\phi_s(\sigma_s)$. This map is a *central projection from the origin*. It preserves collinearity because three collinear points and the origin span a common plane, and the resulting posteriors lie on the line in which this plane meets the simplex. Limited-trust updating adds one step to this picture: it is the composition of a *translation* followed by the same *central projection from the origin*. Indeed, if the agent attaches trust $\tau \in (0, 1]$ to σ and $1 - \tau$ to ρ, the posterior is the normalization of $\tau\sigma_s + (1 - \tau)\rho_s$. Since normalization is unaffected by scaling by the positive constant τ, this is also the normalization of $\sigma_s + \frac{1-\tau}{\tau}\rho_s$:

$$\sigma_s \xrightarrow{\text{translation}} \sigma_s + \tfrac{1-\tau}{\tau}\rho_s \xrightarrow{\text{projection}} \frac{\sigma_s + \frac{1-\tau}{\tau}\rho_s}{\left\|\sigma_s + \frac{1-\tau}{\tau}\rho_s\right\|_1}.$$

Since both transformations preserve collinearity, Bayesian updating with limited trust also preserves collinearity.

4.3 Sufficiency of Axiom 1

The preceding discussion shows that Preservation of Collinearity is (a) the geometric content of Axiom 1 (Informational STP) and (b) a structural feature shared by complete- and limited-trust Bayesian updating. The property does more: it is also *sufficient* to pin down a parameterized family of updating rules that nests complete- and limited-trust Bayesian updating as special cases. Establishing this is the key step in the sufficiency part of the proof; once the family is obtained, the remaining axioms serve to discipline its parameters and single out the representation.

The tool that delivers this step is the *Fundamental Theorem of Projective Geometry*, which we review in Appendix B for the benefit of the reader. The classical versions of the theorem (Faure and Froelicher, 1994; Havlicek, 1994; see also Theorem 1.18 in Sancho de Salas, 2026) apply to mappings that have an unbounded domain, whereas the domain of an updating rule—a set of likelihood vectors—is bounded. The recent extension by Sancho de Salas (2026, Theorem 3.11) covers bounded domains and enables us to establish the following key mathematical result, stated as Theorem 8 in the appendix: *for $n \geq 4$ and open $C \subset \mathbb{R}^n$, if a function $f : C \to \mathbb{R}^n$ preserves collinearity and its image is not contained in a plane, then there exist $c \in \mathbb{R}$, $a, b \in \mathbb{R}^n$, and $M \in \mathbb{R}^{n\times n}$ such that*

$$f(v) = \frac{M \cdot v + a}{b^\top \cdot v + c}, \qquad v \in C. \tag{8}$$

This key abstract result in turn leads to the following representation result for updating rules that preserve collinearity, which is the key step described above. It shows that Axiom 1—through its geometric content, Preservation of Collinearity—together with an appropriate version of the Fundamental Theorem of Projective Geometry, delivers a broad family of updating rules that nests Bayesian updating with both complete and limited trust. This yields the most general representation result this paper has to offer:

Theorem 4. *Suppose the image of $\phi_s : \Sigma_s \to \Delta(\Omega)$ is not contained in a plane in $\Delta(\Omega)$. Then the following statements are equivalent:*

- *ϕ_s satisfies Preservation of Collinearity;*
- *there exist a quasi-experiment ρ, $\tau \in (0, 1]$, and a $|\Omega| \times |\Omega|$ matrix M_s*

such that

$$\phi_s(\sigma_s) = \left(\frac{[\tau M_s \cdot \sigma_s + (1-\tau)\rho_s]_\omega \cdot p(\omega)}{\sum_{\omega'} [\tau M_s \cdot \sigma_s + (1-\tau)\rho_s]_{\omega'} \cdot p(\omega')} \right)_{\omega \in \Omega}, \tag{9}$$

where $\mathrm{rank}[M_s \ \rho_s] = |\Omega|$.

While Theorem 4 is belief-theoretic, its decision-theoretic counterpart (that is, stated in terms of $\succsim_{\sigma,s}$ rather than ϕ_s) is given as Theorem 9 in the appendix. Theorems 1 and 2 are specializations of Theorem 9.

The representation (9) reads as Bayesian updating with prior p applied to the *distorted* likelihood vector

$$\tau M_s \cdot \sigma_s + (1-\tau)\rho_s.$$

Each component has a natural interpretation: $M_s \cdot \sigma_s$ is the likelihood vector in σ as perceived by the agent, ρ_s is her alternative interpretation of the signal realization in case σ is not the true signal structure, and τ is her confidence in σ. The matrix M_s thus permits systematic distortions in how the agent perceives likelihoods—distortions that are not attributable to lack of trust. In our representation results, Axioms 2 and 4 force M_s to be the identity matrix, thereby ruling out such perceptual distortions and reducing the family to Bayesian updating with limited trust.

5 Systematic Non-Homogeneity

Agents often react to surprising information differently from the way Bayes' rule with complete trust would prescribe: they may dismiss a surprise as noise, or take it more seriously than it warrants (see Ortoleva, 2012, 2024, for a discussion of the evidence). Such responses arise naturally when Homogeneity fails. Scaling a likelihood vector changes the ex ante probability of the signal without changing its likelihood ratios. Thus, when Homogeneity is violated, the agent's posterior response may depend on how surprising the signal is, even when its informational content about the state is held fixed. This section defines underreaction and overreaction to surprises behaviorally and characterizes them in terms of the alternative theory ρ.

5.1 Underreaction to Surprises

We begin by providing a behavioral definition.

Axiom 6 (Underreaction to Surprises)**.** *If* $\sigma'_s = \alpha\sigma_s$ *for* $\alpha \in (0,1)$, *then*

$$f \succsim_0 g \text{ and } f \succsim_{\sigma,s} g \implies f \succsim_{\sigma',s} g.$$

Axiom 6 is a weakening of *Homogeneity* that restricts its violations to underreaction to surprises. To illustrate, consider an equivalent formulation of it: if $\sigma'_s = \alpha\sigma_s$ for $\alpha \in (0,1)$, then

$$f \succsim_{\sigma,s} g \text{ and } f \prec_{\sigma',s} g \implies f \prec_0 g.$$

That is, whenever Homogeneity is violated, in the sense that f is weakly chosen over g given σ but g is strictly chosen over f given σ' (which scales down the likelihood of signal s), it must be because g is strictly preferred over f according to the *ex-ante* preference. In this sense, making the signal more unexpected can only move the agent's ranking toward her ex ante ranking, never away from it. The axiom thus concerns how the agent responds when the same information arrives more unexpectedly, rather than prescribing conservatism at every signal, as in Epstein, Noor, and Sandroni (2008) and Kovach (2021).

Axiom 6 has the same form as Consistency with Prior (Axiom 4), but applies to a multiplicative rather than an additive perturbation of the likelihood vector. Axiom 4 makes the signal less informative, whereas Axiom 6 makes it less likely without changing its likelihood ratios.

Proposition 1. *Suppose $\{\succsim_0, \succsim_{\sigma,s}\}_{(\sigma,s)\in\Sigma\times S}$ satisfies (4) with $\tau < 1$. Then Axiom 6 holds if and only if ρ is uninformative:*

$$\rho(s \mid \omega) = \rho(s \mid \omega') \qquad \textit{for all } s \in S \textit{ and } \omega, \omega' \in \Omega.$$

Underreaction to surprises therefore corresponds to an entirely uninformative alternative theory. As a signal becomes less likely, the agent assigns greater relative importance to a process that reveals nothing about the state, pulling her posterior toward her prior.

Axiom 6 is a comparative static on how the agent responds when a signal becomes less likely. The next result relates it to a global notion of underreaction. We say that the agent's beliefs exhibit *conservatism* if her posterior underreacts relative to the Bayesian update for all experiments $\sigma \in \Sigma$ and signals $s \in S$,

$$p^*_\sigma(\omega|s) \le p_\sigma(\omega|s) \le p(\omega) \text{ or } p(\omega) \le p_\sigma(\omega|s) \le p^*_\sigma(\omega|s), \qquad \omega \in \Omega.$$

While violations have been documented in the literature, conservatism is a notably robust finding in experiments (Benjamin, 2019). We observe that

Proposition 2. *Suppose $\{\succsim_0, \succsim_{\sigma,s}\}_{(\sigma,s)\in\Sigma\times S}$ satisfies (4). If Underreaction to Surprises holds then the agent's beliefs exhibit conservatism. The converse holds if Σ_s contains at least one constant likelihood vector, for each s.*

Given Proposition 1, we see that an uninformative ρ generates conservatism in beliefs. The converse does not necessarily hold. Intuitively, even if ρ is informative, the agent will exhibit conservatism as long as it is *less* informative than all $\sigma \in \Sigma$. The proposition shows that the converse holds nevertheless under a very mild richness condition on Σ.

5.2 Overreaction to Surprises

We observe next that the characterization of overreaction is not simply the converse of the one obtained above for underreaction. We first present the overreaction analog of Axiom 6 above.

Axiom 7 (Overreaction to Surprises)**.** *If* $\sigma'_s = \alpha\sigma_s$ *for* $\alpha \in (0,1)$*, then*

$$f \succsim_0 g \text{ and } f \succsim_{\sigma',s} g \implies f \succsim_{\sigma,s} g. \tag{10}$$

While Axiom 7 (like Axiom 6) is "global" in the sense of imposing a property that holds for all experiments $\sigma \in \Sigma$, one can also define a "local" version for a given experiment $\sigma \in \Sigma$: the agent *overreacts to surprises at* σ if (10) holds for all $s \in S$ and all $\sigma' \in \Sigma$ with $\sigma'_s = \alpha\sigma_s$ for some $\alpha \in (0,1)$. To illustrate why (10) restricts violations of Homogeneity at σ to overreaction to surprises, consider an equivalent formulation of it: for all $s \in S$ and all $\sigma' \in \Sigma$ with $\sigma'_s = \alpha\sigma_s$ for some $\alpha \in (0,1)$,

$$f \succsim_0 g \text{ and } g \succ_{\sigma,s} f \implies g \succ_{\sigma',s} f.$$

That is, whenever the agent's ranking departs from her *ex-ante* ranking at σ, it must also depart at σ', which scales down the likelihood of signal s. In this sense, making the signal more unexpected can only move the agent's ranking away from her ex ante ranking, never back toward it.

One could imagine a "pathological" reaction to signals whereby the posterior moves in the opposite direction to the Bayesian update, which is known in the literature as the *backfire effect*.[11] While possible in our model,[12] we exclude it in our results below for a more focused analysis. Say that the agent exhibits *no backfire* at σ if for all $s \in S$ and $\omega \in \Omega$,

$$p^*_\sigma(\omega|s) \le p(\omega) \iff p_\sigma(\omega|s) \le p(\omega).$$

Under this assumption we obtain the following characterization of overreaction to surprises.

Proposition 3. *Suppose* $\{\succsim_0, \succsim_{\sigma,s}\}_{(\sigma,s)\in\Sigma\times S}$ *satisfies* (4) *with* $\tau < 1$*. Then:*

- *Axiom 7 canneot hold;*
- *If the agent exhibits no backfire at* σ*, then she overreacts to surprises at* σ *if and only if for each* $s \in S$ *there are* $\alpha_s, \beta_s \ge 0$ *such that*

$$\rho_s = \alpha_s\sigma_s - \beta_s 1_\Omega.$$

The first part of Proposition 3 states that overreaction to surprises cannot hold at every experiment in Σ. This is a sharp asymmetry with conservatism: an uninformative alternative theory generates conservatism toward surprises at every experiment, whereas no ρ generates overreaction everywhere, for any Σ satisfying our assumptions. The second part gives a local characterization. At a given experiment, overreaction requires the alternative likelihood vector to be an

[11] See Nyhan and Reifler (2010, 2015) and Haglin (2017).

[12] A backfire effect occurs in the model when σ and $\pi := \tau\sigma + (1-\tau)\rho$ push beliefs in different directions at some signal. For instance, if $\sigma(s|\omega) > \sigma_p(s)$ and $\pi(s|\omega) \le \pi_p(s)$ then σ suggests a higher belief in ω after signal s but the agent places a lower belief on it instead because she follows π.

affine transformation of the agent's likelihood vector, with a negative constant component.

The last result confirms that overreaction to surprises implies overreaction to information. We say that the agent exhibits *overreaction* at $\sigma \in \Sigma$, if for all $s \in S$ and $\omega \in \Omega$,

$$p_\sigma(\omega|s) \leq p^*_\sigma(\omega|s) \leq p(\omega) \text{ or } p(\omega) \leq p^*_\sigma(\omega|s) \leq p_\sigma(\omega|s).$$

Proposition 4. *Suppose $\{\succsim_0, \succsim_{\sigma,s}\}_{(\sigma,s)\in\Sigma\times S}$ satisfies* (4)*. Then if the agent exhibits no backfire and overreacts to surprises at σ, then the agent exhibits overreaction at σ.*

6 Prior Dependence

The preceding analysis explores how posteriors vary with $\sigma \in \Sigma$ for a fixed prior p. In this section we explore how posteriors may vary with the prior p at a fixed $\sigma \in \Sigma$. Such dependence is, for instance, the content of *confirmation bias*, the well-known finding that the response to information is biased towards the prior, which we will place some emphasis on below.[13] For simplicity, throughout this section we restrict attention to the 2×2 setting of Section 3.2, where the state space is $\Omega = \{\omega_1, \omega_2\}$ and the signal space is $S = \{s_1, s_2\}$, and σ lies in the set Σ^1 defined in (6), that is, the (interior of the) set of experiments in which s_1 is diagnostic of ω_1.

6.1 Non-Martingale Properties

In this section we assume σ is an objectively true experiment, thereby determining the state-contingent frequencies of signals, but the agents do not fully trust this experiment. We consider a family of agents with the same objective experiment σ, subjective experiment ρ, and the same trust parameter τ, but different priors p, denoted as $(\sigma, \rho, \tau, p)_{p\in(0,1)}$. We define different kinds of behavioral biases via systematic violations of the martingale property of Bayesian updating. For a prior p and a state ω, write

$$\bar{q}_p(\omega) := \sum_{s\in S} p^*_{\tau\sigma+(1-\tau)\rho}(\omega \mid s)\, \sigma_p(s)$$

for the agent's average posterior on ω, where the posterior is formed under the subjective experiment $\tau\sigma + (1-\tau)\rho$ while signals arrive according to the objective experiment σ. A fully trusting Bayesian has $\bar{q}_p(\omega) = p(\omega)$.

Definition 1 (Behavioral Biases)**.** *A family of agents $(\sigma, \rho, \tau, p)_{p\in(0,1)}$ exhibits*

[13] Gentzkow and Shapiro (2006) observe that posterior trust in an information source depends on the prior and suggest a connection between lack of trust and confirmation bias without clarifying the behavioral content of the connection. The results of this section can be viewed as filling this gap.

- ω-biased updating, *for* $\omega \in \{\omega_1, \omega_2\}$, *if* $\bar{q}_p(\omega) > p(\omega)$ *for all* $p \in (0,1)$*;*
- confirmatory updating *if there exists* $\theta \in (0,1)$ *such that, for any* $\omega \in \Omega$, $\bar{q}_p(\omega) > p(\omega)$ *whenever* $p(\omega) > \theta$*;*
- average base-rate neglect *if there exists* $\theta \in (0,1)$ *such that, for any* $\omega \in \Omega$, $\bar{q}_p(\omega) > p(\omega)$ *whenever* $p(\omega) < \theta$*.*

Under ω-biased updating, the agent's beliefs are pushed toward ω regardless of her prior. Under confirmatory updating, they are pushed toward ω when her prior in ω is high. That is, the prior is reinforced. Under average base-rate neglect the pattern is reversed, so that beliefs are pushed away from the state the prior favors, and extreme priors are drawn toward the middle. The next result shows that these three biases are *exhaustive*: in our model, any violation of the martingale property takes one of these forms.

Proposition 5. *Let* $(\sigma, \rho, \tau, p)_{p \in (0,1)}$ *be such that* $\tau < 1$ *and there is no backfire at* σ*. The family of agents:*

- *exhibits* ω_1*-biased updating if and only if*

$$\sigma(s_1 \mid \omega_1) \geq \rho(s_1 \mid \omega_1) \quad \textit{and} \quad \sigma(s_1 \mid \omega_2) \geq \rho(s_1 \mid \omega_2), \tag{11}$$

  where at least one of the two inequalities is strict;

- *exhibits confirmatory updating if and only if*

$$\sigma(s_1 \mid \omega_1) > \rho(s_1 \mid \omega_1) \quad \textit{and} \quad \sigma(s_2 \mid \omega_2) > \rho(s_2 \mid \omega_2); \tag{12}$$

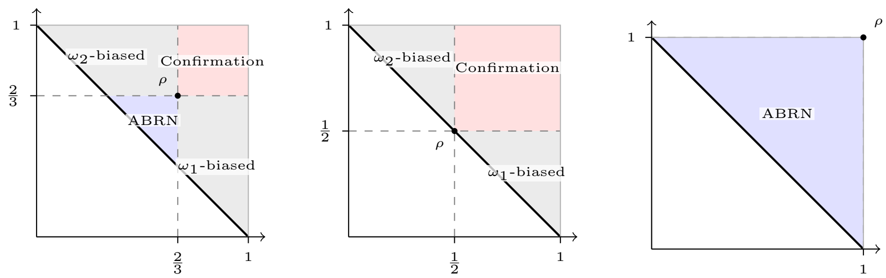
- *exhibits average base-rate neglect if and only if*

$$\sigma(s_1 \mid \omega_1) < \rho(s_1 \mid \omega_1) \quad \textit{and} \quad \sigma(s_2 \mid \omega_2) < \rho(s_2 \mid \omega_2). \tag{13}$$

Proposition 5 delivers three implications. First, the three forms of bias are exhaustive: in our model, every violation of the martingale property falls into one of these categories. Second, the bias exhibited by the agent does not depend on the degree of trust τ, but rather on σ and ρ alone. Trust governs the magnitude of the bias, not its form.

Third, the proposition identifies the region in which each bias arises. ω_1-biased updating occurs when σ is comparatively more skewed toward the signal s_1, in the sense that s_1 is comparatively more likely under σ in both states. For example, let ω_1 and ω_2 indicate whether a student's research is of high or low quality, and let s_1 and s_2 denote favorable and unfavorable feedback. Suppose σ describes an encouraging advisor who gives favorable feedback relatively frequently, regardless of the true quality of the research. The proposition predicts that the agent's average posterior on ω_1 exceeds her prior. Because the agent places some weight on ρ, under which the unfavorable signal s_2 arrives more often, she finds s_2 less surprising than a full-trust Bayesian would. She thus revises her belief less sharply downward upon observing s_2, and her beliefs are tilted toward ω_1 on average.

On the other hand, confirmatory updating occurs if σ generates the state-matching signal more often than ρ in both states, so σ is the more accurate of the two experiments. For an agent who already believes ω_1, the disconfirming signal s_2 is less likely under σ relative to ρ, so observing it lowers her trust in σ and she discounts such disconfirming evidence. Therefore, the posterior beliefs are tilted towards the state that the prior already favors. Conversely, average base-rate neglect occurs if ρ is the more accurate of the two experiments, in which case the tilt is reversed: the agent's posterior beliefs are pulled away from the state her prior favors, and extreme priors are drawn toward the middle.



(a) $\alpha = \frac{2}{3}$ (b) $\alpha = \frac{1}{2}$ (c) $\alpha = 1$

Figure 2: Behavioral biases as regions in $\big(\sigma(s_1 \mid \omega_1), \sigma(s_2 \mid \omega_2)\big)$.

Note: The horizontal axis is $\sigma(s_1 \mid \omega_1)$ and the vertical axis is $\sigma(s_2 \mid \omega_2)$. The point ρ marks $\big(\rho(s_1 \mid \omega_1), \rho(s_2 \mid \omega_2)\big)$, where $\alpha := \rho(s_1 \mid \omega_1) = \rho(s_2 \mid \omega_2)$ is the precision of the alternative theory. "ABRN" denotes average base-rate neglect.

Figure 2 illustrates the proposition. Any objective experiment σ is represented as a point $\big(\sigma(s_1 \mid \omega_1), \sigma(s_2 \mid \omega_2)\big)$ in the unit square, and we restrict attention to the region above the anti-diagonal, so that s_1 is diagnostic of ω_1. In panel (a), the subjective experiment ρ is symmetric with precision 2/3. The four quadrants that ρ induces correspond to the four biases: to its northwest and southeast lie the experiments skewed toward one signal, yielding ω_2-biased and ω_1-biased updating; to its northeast and southwest lie the experiments more and less informative than ρ, yielding confirmatory updating and average base-rate neglect. This correspondence is universal: whatever ρ, the four quadrants it induces map to the same four patterns. Panels (b) and (c) display two important special cases. When ρ is uninformative, limited trust makes the agent underreact to information (Proposition 1), and panel (b) shows that she can then only exhibit confirmatory updating or a bias toward one state—average base-rate neglect is impossible. When ρ is perfectly informative, limited trust makes the agent overreact, and panel (c) shows the reverse: every experiment above the anti-diagonal lies southwest of ρ, so average base-rate neglect is the only possible pattern.

6.2 Confirmatory Reading of Mixed Evidence

The striking finding in the previous subsection is that a prior-dependent notion such as confirmatory updating arises in the model for a family of agents for which ρ is prior-independent. While confirmatory updating is the clearest expression of the notion of "confirmation bias" that we find in our model, the notion is in fact defined differently in the psychology literature. In a classic experiment, Lord, Ross, and Lepper (1979) demonstrate that offering different subjects the same evidence on an issue (the effectiveness of capital punishment in deterring crime) leads them to believe in their prior positions on the issue more strongly, leading to increased polarization. A key feature of this study, which is not captured in Definition 1, is that the evidence presented to subjects was *mixed*. Specifically, subjects were presented with two studies, one that presented empirical evidence on the issue and one that criticised the first study's methodology. We study a definition of confirmation bias that formalizes this.

Consider a family of agents as before, but now allow ρ to depend on the prior. Denote the family by $(\sigma, \rho^p, \tau, p)_{p\in(0,1)}$. Consider again the 2×2 example. Suppose now that each agent is presented with the results of *two* studies. They are told that both results are essentially signals drawn from one experiment $\sigma \in \Sigma^1$. The agent, however, places only probability τ on a given signal being drawn from σ, placing the remaining probability $1-\tau$ on it being drawn from ρ^p. Thus, a pair of signals $s, s' \in S$ is viewed as independently generated by the subjective experiment $\tau\sigma + (1-\tau)\rho^p$.

The pair of signals s, s' are *mixed* if they are distinct. Since the order of signals will not matter in the model we will state our result just for the mixed signals $s_1 s_2$. Although the posterior $p(\cdot|s_1 s_2)$ depends on the experiment σ, we will suppress this in the notation since the context will be clear. In contrast to Definition 1, we define confirmatory reading of mixed evidence as a statement about a given agent.

Definition 2. *(Confirmatory Reading of Evidence) An agent with prior p exhibits* ω-confirmatory reading of mixed evidence *at σ if*

$$p_p(\omega|s_1 s_2) > p(\omega).$$

There is no confirmatory reading if $p(\omega|s_1 s_2) = p(\omega)$ for all $\omega \in \Omega$.

Being subjectively Bayesian, our agent exhibits ω_1-confirmatory reading if and only if $s_1 s_2$ is more likely to be generated by the subjective experiment $\tau\sigma + (1-\tau)\rho^p$ in state ω_1 than ω_2. This yields the first part of the following proposition. The second part presents a sharper characterization by restricting attention to the case where the agent never exhibits a backfire effect and to the responses to symmetric σ (that is, σ such that $\sigma(s_1|\omega_1) = \sigma(s_2|\omega_2)$).

Proposition 6. *(a) Consider any $\omega_i \in \Omega$. An agent with prior p satisfying our model* (4) *exhibits ω_i-confirmatory reading of mixed evidence at σ if and only if*

$$\left|\tau\sigma(s_1|\omega_i) + (1-\tau)\rho(s_1|\omega_i) - \frac{1}{2}\right| < \left|\tau\sigma(s_1|\omega_{-i}) + (1-\tau)\rho(s_1|\omega_{-i}) - \frac{1}{2}\right|.$$

(b) Suppose that the agent with prior p exhibits no backfire at any $\sigma \in \Sigma^1$. Then the agent exhibits ω_1-confirmatory reading (resp. ω_2-confirmatory reading, no confirmatory reading) for all symmetric $\sigma \in \Sigma^1$ iff

$$\rho^p(s_1|\omega_1) + \rho^p(s_1|\omega_2) < (\textit{resp.} \; >, =) 1.$$

Our key takeaways are as follows. First, while confirmatory updating arises even when all agents share the same ρ, the only way to produce confirmatory reading is for ρ^p to depend on p. This is evident from the fact that the expression in (a) does not involve the prior. Second, confirmatory bias and confirmatory reading are not equivalent in a model of limited trust. Confirmatory updating can exist without confirmatory reading – for instance, when $\rho(s_1 \mid \omega_1) = \rho(s_1 \mid \omega_2) = \frac{1}{2}$, no agent will exhibit confirmatory reading towards symmetric σ, but agents with strong priors will exhibit confirmatory updating. Similarly, confirmatory reading can exist without confirmatory updating – for instance, when $\frac{1}{2} - \rho(s_1|\omega_2) > \rho(s_1|\omega_1) - \frac{1}{2} > 0$ then the agent exhibits ω_1-confirmatory reading at all symmetric $\sigma \in \Sigma_1$ but does not exhibit confirmatory updating at any $\sigma \in \Sigma_1$ that are close to being uninformative.

A Dynamic Subjective Expected Utility: Review

Consider a finite state space Ω and a finite signal space S. An experiment is a mapping $\sigma : \Omega \to \Delta(S)$. Let X be a convex subset of a normed vector space. Let $\mathcal{F} = \{f : \Omega \to X\}$ and $\mathcal{F}_0 = \{F : S \to \mathcal{F}\}$. As usual, for any acts f, h and event $E \subset \Omega$,the act fEh pays according to f for all $\omega \in E$ and according to h otherwise. Similarly $f\{s\}H$ is a signal-act in $\mathcal{F}_0$ that yields f if the signal s obtains and otherwise yields acts according to H. Such acts allow the analyst to observe how the agent would bet on events and therefore elicit beliefs.

Consider an "ex ante" preference $\succsim_0$ over $\mathcal{F}_0$ and a family $\{\succsim_s\}_{s\in S}$ of "ex post" preferences over $\mathcal{F}$ that admit the following Subjective Expected Utility representations: For some non-constant affine $u : X \to \mathbb{R}$ and full-support distributions $p_0(\cdot) \in \Delta(\Omega \times S)$ and $q(\cdot|s) \in \Delta(\Omega \times S)$, $\succsim_0$ admits a representation

$$U(F) = \sum_{s\in S} \sum_{\omega\in\Omega} u(F_s(\omega)) p_0(\omega, s), \qquad F \in \mathcal{F}_0,$$

and for each $s \in S$, $\succsim_s$ admits a representation

$$U_s(f) = \sum_{\omega\in\Omega} u(f(\omega)) p(\omega \mid s), \qquad f \in \mathcal{F}.$$

To establish a connection between $p_0(\cdot)$ and each $p(\cdot|s)$, consider:

Definition 3 (Dynamic Consistency). *For all $f \in \mathcal{F}$, $H \in \mathcal{F}_0$ and $s \in S$,*

$$f\{s\}H \succsim_0 g\{s\}H \implies f \succsim_s g.$$

The following result is the standard justification for the use of Bayesian updating in economics.[14] The result states that any agent satisfying Dynamic Consistency can be represented as a Bayesian updater who subjectively perceives signals to be generated by some π. The subjective experiment π is uniquely identified. This implies that if the analyst observes that an agent appears to update in a non-Bayesian fashion with respect to some objective σ, then they need only check Dynamic Consistency to determine whether the agent can be represented as a Bayesian with some (misspecified) subjective experiment π.

Theorem 5. *$\succsim_0$ and $\{\succsim_s\}_{s\in S}$ satisfies Dynamic Consistency if and only if there exists a subjective experiment π such that for all $s \in S$,*

$$p(\cdot \mid s) = \frac{\pi(s|\omega) p_0(\omega)}{\sum_{\omega'} \pi(s|\omega') p_0(\omega')}.$$

The subjective experiment is unique and satisfies $\pi(s|\omega) := \frac{p_0(\omega,s)}{p_0(\omega)}$ for all s, ω.

[14]See Ghirardato (2002) for the earliest formalization of this folk wisdom in the context of partitional information.

Proof. Write $U(F) = \sum_{s\in S}\left[\sum_{\omega\in\Omega} u(F_s(\omega))p_0(\omega|s)\right] p_0(s)$, where $p_0(s) = \sum_{\omega\in\Omega} p_0(\omega, s)$ is the marginal over S and $p_0(\omega|s) := \frac{p_0(\omega,s)}{p_0(s)}$ is the conditional probability. Then by Dynamic Consistency, for all s, f, g,

$$\sum_{\omega\in\Omega} u(f(\omega))p_0(\omega|s) \geq \sum_{\omega\in\Omega} u(g(\omega))p_0(\omega|s) \iff f\{s\}H \succsim_0 g\{s\}H$$

$$\iff f \succsim_s g \iff \sum_{\omega\in\Omega} u(f(\omega))p(\omega|s) \geq \sum_{\omega\in\Omega} u(g(\omega))p(\omega|s).$$

By the uniqueness of the SEU representation, we obtain $p(\omega|s) = \frac{p_0(\omega,s)}{p_0(s)}$.Define a subjective experiment by $\pi(s|\omega) := \frac{p_0(\omega,s)}{p_0(\omega)}$. Then the posterior is a Bayesian update wrt to π:

$$p(\omega|s) = \frac{p_0(\omega, s)}{p_0(s)} = \frac{p_0(\omega, s)/p_0(\omega)}{\sum_{\omega'} p_0(\omega', s)} p_0(\omega)$$

$$= \frac{p_0(\omega, s)/p_0(\omega)}{\sum_{\omega'} \pi(s|\omega')p_0(\omega')} p_0(\omega) = \frac{\pi(s|\omega)p_0(\omega)}{\sum_{\omega'} \pi(s|\omega')p_0(\omega')}.$$

□

B Projective Geometry: A Review

B.1 Projective Spaces

A (non-zero) *ray* in Euclidean space $\mathbb{R}^m$ is the uni-dimentional subspace $[x] = \{\lambda x : \lambda \neq 0\}$ generated by some $x \in \mathbb{R}^m \backslash \{0\}$. We use $[x]$ to refer to a generic non-zero ray and, with some abuse of notation, we use "$x \in [x]$" to refer to any generic vector x in a given ray $[x]$ generated any generic x in the ray.

The *projective space* $\boldsymbol{P}(\mathbb{R}^m)$ *associated with* $\mathbb{R}^m$ is defined as the set of non-zero rays in $\mathbb{R}^m$,

$$\boldsymbol{P}(\mathbb{R}^m) = \{[x] : x \in \mathbb{R}^m \backslash \{0\}\}.$$

The mapping $x \mapsto [x]$ is the "canonical projection". With a slight abuse of notation, a non-zero ray $[x]$ in $\mathbb{R}^m$ corresponds to a *point* $[x]$ in $\boldsymbol{P}(\mathbb{R}^m)$. While not a vector space, projective spaces have a dimension and (the obvious) addition operation, as well as other notions such as a "basis".[15]

Say that the points $[x], [y], [z]$ in $\boldsymbol{P}(\mathbb{R}^m)$ are *collinear* if for any $x \in [x], y \in [y], z \in [z]$ there exists $\alpha, \beta \in \mathbb{R}$, not both 0, s.t. $x = \alpha y + \beta z$. Thus, $[x], [y], [z]$ are collinear in the projective space if they correspond to *coplanar* rays in $\mathbb{R}^m$ (rays that lie in two-dimensional subspace of $\mathbb{R}^m$), or equivalently, if any vectors x, y, z contained in the respective rays are linearly dependent. Note that by this definition, any pair of points $[x], [y] \in \boldsymbol{P}(\mathbb{R}^m)$ are collinear.

[15]The dimension of $\boldsymbol{P}(\mathbb{R}^m)$ is defined to be $m - 1$. For instance, if $m = 1$ then $\boldsymbol{P}(\mathbb{R}^m)$ contains one point (corresponding to the non-zero ray $\mathbb{R}\backslash\{0\}$), and is therefore 0 dimensional.

A subset $L \subseteq \boldsymbol{P}(\mathbb{R}^m)$ is a *line* if L corresponds to the set of all rays passing through some plane in $\mathbb{R}^m$. The line *passing through* $[x], [y]$, denoted $[x] \wedge [y]$, corresponds to the set of rays associated with the plane in $\mathbb{R}^m$ passing through $[x], [y]$. Similarly $L \subseteq \boldsymbol{P}(\mathbb{R}^m)$ is a *plane* if it corresponds to the set of rays passing through a 3 dimensional subspace of $\mathbb{R}^m$. A subset $S \subseteq \boldsymbol{P}(\mathbb{R}^m)$ is a *subspace* if $[x], [y] \in S \implies [x] \wedge [y] \subset S$, that is, it contains the line passing through each pair of points in S.

A subset $Z \subseteq \boldsymbol{P}(\mathbb{R}^m)$ is a *locally projective subspace of* $\boldsymbol{P}(\mathbb{R}^m)$ if (a) its subspaces are of the form $S \cap Z$ where S is a subspace of $\boldsymbol{P}(\mathbb{R}^m)$, and (b) for any line $L \subseteq \boldsymbol{P}(\mathbb{R}^m)$ either $L \cap Z = \emptyset$ or the cardinality of $L \cap Z$ is at least 2.

B.2 Morphisms

We defined collinear points in projective space above. For any vector space X, say that $x, y, z \in X$ are *collinear* if there exists $\alpha \in \mathbb{R}$ s.t. $x = \alpha y + (1 - \alpha)z$. Note that any pair of points $x, y \in X$ are collinear by this definition.

Definition 4. *Let X denote either a vector space or a projective space. Let Y denote any subset Y of X. A map $h : Y \to X$ is a* morphism *if for any collinear points $x, y, z \in Y$, the images $f(x), f(y)$ and $f(z)$ are collinear.*

The notion of "partial morphisms" that we define next allows us state to the Fundamental Theorem of Projective Geometry (Theorem 6) for the benefit of the reader. This can be skipped without loss of continuity, however, since our main result will use a "local" version of the Fundamental Theorem (Theorem 7) that does not make reference to partial morphisms.

A function φ with domain $dom(\varphi) \subseteq \boldsymbol{P}(\mathbb{R}^m)$ and image $im(\varphi) \subseteq \boldsymbol{P}(\mathbb{R}^m)$ is referred to as a *partial map* and denoted $\varphi : \boldsymbol{P}(\mathbb{R}^m) \dashrightarrow \boldsymbol{P}(\mathbb{R}^m)$. The set $E := \boldsymbol{P}(\mathbb{R}^m) - dom(\varphi)$ is referred to as the *exceptional set.* For any $S \subseteq \boldsymbol{P}(\mathbb{R}^m)$, define $\varphi(S) = \varphi(S \cap dom(\varphi)\}$.

Definition 5. *A partial map $\varphi : \boldsymbol{P}(\mathbb{R}^m) \dashrightarrow \boldsymbol{P}(\mathbb{R}^m)$ is a partial morphism if, for any $[x], [y] \in \boldsymbol{P}(\mathbb{R}^m)$,*

$$\varphi([x] \wedge [y]) \subseteq \varphi([x]) \wedge \varphi([y]).$$

That is, a partial morphism φ maps a line $[x] \wedge [y]$ to a set $\varphi([x] \wedge [y])$ which is contained in the line $\varphi([x]) \wedge \varphi([y])$. This corresponds to mapping a line to a line while taking care of the partial nature of the mapping.

B.3 Fundamental Theorem of Projective Geometry

The "Fundamental Theorem of Projective Geometry" is stated in various ways and different degrees of generality, and thus refers to a family of related results. The classic statements involve *collineations* (morphisms that are bijections whose inverse is also a morphism), which are not suitable for our purposes.[16] The key generalization to (partial) morphisms is:

[16]We study updating rules that are morphisms, but not bijections (for a given signal, they map a likelihood vector to a posterior, and thus have an n-dimensional unit cube as domain

Theorem 6 (The Fundamental Theorem of Projective Geometry)**.** *Suppose $m \geq 3$. If $\varphi : \boldsymbol{P}(\mathbb{R}^m) \dashrightarrow \boldsymbol{P}(\mathbb{R}^m)$ is a partial morphism for which the image is not contained in a line, then there exists an $m \times m$ matrix M such that*

$$\varphi([x]) = [M \cdot x] \text{ for all } [x] \in dom(\varphi).$$

Moreover, M is unique up to non-zero scalar multiplication.

See for instance Faure (2002, Theorem 3.1) or Sancho de Salas (2026, Theorem 1.18). The result stated in those papers asserts that φ is generated by a semilinear map on some vector space. But semilinear maps on Euclidean space are linear, and this yields the matrix M in the Theorem.

This formulation will prove not to be directly applicable in our setup since we need a version of the Fundamental Theorem that allows for a bounded domain.[17] The following formulation, due to a very recent paper by Sancho de Salas (2026, Theorems 1.18 and 3.11), weakens the domain requirement on φ, but strengthens the lower bound on m and the image requirement.

Theorem 7 (The Fundamental Theorem of Locally Projective Geometry)**.** *Suppose $m \geq 4$. Let $P \subseteq \boldsymbol{P}(\mathbb{R}^m)$ be a locally projective subspace and let $\varphi : P \to \boldsymbol{P}(\mathbb{R}^m)$ be a morphism whose image is not contained in a plane. Then there exists a $m \times m$ matrix M such that*

$$\phi([x]) = [M \cdot x] \text{ for all } [x] \in P.$$

Moreover, M is unique up to non-zero scalar multiplications.

Proof. Our definition of a locally projective subspace is in fact Sancho de Salas (2026, Proposition 3.7). P and $\boldsymbol{P}(\mathbb{R}^m)$ satisfy the conditions of Sancho de Salas (2026, Theorem 3.11) by the "elementary facts" stated in Sancho de Salas (2026, pg 15). Therefore by Sancho de Salas (2026, Theorem 3.11), the morphism φ extends uniquely to a partial morphism $\hat{\varphi} : \boldsymbol{P}(\mathbb{R}^m) \dashrightarrow \boldsymbol{P}(\mathbb{R}^m)$. Apply Theorem 6. □

Early precursors of the local result are Lenz (1958) and Lowen (1982).[18]

---

and $n-1$ dimensional simplex as range), and therefore their projection into projective space is not a collineation.

[17]When φ is a partial morphism, its domain must have particular structure: the so-called "exceptional set" $E := \boldsymbol{P}(\mathbb{R}^m) \backslash dom(\varphi)$ must be subspace of $\boldsymbol{P}(\mathbb{R}^m)$. This rules out, for instance, that $dom(\varphi)$ is a bounded subset of $\boldsymbol{P}(\mathbb{R}^m)$. In our setup, we will generate a mapping φ through an updating rule, which (for a given signal s) maps a likelihood vector $v \in [0,1]^m$ to a posterior $p_v(\cdot|s) \in [0,1]^m$. Since the domain of the updating rule, namely $[0,1]^m$, is bounded it will correspond only to some bounded subset of $\boldsymbol{P}(\mathbb{R}^m)$.

[18]Lenz (1958, Hilfssatz 3, p. 348) states and proves that "Jede geradentreue Abbildung eines abgeschlossenen Simplex laesst sich, wenn nicht alle Bildpunkte in einer Geraden liegen, eindeutig zu einer – unter Umstaenden ausgearteten – projektiven Kollineation des Raumes fortsetzen," that is, "Any line-preserving map on a closed simplex, provided that its image is not contained in a line, extends uniquely to a projective collineation of the ambient space, possibly a degenerate one." When stated for Euclidean spaces, Lowen (1982) proves that for $m \geq 3$ and any open subset $U \subset \mathbb{R}^m$, a continuous injective morphism $\varphi : \boldsymbol{P}(U) \to \mathbf{P}(\mathbb{R}^m)$ extends uniquely to a morphism on $\boldsymbol{P}(\mathbb{R}^m)$.

C A Representation for Local Morphisms

Using a proof strategy that is familiar from applications of the Fundamental Theorem, we prove a representation result for functions $f : C \to \mathbb{R}^n$ that belong to the following class:

Definition 6. *(Local Morphism) A mapping $f : C \to \mathbb{R}^n$ is a local morphism if*
a) $\emptyset \neq C \subset \mathbb{R}^n$ is open.
b) $f : C \to \mathbb{R}^n$ is a morphism.
c) $f(C)$ is not contained in a plane in $\mathbb{R}^n$.

Our axiomatization of the limited trust model will use the representation result. We proceed by projecting f to some function φ on a locally projective subset of $\boldsymbol{P}(\mathbb{R}^{n+1})$. Note that f is defined on a subset of $\mathbb{R}^n$ and, instead of $\boldsymbol{P}(\mathbb{R}^n)$, we define φ on $\boldsymbol{P}(\mathbb{R}^{n+1})$. Applying the Fundamental Theorem for Locally Projective Spaces (Theorem 7) yields a representation for φ, which induces a representation for f.

C.1 Defining φ

Define $\pi : \mathbb{R}^n \to \mathbb{R}^{n+1} \backslash \{0\}$ by mapping each $v = (v_1, ..., v_n) \in \mathbb{R}^n$ to $\pi(v) = (v_1, ..., v_n, 1) \in \mathbb{R}^{n+1} \backslash \{0\}$. Refer to this as the "Homogeneous coordinate" representation for $v \in \mathbb{R}^n$ and, as conventionally done, denote it using colons as follows:

$$\pi(v) = (v_1 : ... : v_m : 1) \in \mathbb{R}^{n+1} \backslash \{0\}.$$

Conversely, let $E := \mathbb{R}^n \times \{0\}$ and define an "inverse" $\pi^{-1} : \mathbb{R}^{n+1} \backslash E \to \mathbb{R}^n$ by mapping each $\tilde{v} = (v_1 : ... : v_n : v_{n+1}) \in \mathbb{R}^{n+1} \backslash E$ to its so-called "Cartesian coordinate" representation in $\mathbb{R}^m$ given by

$$\pi^{-1}(\tilde{v}) = \left(\frac{v_1}{v_{m+1}}, \ldots, \frac{v_n}{v_{n+1}} \right) \in \mathbb{R}^n.$$

Note that there is an abuse of notation since π^{-1} is not formally an inverse of π.[19]

Consider the projective space $\boldsymbol{P}(\mathbb{R}^{n+1})$ and the subset defined by

$$P = \{[(v_1, ..., v_n, 1)] \in \boldsymbol{P}(\mathbb{R}^{n+1}) : (v_1, ..., v_n) \in C\}.$$

To express this in words, define the set $(C, 1) := \{(v_1, ..., v_n, 1) \in \mathbb{R}^{n+1} : (v_1, ..., v_n) \in C\}$, which can be thought of as the embedding of $C \subset \mathbb{R}^n$ in the hyperplane $H \subset \mathbb{R}^{n+1}$ defined by vectors of the form $(v_1, ..., v_n, 1)$. Then, P corresponds to all rays in $\mathbb{R}^{n+1}$ passing through $(C, 1) \subseteq \mathbb{R}^{n+1}$. Endow P with the family of subspaces that are of the form $S \cap P$ where S is a subspace

[19] The homogeneous coordinates of the form $(v_1 : \cdots : v_m : 0)$ do not correspond to points in Cartesian space. Rather, they correspond to *directions* in $\mathbb{R}^{n+1}$ and are interpreted geometrically as points "at infinity" in projective space. Because such points are included in projective space, any two distinct lines in $\boldsymbol{P}(\mathbb{R}^{n+1})$ intersect (parallel lines meet at a point at infinity).

of $\boldsymbol{P}(\mathbb{R}^m)$. Then a *line* in P corresponds to all the rays in the intersection of a plane in $\mathbb{R}^{n+1}$ with the set of rays that pass through $(C, 1)$.

Define a map $\varphi : P \to \boldsymbol{P}(\mathbb{R}^{n+1})$ by

$$\varphi([\tilde{v}]) = [\pi\left(f(\pi^{-1}(\tilde{v}))\right)], \qquad [\tilde{v}] \in P.$$

To verify that this is well-defined, take any $[(v_1, ..., v_n, 1)] \in P$. By definition, $(v_1, ..., v_n) \in C$. Given the definition of π^{-1}, we see that each $\tilde{v} \in [(v_1, ..., v_n, 1)]$ maps uniquely to $\pi^{-1}(\tilde{v}) = (v_1, ..., v_n) \in C$. Therefore $f(\pi^{-1}(\tilde{v})) \in \mathbb{R}^n$ does not depend on the choice of $\tilde{v} \in [\tilde{v}]$. Map $[\tilde{v}]$ into the ray $[\pi\left(f(\pi^{-1}(\tilde{v}))\right)]$ in $\mathbb{R}^{n+1}$. This defines the point $\varphi([\tilde{v}])$ in $\boldsymbol{P}(\mathbb{R}^{n+1})$. Note that the image of φ is $\varphi(P) = \{[(v_1, ..., v_n, 1)] \in \boldsymbol{P}(\mathbb{R}^{n+1}) : (v_1, ..., v_n) \in f(C)\}$.

Lemma 1. *Suppose $f : C \to \Delta$ is a local morphism. Then the following hold:*
(i) P is a locally projective subspace of $\boldsymbol{P}(\mathbb{R}^{n+1})$.
(ii) The image $\varphi(P)$ is not contained in a plane in P.
(iii) $\varphi : P \to \boldsymbol{P}(\mathbb{R}^{n+1})$ is a morphism.

Proof. (i) We need only show that a nonempty line in P has at least 2 points in it. By definition, a line in P is the intersection $L \cap P$ between P and some line L in $\boldsymbol{P}(\mathbb{R}^{n+1})$. This intersection corresponds to the rays in $\mathbb{R}^{n+1}$ passing through the intersection of $(C, 1)$ and some plane in $\mathbb{R}^{n+1}$. Since C is open, this intersection contains infinitely many points. Therefore there are infinitely many rays in the intersection, and in turn infinitely many points in the line $L \cap P$.

(ii) As noted earlier, $\varphi(P) = \{[(v_1, ..., v_n, 1)] \in \boldsymbol{P}(\mathbb{R}^{n+1}) : (v_1, ..., v_n) \in f(C)\}$. The union of these rays is a set S in $\mathbb{R}^{n+1}$ with dimensionality that equals the dimensionality of $f(C)$ plus 1. By hypothesis, $f(C)$ has dimensionality at least 3, and therefore S has dimensionality at least 4. It follows that $\varphi(P)$ does not lie in a plane in $\boldsymbol{P}(\mathbb{R}^{n+1})$.

(iii) Take any $[\tilde{u}], [\tilde{v}], [\tilde{w}] \in P$ that are collinear. Since they are in P, we can pick $\tilde{u}, \tilde{v}, \tilde{w}$ so that $\tilde{u}_{n+1} = \tilde{v}_{n+1} = \tilde{w}_{n+1} = 1$. Since they are collinear, there exists λ, γ such that $\tilde{u} = \lambda\tilde{v} + \gamma\tilde{w}$. Since $\tilde{u}_{n+1} = \tilde{v}_{n+1} = \tilde{w}_{n+1} = 1$ it must be that $\gamma = 1 - \lambda$. Take $z := \pi^{-1}(\tilde{z}) \in \mathbb{R}^n$ for $\tilde{z} = \tilde{u}, \tilde{v}, \tilde{w}$. It follows that $u = \lambda v + (1 - \lambda)w$. Then u, v, w are collinear. Since f is a morphism, we obtain that the vectors $f(u), f(v), f(w)$ are collinear, that is, there exists θ such that $f(u) = \theta f(v) + (1 - \theta)f(w)$. The projection $\tilde{z} = \pi(z)$ for $z = u, v, w$ then yields

$$\pi\left(f(u)\right) = \pi\left(\theta f(v) + (1 - \theta)f(w)\right) = \theta\pi(f(v)) + (1 - \theta)\pi(f(w)),$$

that is, $\pi(f(u)), \pi(f(v)), \pi(f(w))$ are collinear in $\mathbb{R}^{n+1}$, and in particular coplanar. Thus the rays $[\pi(f(u))], [\pi(f(v))], [\pi(f(w))]$ respectively correspond to points $\varphi([\tilde{u}]), \varphi([\tilde{v}]), \varphi([\tilde{w}])$ that are collinear in $\boldsymbol{P}(\mathbb{R}^{n+1})$. This estabishes that φ is a morphism. $\square$

C.2 Representation

For any matrix $M \in \mathbb{R}^{n \times n}$, let M_i denote the i^{th} row.

Theorem 8. *Suppose $n \geq 4$. If $f : C \to \mathbb{R}^n$ is a local morphism then there exist $c \in \mathbb{R},\ a, b \in \mathbb{R}^n,\ M \in \mathbb{R}^{n\times n}$ such that*

$$f(v) = \frac{M \cdot v + a}{b^\top \cdot v + c}, \qquad v \in C.$$

that is, $f(v) = \left(\frac{M_1 \cdot v + a_1}{b^\top \cdot v + c}, ..., \frac{M_n \cdot v + a_n}{b^\top \cdot v + c}\right)$ for any $v \in C$.

Proof. By Lemma 1, if $n \geq 4$ and f is a local morphism then P, φ satisfy the conditions of the Fundamental Theorem for Locally Projective Spaces (Theorem 7). Therefore there exists a $(n+1) \times (n+1)$ matrix $\tilde{M}$ such that $\phi([\tilde{v}]) = [\tilde{M} \cdot \tilde{v}]$ for all $[\tilde{v}] \in \boldsymbol{P}(\mathbb{R}^{n+1})$. In particular, for all $[\tilde{v}] \in P \subset \boldsymbol{P}(\mathbb{R}^{n+1})$ we have

$$[\pi \left(f(\pi^{-1}(\tilde{v}))\right)] = \varphi([\tilde{v}]) = [\tilde{M} \cdot \tilde{v}].$$

Write $\tilde{M} = \begin{pmatrix} M & a \\ b^\top & c \end{pmatrix}$, where $c \in \mathbb{R},\ a, b \in \mathbb{R}^n,\ M \in \mathbb{R}^{n\times n}$.

Take any $v = (v_1, .., v_n) \in C$. Write $\tilde{v} = (v_1, .., v_n, 1)$. Since $[\tilde{v}] \in P$, we have $[\pi \left(f(\pi^{-1}(\tilde{v}))\right)] = [\tilde{M} \cdot \tilde{v}]$, and so there exists a scalar $\lambda \neq 0$ such that

$$\pi(f(\pi^{-1}(\tilde{v})) = \lambda \tilde{M} \cdot \tilde{v} = \lambda(M_1 \cdot v + a_1, \cdots, M_n \cdot v + a_n,\ b^\top \cdot v + c).$$

However, the last coordinate must be 1 since

$$\pi(f(\pi^{-1}(\tilde{v})) = (f(v)_1, ..., f(v)_n, 1).$$

It follows that $\lambda = \frac{1}{b^\top \cdot v + c}$, and in particular, $f(v) = \left(\frac{M_1 \cdot v + a_1}{b^\top \cdot v + c}, ..., \frac{M_n \cdot v + a_n}{b^\top \cdot v + c}\right)$, as was to be shown. □

D Proofs for Main Results

Let $n := |\Omega| \geq 4$. Recall that an experiment σ is a mapping $\Omega \to \Delta S$. For each $s \in S$, let

$$\sigma_s := \left(\sigma(s \mid \omega)\right)_{\omega\in\Omega} \in \mathbb{R}^\Omega$$

denote the likelihood vector of signal s under experiment σ. We also write $\sigma_{s\omega}$ as shorthand for $\sigma(s \mid \omega)$.

The following result is a preference-based analog of Theorem 4, and is the main stepping stone for Theorems 1 and 2. It requires only minimal additional assumptions beyond Axiom 1: Axiom 5 is a technical richness condition, and Axiom 8 below is a weakening of Axiom 2 which states that $\succsim_{\sigma,s}$ depends on σ only through the likelihood vector σ_s.

Axiom 8 (Likelihood Consistency, LC)**.** *If $\sigma_s = \sigma'_s$, then $\succsim_{\sigma,s} = \succsim_{\sigma',s}$.*

Theorem 9. $\{\succsim_0, \succsim_{\sigma,s}\}_{(\sigma,s)\in(\Sigma\times S)}$ *satisfies Axioms 1, 5, and 8 if and only if there exist a quasi-experiment* ρ, $\tau \in (0,1]$, *and a family of* $|\Omega| \times |\Omega|$ *matrices* $(M_s)_{s\in S}$ *with weakly positive entries such that*

$$p_\sigma(\omega \mid s) = \frac{[\tau M_s \cdot \sigma_s + (1-\tau)\rho_s]_\omega \cdot p(\omega)}{\sum_{\omega'} [\tau M_s \cdot \sigma_s + (1-\tau)\rho_s]_{\omega'} \cdot p(\omega')}, \tag{14}$$

where $\mathrm{rank}[M_s \ \rho_s] = |\Omega|$ *for each* $s \in S$.

D.1 Proof for Theorem 9

Fix the family of binary relations over $\mathcal{F}$:

$$\{\succsim_0, \succsim_{\sigma,s}\}_{(\sigma,s)\in\Sigma\times S}$$

that satisfies Axiom 0. By Axiom 0, for some non-constant and affine $u : X \to \mathbb{R}$, $\succsim_0$ admits a representation $U(f) = \int_\Omega u \circ f \, dp$, and for each $(\sigma, s) \in \Sigma \times S$, $\succsim_{\sigma,s}$ admits a representation $U_{\sigma,s}(f) = \int_\Omega u \circ f \, dp_\sigma(\cdot \mid s)$. In later proofs, we use the notation $\succsim^E_{\sigma,s}$ for the conditional preference on an event $E \subseteq \Omega$, defined as in Ghirardato (2002): for $f, g \in \mathcal{F}$, write $f \succsim^E_{\sigma,s} g$ if there exists $h \in \mathcal{F}$ such that

$$fEh \succsim_{\sigma,s} gEh,$$

where fEh denotes the act that agrees with f on E and with h on $\Omega \setminus E$, and gEh is defined analogously.

By Axiom 8 (Likelihood consistency), if $\sigma_s = \sigma'_s$, then $p_\sigma(\cdot \mid s) = p_{\sigma'}(\cdot \mid s)$. Therefore, for each $\sigma_s \in \Sigma_s$ and $s \in S$, there exists a distribution $p_{\sigma_s}(\cdot \mid s) \in \Delta(\Omega)$ such that $\succsim_{\sigma,s}$ is represented by

$$U_{\sigma,s}(f) = \int_\Omega u \circ f \, dp_{\sigma_s}(\cdot \mid s). \tag{15}$$

Without loss of generality, assume 0 lies in the interior of $u(X)$. Fix $s \in S$. Define the mapping $\phi_s : \Sigma_s \to \Delta(\Omega) \subset \mathbb{R}^n$ such that for all $\sigma \in \Sigma$,

$$\phi_s(\sigma_s) := p_{\sigma_s}(\cdot \mid s).$$

Next, we claim that ϕ_s preserves collinearity. Let $\mu := \phi_s(\sigma_s)$, $\mu' := \phi_s(\sigma'_s)$, and $\mu'' := \phi_s(\sigma''_s)$, where $\sigma'' = \alpha\sigma + (1-\alpha)\sigma'$ for some $\alpha \in (0,1)$. It suffices to show that $\mu'' \in \mathrm{co}\{\mu, \mu'\}$.

For each $f \in \mathcal{F}$, let $u_f \in \mathbb{R}^n$ denote the utility act. Since $\succsim_{\sigma,s}$ is represented by μ, we have $f \succsim_{\sigma,s} g \iff \mu \cdot (u_f - u_g) \geq 0$, and similarly $f \succsim_{\sigma',s} g \iff \mu' \cdot (u_f - u_g) \geq 0$ and $f \succsim_{\sigma'',s} g \iff \mu'' \cdot (u_f - u_g) \geq 0$.

Thus Axiom 1 implies that for all $f, g \in \mathcal{F}$, $\mu \cdot (u_f - u_g) \geq 0$ and $\mu' \cdot (u_f - u_g) \geq 0$ together imply $\mu'' \cdot (u_f - u_g) \geq 0$. Because u is non-constant and affine and X is convex, $u(X)$ contains a non-degenerate interval. Hence, for every $v \in \mathbb{R}^n$, there exist acts $f, g \in \mathcal{F}$ and $\lambda > 0$ such that

$$u_f - u_g = \lambda v.$$

Therefore the implication above extends from utility-difference vectors to all $v \in \mathbb{R}^n$. That is, for every $v \in \mathbb{R}^n$, if $\mu \cdot v \geq 0$ and $\mu' \cdot v \geq 0$, then $\mu'' \cdot v \geq 0$.

Now suppose $\mu'' \notin \text{co}\{\mu, \mu'\}$. Since $\text{co}\{\mu, \mu'\}$ is closed and convex, the separating hyperplane theorem yields $v \in \mathbb{R}^n$ and $c \in \mathbb{R}$ such that

$$\mu \cdot v \geq c, \qquad \mu' \cdot v \geq c, \qquad \mu'' \cdot v < c.$$

Let $\tilde{v} := v - c1_\Omega$. Since $\mu, \mu', \mu'' \in \Delta\Omega$, we have

$$\mu \cdot \tilde{v} \geq 0, \qquad \mu' \cdot \tilde{v} \geq 0, \qquad \mu'' \cdot \tilde{v} < 0,$$

contradicting the preceding implication. Hence $\mu'' \in \text{co}\{\mu, \mu'\}$. Therefore ϕ_s preserves collinearity.

Moreover, I claim that the image of ϕ_s is not contained in a plane in $\mathbb{R}^n$. Suppose towards a contradiction that this is the case. Since $n := |\Omega| \geq 4$, the affine hull of the probability simplex $\Delta(\Omega) := \{v \in \mathbb{R}^n_+ \mid \sum_{m=1}^n v_m = 1\}$ is at least a 3-dimensional subset of $\mathbb{R}^n$. The assumption that ϕ_s is contained in a plane in $\mathbb{R}^n$, which is at least one-dimensional lower than the probability simplex, would imply that there exists $f^* \in \mathcal{F}$ and $c \in \mathbb{R}$ such that $u_{f^*} \notin \text{Span}\{1_\Omega\}$ and

$$p_{\sigma_s}(\cdot \mid s) \cdot u_{f^*} = c \qquad \text{for all } \sigma \in \Sigma.$$

Hence, for all $\sigma, \sigma' \in \Sigma$ and all $x \in X$,

$$f^* \succ_{\sigma,s} x \iff f^* \succ_{\sigma',s} x.$$

Since $u_{f^*} \notin \text{Span}\{1_\Omega\}$, there exist $\omega, \psi \in \Omega$ such that $f^*(\omega) \not\sim_0 f^*(\psi)$, contradicting Axiom 5 (Richness).

By Theorem 8, since ϕ_s preserves collinearity and $\phi_s(\Sigma_s)$ is not contained in a two-dimensional affine subspace of $\mathbb{R}^n$, there exists a nonzero $(n+1) \times (n+1)$ matrix $T_s = \begin{pmatrix} A_s & b_s \\ c_s & d_s \end{pmatrix}$, where A_s is an $n \times n$ matrix, b_s is an n-dimensional column vector, c_s is an n-dimensional row vector, and $d_s \in \mathbb{R}$ are such that for all $\zeta \in \Sigma_s$,

$$\phi_s(\zeta) = \frac{A_s\zeta + b_s}{c_s\zeta + d_s}.$$

Moreover, since $\phi_s(\zeta) \in \Delta(\Omega)$ for all $\zeta \in \Sigma_s$, we have $\left\| \frac{A_s\zeta + b_s}{c_s\zeta + d_s} \right\|_1 = 1$ where $||v||_1 = \sum v_i$ for any vector $v \in \mathbb{R}^n$. Therefore,

$$|c_s\zeta + d_s| = \|A_s\zeta + b_s\|_1 \qquad \text{for all } \zeta \in \Sigma_s.$$

Given the intermediate value theorem, since $c_s\zeta + d_s$ is a continuous function that never vanishes on the convex (and hence path-connected) set Σ_s, it must have a constant sign on Σ_s. Multiplying A_s, b_s, c_s, d_s by -1 if necessary, we may assume that $c_s\zeta + d_s = \|A_s\zeta + b_s\|_1$ for all $\zeta \in \Sigma_s$Therefore, for all $\zeta \in \Sigma_s$,

$$\phi_s(\zeta) = \frac{A_s\zeta + b_s}{\|A_s\zeta + b_s\|_1}.$$

Moreover, since 0 lies in the closure of Σ_s, take any sequence $\zeta_k \in \Sigma_s$ with $\zeta_k \to 0$. Since $\phi_s(\zeta_k) = \frac{A_s\zeta_k + b_s}{\|A_s\zeta_k + b_s\|_1} \in \Delta(\Omega)$, we have $A_s\zeta_k + b_s \geq 0$ for every k. Taking limits yields $b_s \geq 0$.

We next rewrite the representation in the form stated in Theorem 9. First, notice that since $p \in \mathrm{Int}(\Delta(\Omega))$, each component of p is strictly positive. Thus the diagonal matrix $D_p := \mathrm{diag}(p)$ is invertible. If $b_s = 0$ for all $s \in S$, then take $M_s := D_p^{-1}A_s$ for all $s \in S$, and

$$\phi_s(\sigma_s) = \frac{p \odot [M_s\sigma_s]}{\|p \odot [M_s\sigma_s]\|_1}$$

for all $\sigma \in \Sigma$ and $s \in S$, yielding the representation in Theorem 9, specifically the special case where the agent assigns weight zero to the quasi-experiment ρ. Proceed to establish the representation for the case where $b_s \neq 0$ for some $s \in S$.

Step 1. Constructing τ, ρ, and M_s Choose $\alpha > 0$ such that $\alpha \leq \min_{\omega:\sum_s b_{s\omega} \neq 0} \frac{p_\omega}{\sum_s b_{s\omega}}$ and define

$$\rho_s := \alpha D_p^{-1} b_s \quad \text{and} \quad M_s := D_p^{-1}A_s \qquad \text{for all } s \in S.$$

Let

$$\tau := \frac{\alpha}{1+\alpha} \in (0,1).$$

Then $1-\tau = \frac{1}{1+\alpha}$ and $(1-\tau)\rho_s = \frac{\alpha}{1+\alpha} D_p^{-1} b_s = \tau D_p^{-1} b_s$. Therefore,

$$\tau M_s\sigma_s + (1-\tau)\rho_s = \tau D_p^{-1}A_s\sigma_s + \tau D_p^{-1} b_s = \tau D_p^{-1}(A_s\sigma_s + b_s).$$

Multiplying componentwise by p, we obtain

$$p \odot [\tau M_s\sigma_s + (1-\tau)\rho_s] = \tau(A_s\sigma_s + b_s).$$

Since normalization is invariant under multiplication by a positive scalar,

$$\frac{p \odot [\tau M_s\sigma_s + (1-\tau)\rho_s]}{\|p \odot [\tau M_s\sigma_s + (1-\tau)\rho_s]\|_1} = \frac{A_s\sigma_s + b_s}{\|A_s\sigma_s + b_s\|_1}.$$

Since $A_s\sigma_s + b_s \geq 0$, the l_1 norm equals the coordinate sum, matching the denominator in Theorem 9.

Step 2. Verifying that ρ is a quasi-experiment Recall that $\rho_s = \alpha D_p^{-1} b_s$ for all $s \in S$. Since $b_s \geq 0$ for every $s \in S$ and $p_\omega > 0$ for every $\omega \in \Omega$, we have

$$\rho_{s\omega} = \alpha \frac{b_{s\omega}}{p_\omega} \geq 0 \qquad \text{for all } s \in S,\ \omega \in \Omega.$$

It remains to verify the column-sum condition. For each $\omega \in \Omega$, we have $\sum_{s\in S} \rho_{s\omega} = \alpha \sum_{s\in S} \frac{b_{s\omega}}{p_\omega} = \alpha \frac{\sum_{s\in S} b_{s\omega}}{p_\omega}$, and by our choice of α, we have $\alpha \leq$

$\min_{\omega:\sum_s b_{s\omega}\neq 0} \frac{p_\omega}{\sum_s b_{s\omega}}$. Hence, if $\sum_s b_{s\omega} > 0$ then $\sum_{s\in S}\rho_{s\omega} = \alpha\frac{\sum_s b_{s\omega}}{p_\omega} \leq 1$, and if $\sum_s b_{s\omega} = 0$ then $b_{s\omega} = 0$ for every s so that $\sum_{s\in S}\rho_{s\omega} = 0 \leq 1$. Therefore

$$\rho_{s\omega} \geq 0 \quad \text{and} \quad \sum_{s\in S}\rho_{s\omega} \leq 1 \qquad \text{for all } s \in S,\ \omega \in \Omega.$$

Finally, since $b_s \neq 0$ for some $s \in S$, it follows that $\rho \neq 0$. This shows that ρ is a quasi-experiment.

Step 3. Show the Rank Condition Finally, it remains to show that,

$$\operatorname{rank}[\,M_s\ \rho_s\,] = n := |\Omega| \qquad \text{for all } s \in S,$$

where $[\,M_s\ \rho_s\,]$ denotes the $n\times(n+1)$ matrix obtained by appending the column vector ρ_s to M_s.

Suppose towards a contradiction that $\operatorname{rank}[\,M_{\hat{s}}\ \rho_{\hat{s}}\,] < n$ for some $\hat{s} \in S$. Since the matrix has n rows, its rows are linearly dependent, so there exists a nonzero vector $w \in \mathbb{R}^n$ in its left null space, that is,

$$w^\top M_{\hat{s}} = 0 \qquad \text{and} \qquad w^\top \rho_{\hat{s}} = 0.$$

Consequently, for every $\zeta \in \Sigma_{\hat{s}}$,

$$w^\top\big[\tau M_{\hat{s}}\zeta + (1-\tau)\rho_{\hat{s}}\big] = \tau\,(w^\top M_{\hat{s}})\zeta + (1-\tau)\,w^\top\rho_{\hat{s}} = 0.$$

Let $\hat{w} := D_p^{-1}w$, which is nonzero since D_p is invertible. Then, for every $\zeta \in \Sigma_{\hat{s}}$,

$$\hat{w}^\top\Big(p\odot\big[\tau M_{\hat{s}}\zeta+(1-\tau)\rho_{\hat{s}}\big]\Big) = \hat{w}^\top D_p\big[\tau M_{\hat{s}}\zeta+(1-\tau)\rho_{\hat{s}}\big] = w^\top\big[\tau M_{\hat{s}}\zeta+(1-\tau)\rho_{\hat{s}}\big] = 0.$$

Since $\phi_{\hat{s}}(\zeta)$ is a positive multiple of $p\odot[\tau M_{\hat{s}}\zeta + (1-\tau)\rho_{\hat{s}}]$, it follows that

$$\hat{w}^\top\phi_{\hat{s}}(\zeta) = 0 \qquad \text{for all } \zeta \in \Sigma_{\hat{s}}.$$

We may take $\hat{w} \notin \operatorname{span}\{1_\Omega\}$; otherwise, if $\hat{w} \in \operatorname{span}\{1_\Omega\}$, then the fact that $\hat{w}^\top\phi_{\hat{s}}(\zeta) = 0$ and $\phi_{\hat{s}}(\zeta) \in \Delta\Omega$ for all $\zeta \in \Sigma_{\hat{s}}$ implies that $\hat{w} = 0$, a contradiction.

As a result, there exists $\hat{f} \in \mathcal{F}$ with $u_{\hat{f}} = \hat{w}$, so that $u_{\hat{f}} \notin \operatorname{span}\{1_\Omega\}$ and

$$\phi_{\hat{s}}(\sigma_{\hat{s}}) \cdot u_{\hat{f}} = p_{\sigma_{\hat{s}}}(\cdot \mid \hat{s}) \cdot u_{\hat{f}} = 0 \qquad \text{for all } \sigma \in \Sigma.$$

It follows that, $\hat{f} \succsim_{\sigma,\hat{s}} x \iff \hat{f} \succsim_{\sigma',\hat{s}} x$ for all $x \in X$ and all $\sigma, \sigma' \in \Sigma$. Since $u_{\hat{f}} \notin \operatorname{span}\{1_\Omega\}$, there exist $\omega, \psi \in \Omega$ such that $\hat{f}(\omega) \not\sim_0 \hat{f}(\psi)$, contradicting Axiom 5 (Richness). Therefore $\operatorname{rank}[\,M_{\hat{s}}\ \rho_{\hat{s}}\,] = n$ for every $s \in S$, completing the proof.

D.2 Proof for Theorem 2

Lemma 2. *For $\sigma, \sigma' \in \Sigma$, $\omega, \omega' \in \Omega$, and $s \in S$, suppose $\sigma(s \mid \omega) = \sigma'(s \mid \omega)$ and $\sigma(s \mid \omega') = \sigma'(s \mid \omega')$. Then*

$$p_{\sigma_s}(\omega' \mid s) > 0 \quad \textit{and} \quad p_{\sigma'_s}(\omega' \mid s) > 0 \implies \frac{p_{\sigma_s}(\omega \mid s)}{p_{\sigma_s}(\omega' \mid s)} = \frac{p_{\sigma'_s}(\omega \mid s)}{p_{\sigma'_s}(\omega' \mid s)}.$$

Proof. Let $E = \{\omega, \omega'\}$. Since $\sigma(s \mid \psi) = \sigma'(s \mid \psi)$ for all $\psi \in E$, Axiom 2 implies that the conditional preferences $\succsim^E_{\sigma,s}$ and $\succsim^E_{\sigma',s}$ coincide, which are represented by $f \mapsto \sum_{\psi \in E} p_{\sigma_s}(\psi \mid s)\, u(f(\psi))$ and $f \mapsto \sum_{\psi \in E} p_{\sigma'_s}(\psi \mid s)\, u(f(\psi))$, respectively. Since both representations use the same non-constant affine utility index u, the induced likelihood ratios on E must coincide. Therefore, $\frac{p_{\sigma_s}(\omega|s)}{p_{\sigma_s}(\omega'|s)} = \frac{p_{\sigma'_s}(\omega|s)}{p_{\sigma'_s}(\omega'|s)}$, as desired. □

Lemma 3. *Suppose that, for each $s \in S$,*

$$p_{\sigma_s}(\cdot \mid s) = \frac{p \odot [\tau M_s \sigma_s + (1-\tau)\rho_s]}{\|p \odot [\tau M_s \sigma_s + (1-\tau)\rho_s]\|_1} \qquad \textit{for all } \sigma \in \Sigma,$$

where $p \in \mathrm{Int}(\Delta(\Omega))$, $\tau \in (0,1]$, and $\mathrm{rank}[\, M_s \ \rho_s \,] = n$. Then, for every $s \in S$, the matrix M_s is diagonal.

Proof. Fix $s \in S$. Recall that

$$\sigma_s = (\sigma_{s\psi})_{\psi \in \Omega} \in \Sigma_s \subset \mathbb{R}^n$$

is viewed as a vector indexed by states. Fix two distinct states $\omega, \omega' \in \Omega$. We first show that, for every $\kappa \notin \{\omega, \omega'\}$,

$$(M_s)_{\omega\kappa} = (M_s)_{\omega'\kappa} = 0.$$

By Lemma 2, if two experiments $\sigma, \sigma' \in \Sigma$ satisfy $\sigma_{s\omega} = \sigma'_{s\omega}$ and $\sigma_{s\omega'} = \sigma'_{s\omega'}$, then, whenever the denominator is positive, it must be that $\frac{p_{\sigma_s}(\omega|s)}{p_{\sigma_s}(\omega'|s)} = \frac{p_{\sigma'_s}(\omega|s)}{p_{\sigma'_s}(\omega'|s)}$. Hence the ratio

$$R_{\omega,\omega'}(\sigma_s) := \frac{p_{\sigma_s}(\omega \mid s)}{p_{\sigma_s}(\omega' \mid s)}$$

depends only on the two coordinates $\sigma_{s\omega}$ and $\sigma_{s\omega'}$.

Using the representation above,

$$R_{\omega,\omega'}(\sigma_s) = \frac{p_\omega \left[\tau \sum_{\psi \in \Omega} (M_s)_{\omega\psi} \sigma_{s\psi} + (1-\tau)\rho_{s\omega}\right]}{p_{\omega'} \left[\tau \sum_{\psi \in \Omega} (M_s)_{\omega'\psi} \sigma_{s\psi} + (1-\tau)\rho_{s\omega'}\right]}.$$

The denominator is positive on a nonempty open subset of Σ_s: otherwise $p_{\sigma_s}(\omega' \mid s) = 0$ on a nonempty open set, so $\phi_s(\Sigma_s)$ would be locally contained in the hyperplane $\{x \in \Delta\Omega : x_{\omega'} = 0\}$, contradicting the full-rank condition established above.

Now fix $\kappa \notin \{\omega, \omega'\}$. Since Σ_s has nonempty interior, we may vary $\sigma_{s\kappa}$ on a nonempty open set while holding the other coordinates fixed. On the intersection with the open set where the denominator is positive, $R_{\omega,\omega'}$ is well-defined and, by the preceding paragraph, independent of $\sigma_{s\kappa}$. Hence

$$\frac{\partial R_{\omega,\omega'}(\sigma_s)}{\partial \sigma_{s\kappa}} = 0$$

on a full-dimensional open set.

Since p_ω and $p_{\omega'}$ are positive constants, the quotient rule yields

$$0 = \frac{\tau(M_s)_{\omega\kappa}\left[\tau\sum_{\psi\in\Omega}(M_s)_{\omega'\psi}\sigma_{s\psi} + (1-\tau)\rho_{s\omega'}\right] - \tau(M_s)_{\omega'\kappa}\left[\tau\sum_{\psi\in\Omega}(M_s)_{\omega\psi}\sigma_{s\psi} + (1-\tau)\rho_{s\omega}\right]}{\left[\tau\sum_{\psi\in\Omega}(M_s)_{\omega'\psi}\sigma_{s\psi} + (1-\tau)\rho_{s\omega'}\right]^2}.$$

Since $\tau > 0$, it follows that

$$(M_s)_{\omega\kappa}\left[\tau\sum_{\psi\in\Omega}(M_s)_{\omega'\psi}\sigma_{s\psi} + (1-\tau)\rho_{s\omega'}\right] = (M_s)_{\omega'\kappa}\left[\tau\sum_{\psi\in\Omega}(M_s)_{\omega\psi}\sigma_{s\psi} + (1-\tau)\rho_{s\omega}\right].$$

This identity holds on a full-dimensional open set. Since both sides are affine functions of $\sigma_s \in \mathbb{R}^\Omega$, their coefficients must coincide. Hence, for every $\psi \in \Omega$,

$$(M_s)_{\omega\kappa}(M_s)_{\omega'\psi} = (M_s)_{\omega'\kappa}(M_s)_{\omega\psi},$$

and

$$(M_s)_{\omega\kappa}\rho_{s\omega'} = (M_s)_{\omega'\kappa}\rho_{s\omega}.$$

Therefore, if either $(M_s)_{\omega\kappa} \neq 0$ or $(M_s)_{\omega'\kappa} \neq 0$, then the ω-row and the ω'-row of $[\,M_s\ \rho_s\,]$ are proportional, contradicting

$$\operatorname{rank}[\,M_s\ \rho_s\,] = n.$$

Hence

$$(M_s)_{\omega\kappa} = (M_s)_{\omega'\kappa} = 0 \qquad \text{for all } \kappa \notin \{\omega, \omega'\}.$$

Now fix $\omega \in \Omega$. Since $n \geq 4$, for every $\kappa \neq \omega$, we may choose some $\omega' \notin \{\omega, \kappa\}$. Applying the preceding conclusion to the pair (ω, ω') gives

$$(M_s)_{\omega\kappa} = 0.$$

Therefore,

$$(M_s)_{\omega\kappa} = 0 \qquad \text{for all } \kappa \neq \omega.$$

Since ω was arbitrary, M_s is diagonal. $\square$

Lemma 4. *Suppose that, for each $s \in S$,*

$$p_{\sigma_s}(\cdot \mid s) = \frac{p \odot [\tau M_s\sigma_s + (1-\tau)\rho_s]}{\|p \odot [\tau M_s\sigma_s + (1-\tau)\rho_s]\|_1} \qquad \textit{for all } \sigma \in \Sigma,$$

where $p \in \operatorname{Int}(\Delta(\Omega))$, $\tau \in (0,1]$, and M_s is diagonal. Suppose also that Axiom 4 holds. Then, for each $s \in S$, there exists $k_s > 0$ such that

$$M_s = k_s I.$$

Proof. Fix $s \in S$. Since M_s is diagonal, write

$$M_s = \mathrm{Diag}(m_s),$$

where $m_s \in \mathbb{R}^n \setminus \{0\}$.[20] Suppose, towards a contradiction, that $p \odot m_s$ is not proportional to p. We first focus on the scenario where $\sum_{\omega\in\Omega} p_\omega \cdot m_{s\omega} \neq 0$. Then at the end of the proof, we consider the scenario where $\sum_{\omega\in\Omega} p_\omega \cdot m_{s\omega} = 0$.

If $\sum_{\omega\in\Omega} p_\omega \cdot m_{s\omega} \neq 0$, then take

$$r_s := \frac{p \odot m_s}{\sum_{\omega\in\Omega} p_\omega \cdot m_{s\omega}} \neq p.$$

Since the image $\phi_s(\Sigma_s)$ is not contained in a one-dimensional affine subspace of $\Delta(\Omega)$, we may choose $\sigma \in \Sigma$ such that $\sigma_s \in \mathrm{Int}(\Sigma_s)$ and the three points

$$p_{\sigma_s}(\cdot \mid s), \qquad p, \qquad r_s$$

are not collinear.

Choose $\alpha > 0$ such that

$$\sigma_s' := \sigma_s + \alpha 1_\Omega \in \Sigma_s,$$

and let $\sigma' \in \Sigma$ satisfy this equality.

For any two points $x, y \in \mathbb{R}^n$, let $l(x,y) := \{\lambda x + (1-\lambda)y \mid \lambda \in \mathbb{R}\}$ denote the collection of all points collinear with x and y.

Claim 1.

$$p_{\sigma_s'}(\cdot \mid s) \in l[q_{\sigma_s}(\cdot \mid s), r_s].$$

Proof of Claim 1. Since $M_s = \mathrm{Diag}(m_s)$, we have

$$M_s 1_\Omega = m_s.$$

Hence

$$\begin{aligned} p_{\sigma_s'}(\cdot \mid s) &= \frac{p \odot [\tau M_s(\sigma_s + \alpha 1_\Omega) + (1-\tau)\rho_s]}{\sum_{\omega\in\Omega} p_\omega \left[\tau(M_s(\sigma_s + \alpha 1_\Omega))_\omega + (1-\tau)\rho_{s\omega}\right]} \\ &= \frac{p \odot [\tau M_s \sigma_s + (1-\tau)\rho_s] + \tau\alpha(p \odot m_s)}{\sum_{\omega\in\Omega} \{p_\omega[\tau(M_s\sigma_s)_\omega + (1-\tau)\rho_{s\omega}] + \tau\alpha p_\omega m_{s\omega}\}}. \qquad (16) \end{aligned}$$

Therefore,

$$\begin{aligned} p_{\sigma_s'}(\cdot \mid s) = &\frac{\sum_{\omega\in\Omega} p_\omega[\tau(M_s\sigma_s)_\omega + (1-\tau)\rho_{s\omega}]}{\sum_{\omega\in\Omega} \{p_\omega[\tau(M_s\sigma_s)_\omega + (1-\tau)\rho_{s\omega}] + \tau\alpha p_\omega m_{s\omega}\}} p_{\sigma_s}(\cdot \mid s) \qquad (17) \\ &+ \frac{\tau\alpha \sum_{\omega\in\Omega} p_\omega m_{s\omega}}{\sum_{\omega\in\Omega} \{p_\omega[\tau(M_s\sigma_s)_\omega + (1-\tau)\rho_{s\omega}] + \tau\alpha p_\omega m_{s\omega}\}} r_s \qquad (18) \end{aligned}$$

[20] Indeed, if $m_s = 0$, then $M_s = 0$, so the matrix $[\, M_s \ \rho_s \,]$ has rank at most one, contradicting $\mathrm{rank}[\, M_s \ \rho_s \,] = n$.

By construction, the two coefficients sum to one. Hence

$$p_{\sigma'_s}(\cdot \mid s) \in l[q_{\sigma_s}(\cdot \mid s), r_s].$$

Claim 2.

$$p_{\sigma'_s}(\cdot \mid s) \in l[p, p_{\sigma_s}(\cdot \mid s)], \text{ and } p_{\sigma'_s}(\cdot \mid s) \neq q_{\sigma_s}(\cdot \mid s).$$

Proof of Claim 2. By Axiom 4, since $\sigma'_s = \sigma_s + \alpha 1_\Omega$, for all $f, g \in \mathcal{F}$,

$$f \succsim_0 g \quad \text{and} \quad f \succsim_{\sigma,s} g \implies f \succsim_{\sigma',s} g.$$

Since $\succsim_0$, $\succsim_{\sigma,s}$, and $\succsim_{\sigma',s}$ are represented by the beliefs p, $p_{\sigma_s}(\cdot \mid s)$, and $p_{\sigma'_s}(\cdot \mid s)$, respectively, the same separating-hyperplane argument used above implies

$$p_{\sigma'_s}(\cdot \mid s) \in \text{co}\{p, p_{\sigma_s}(\cdot \mid s)\}.$$

Moreover, $p_\sigma(\cdot \mid s) \neq p_{\sigma'}(\cdot \mid s)$. Indeed, since $p \neq p_{\sigma_s}(\cdot \mid s)$, there would exists $f, g \in \mathcal{F}$ such that

$$f \succ_0 g \text{ and } f \sim_{\sigma,s} g.$$

However, $p_\sigma(\cdot \mid s) = p_{\sigma'}(\cdot \mid s)$ would imply that $f \sim_{\sigma',s} g$, which contradicts Axiom 4.

By Claims 1 and 2,

$$p_{\sigma'_s}(\cdot \mid s) \in l[p_{\sigma_s}(\cdot \mid s), r_s]. \cap l[p, p_{\sigma_s}(\cdot \mid s)].$$

But the three points

$$p_{\sigma_s}(\cdot \mid s), \qquad p, \qquad r_s$$

are not collinear, so the two lines intersect only at $p_{\sigma_s}(\cdot \mid s)$. Hence

$$p_{\sigma'_s}(\cdot \mid s) = p_{\sigma_s}(\cdot \mid s),$$

contradicting Claim 2.

Therefore $p \odot m_s$ must be proportional to p. Since p has full support, this implies that m_s is proportional to 1_Ω. Thus there exists $k_s \in \mathbb{R}$ such that

$$m_s = k_s 1_\Omega.$$

Moreover, if $k_s = 0$, $M_s = 0$, so $[M_s \ \rho_s]$ has rank at most one, which contradicts $\text{rank}[M_s \ \rho_s] = n$. If $k_s < 0$, then $m_s < 0$ for all $s \in S$, thus, for σ_s sufficiently close to 0, the weight

$$\frac{\tau\alpha \sum_{\omega\in\Omega} p_\omega m_{s\omega}}{\sum_{\omega\in\Omega} \{p_\omega[\tau(M_s\sigma_s)_\omega + (1-\tau)\rho_{s\omega}] + \tau\alpha p_\omega m_{s\omega}\}}$$

in (17) must be negative. That is,

$$p_{\sigma'_s}(\cdot \mid s) = (1-\theta)p_{\sigma_s}(\cdot \mid s) + \theta p$$

for some $\theta \leq 0$. This is contradictory to that

$$p_{\sigma'_s}(\cdot \mid s) \in \text{co}\{p, p_{\sigma_s}(\cdot \mid s)\} \text{ and } p_{\sigma_s}(\cdot \mid s) \neq p_{\sigma'_s}(\cdot \mid s).$$

As a result, $k_s > 0$. In conclusion,

$$m_s = k_s 1_\Omega,$$

as desired. Finally, recall that the preceding argument assumed

$$\sum_{\omega \in \Omega} p_\omega m_{s\omega} \neq 0.$$

It remains to consider the case

$$\sum_{\omega \in \Omega} p_\omega m_{s\omega} = 0. \tag{19}$$

We show that this case is impossible. Suppose, towards a contradiction, that (19) holds. For any $\sigma_s \in \text{Int}(\Sigma_s)$ and any sufficiently small $\alpha > 0$, let

$$\sigma'_s := \sigma_s + \alpha 1_\Omega \in \Sigma_s.$$

By (16) and (19),

$$p_{\sigma'_s}(\cdot \mid s) = p_{\sigma_s}(\cdot \mid s) + \frac{\tau\alpha(p \odot m_s)}{\sum_{\omega \in \Omega} p_\omega[\tau(M_s\sigma_s)_\omega + (1-\tau)\rho_{s\omega}]}.$$

Since $m_s \neq 0$ and $p \in \text{Int}(\Delta\Omega)$, the direction of the displacement

$$p_{\sigma'_s}(\cdot \mid s) - p_{\sigma_s}(\cdot \mid s)$$

is the fixed direction $p \odot m_s$, independent of σ_s.

On the other hand, by Axiom 4 and the same separating-hyperplane argument used above,

$$p_{\sigma'_s}(\cdot \mid s) \in \text{co}\{p, p_{\sigma_s}(\cdot \mid s)\}.$$

Since $p_{\sigma'_s}(\cdot \mid s) \neq p_{\sigma_s}(\cdot \mid s)$, it follows that

$$p - p_{\sigma_s}(\cdot \mid s)$$

must be collinear with $p \odot m_s$. Because this holds for every $\sigma_s \in \Sigma_s$, we obtain

$$\phi_s(\Sigma_s) \subseteq p + \text{span}\{p \odot m_s\}.$$

Thus $\phi_s(\Sigma_s)$ is contained in a one-dimensional affine subspace of $\Delta(\Omega)$, contradicting the non-degeneracy of the image. Hence (19) cannot hold. □

By Lemma 4, for each $s \in S$, there exists $k_s > 0$ such that

$$M_s = k_s I.$$

Hence

$$p_{\sigma_s}(\cdot \mid s) = \frac{p \odot [\tau k_s \sigma_s + (1-\tau)\rho_s]}{\sum_{\omega \in \Omega} p_\omega [\tau k_s \sigma_{s\omega} + (1-\tau)\rho_{s\omega}]}.$$

Let

$$\underline{k} := \min_{s \in S} k_s > 0, \qquad \hat{\tau} := \frac{\tau \underline{k}}{1 - \tau + \tau \underline{k}} \in (0, 1],$$

and define, for each $s \in S$,

$$\hat{\rho}_s := \frac{\underline{k}}{k_s} \rho_s.$$

Since $0 < \underline{k}/k_s \leq 1$ for every s, $\hat{\rho}$ is again a quasi-experiment. Moreover,

$$\tau k_s \sigma_s + (1-\tau)\rho_s = \frac{k_s}{\underline{k}}(1 - \tau + \tau \underline{k})\big[\hat{\tau}\sigma_s + (1-\hat{\tau})\hat{\rho}_s\big].$$

Substituting this identity into the representation above, and using the fact that normalization is invariant under multiplication by a positive scalar, yields

$$p_{\sigma_s}(\cdot \mid s) = \frac{p \odot [\hat{\tau}\sigma_s + (1-\hat{\tau})\hat{\rho}_s]}{\sum_{\omega \in \Omega} p_\omega [\hat{\tau}\sigma_{s\omega} + (1-\hat{\tau})\hat{\rho}_{s\omega}]}.$$

Renaming $(\hat{\tau}, \hat{\rho})$ as (τ, ρ) gives the desired representation.

It remains to establish the uniqueness claim. First, fix $\tau \in (0, 1]$ and suppose two quasi-experiments ρ and ρ' both satisfy (3). Fix $s \in S$. Since 0 lies in the closure of Σ_s, letting $\sigma_s \to 0$ in (3) shows that the normalizations of $p \odot \rho_s$ and $p \odot \rho'_s$ coincide; as p has full support, $\rho'_s = c\rho_s$ for some $c > 0$. If $c \neq 1$, then for any $\sigma_s \in \Sigma_s$ not proportional to ρ_s, the vectors $\tau\sigma_s + (1-\tau)\rho_s$ and $\tau\sigma_s + (1-\tau)c\rho_s$ are not proportional and hence induce different posteriors, contradicting (3); such a σ_s exists since Σ_s is open. Hence $c = 1$ and $\rho = \rho'$: for each τ, the quasi-experiment in (3) is unique.

Next, suppose (τ, ρ) and (τ', ρ') both satisfy (3). Fix $s \in S$. Since the two representations induce the same posterior at every $\sigma_s \in \Sigma_s$ and p has full support, the vectors $\tau'\sigma_s + (1-\tau')\rho'_s$ and $\tau\sigma_s + (1-\tau)\rho_s$ are proportional at each $\sigma_s \in \Sigma_s$:

$$\tau'\sigma_s + (1-\tau')\rho'_s = d\,\big[\tau\sigma_s + (1-\tau)\rho_s\big].$$

The factor $d > 0$ cannot depend on σ_s: both sides are affine in σ_s, so comparing any two points of the open set Σ_s with their midpoint forces a common factor. Matching coefficients on Σ_s then yields $\tau' = d\tau$ and $(1-\tau')\rho'_s = d(1-\tau)\rho_s$; in particular $d = \tau'/\tau$ is the same for all s, and eliminating it gives

$$\frac{1-\tau'}{\tau'}\,\rho'_s = \frac{1-\tau}{\tau}\,\rho_s \quad \text{for all } s \in S. \tag{20}$$

Thus any two representations are related by (20): a higher trust is compensated by a proportional enlargement of the alternative theory. Fix one representation (τ, ρ) with $\tau < 1$. By (20), the candidate representations are exactly $\left(\tau', \frac{\tau'(1-\tau)}{\tau(1-\tau')}\rho\right)$ for $\tau' \in (0,1)$, and such ρ' is a quasi-experiment if and only if, for every $\omega \in \Omega$,

$$\sum_{s\in S} \rho'(s \mid \omega) = \frac{\tau'(1-\tau)}{\tau(1-\tau')} \sum_{s\in S} \rho(s \mid \omega) \leq 1,$$

that is, if and only if

$$\tau' \;\leq\; \tau^* := \Big(1 + \tfrac{1-\tau}{\tau} \max_{\omega\in\Omega} \sum_{s\in S} \rho(s \mid \omega)\Big)^{-1}.$$

The bound is attained: at $\tau' = \tau^*$ the column sum equals 1 at the maximizing ω. Hence the set of admissible trust parameters is $(0, \tau^*]$, so a unique maximal trust τ^* exists, and the associated $\bar{\rho}$ is unique by the first part of the argument. Finally, if some representation has $\tau = 1$, then (20) forces $\tau' = 1$ for every representation, and $\tau^* = 1$ with ρ playing no role in (3).

D.3 Proof for Theorem 3

Suppose Idempotence (given by (2)) is satisfied. Then, for some $\sigma^* \in \Sigma$,

$$p_{\sigma^*_s}(\cdot \mid s) = p \qquad \text{for all } s \in S.$$

By Theorem 2, there exist a quasi-experiment ρ and $\tau \in (0,1]$ such that

$$p_{\sigma_s}(\cdot \mid s) = \frac{p \odot [\tau\sigma_s + (1-\tau)\rho_s]}{\sum_{\omega\in\Omega} p_\omega[\tau\sigma_{s\omega} + (1-\tau)\rho_{s\omega}]} \qquad \text{for all } (\sigma, s) \in \Sigma \times S.$$

If $\tau = 1$, then ρ does not enter the posterior. Hence we may choose ρ to be any experiment, and the desired representation follows. It therefore suffices to consider the case $\tau \in (0,1)$.

Since $p_{\sigma^*_s}(\cdot \mid s) = p$, and $p \in \mathrm{Int}(\Delta(\Omega))$, for each $s \in S$ there exists $l_s > 0$ such that

$$\tau\sigma^*_s + (1-\tau)\rho_s = l_s 1_\Omega.$$

Equivalently, for all $s \in S$ and $\omega \in \Omega$,

$$\tau\sigma^*_{s\omega} + (1-\tau)\rho_{s\omega} = l_s.$$

Summing over $s \in S$, we obtain, for every $\omega \in \Omega$,

$$\tau \sum_{s\in S} \sigma^*_{s\omega} + (1-\tau) \sum_{s\in S} \rho_{s\omega} = \sum_{s\in S} l_s.$$

Since σ^* is an experiment,

$$\sum_{s\in S} \sigma^*_{s\omega} = 1 \qquad \text{for every } \omega \in \Omega.$$

Hence

$$\sum_{s\in S}\rho_{s\omega}=\frac{\sum_{s\in S}l_s-\tau}{1-\tau}=:\gamma \qquad \text{for every } \omega\in\Omega.$$

Since ρ is a nonzero quasi-experiment and $\rho_{s\omega}\geq 0$, we have $\gamma>0$. Therefore

$$\tilde{\rho}_s:=\frac{1}{\gamma}\rho_s \qquad \text{for each } s\in S$$

defines an experiment, since

$$\sum_{s\in S}\tilde{\rho}_{s\omega}=1 \qquad \text{for every } \omega\in\Omega.$$

Now define

$$\tilde{\tau}:=\frac{\tau}{\tau+(1-\tau)\gamma}.$$

Then $\tilde{\tau}\in(0,1]$, and for every $\sigma\in\Sigma$ and $s\in S$,

$$\tau\sigma_s+(1-\tau)\rho_s=\tau\sigma_s+(1-\tau)\gamma\tilde{\rho}_s=\big[\tau+(1-\tau)\gamma\big]\big[\tilde{\tau}\sigma_s+(1-\tilde{\tau})\tilde{\rho}_s\big].$$

Since posterior normalization is invariant under multiplication by a positive scalar, we have

$$\frac{p\odot[\tau\sigma_s+(1-\tau)\rho_s]}{\sum_{\omega\in\Omega}p_\omega[\tau\sigma_{s\omega}+(1-\tau)\rho_{s\omega}]}=\frac{p\odot[\tilde{\tau}\sigma_s+(1-\tilde{\tau})\tilde{\rho}_s]}{\sum_{\omega\in\Omega}p_\omega[\tilde{\tau}\sigma_{s\omega}+(1-\tilde{\tau})\tilde{\rho}_{s\omega}]}.$$

Thus the same posterior rule can be represented using the experiment $\tilde{\rho}$. Renaming $(\tilde{\tau},\tilde{\rho})$ as (τ,ρ) gives the desired representation. The uniqueness of the representation follows from the previous proof.

D.4 Proof for Theorem 1

Suppose first that $\{\succsim_0,\succsim_{\sigma,s}\}_{(\sigma,s)\in\Sigma\times S}$ satisfies Axioms 1–5. Then it satisfies Axioms 1, 2, 4, and 5. By Theorem 2, there exist a quasi-experiment ρ and $\tau\in(0,1]$ such that, for all $(\sigma,s)\in\Sigma\times S$,

$$p_{\sigma_s}(\cdot\mid s)=\frac{p\odot[\tau\sigma_s+(1-\tau)\rho_s]}{\sum_{\omega\in\Omega}p_\omega[\tau\sigma_{s\omega}+(1-\tau)\rho_{s\omega}]}.$$

We show that necessarily $\tau=1$.

Suppose, towards a contradiction, that $\tau<1$. Fix $s\in S$, and write

$$b_s:=(1-\tau)\rho_s.$$

Since Σ_s has nonempty interior, choose two non-collinear points $v,w\in\Sigma_s$. Since Σ_s is open, for some $\alpha\neq 1$ sufficiently close to 1, we have

$$\alpha v,\alpha w\in\Sigma_s.$$

By Axiom 3, the posterior induced by v must coincide with the posterior induced by αv. Since p has full support, this implies that the unnormalized posterior vectors are proportional:

$$\tau\alpha v + b_s = \lambda_v(\tau v + b_s)$$

for some $\lambda_v > 0$. Since $\alpha \neq 1$, we cannot have $\lambda_v = 1$. Hence b_s is proportional to v. The same argument applied to w implies that b_s is proportional to w. This contradicts the choice of non-collinear v and w, unless $b_s = 0$.

Thus $b_s = 0$ for every $s \in S$. Since $b_s = (1-\tau)\rho_s$ and $\tau < 1$, it follows that $\rho_s = 0$ for every s, contradicting the definition of a quasi-experiment as nonzero. Therefore $\tau = 1$.

E Proofs for Other Results

E.1 Proof for Proposition 1

The "if" side is trivial, and we focus on the "only if" side. Fix $s \in S$ and assume $\rho_s \neq 0$; otherwise the claim is immediate. By a standard separating hyperplane argument, Axiom 6 implies that, whenever $\sigma_s = \alpha\sigma'_s$ with $\alpha \in (0,1)$,

$$p_\sigma(\cdot \mid s) \in \mathrm{co}\{p,\, p_{\sigma'}(\cdot \mid s)\}.$$

Suppose, towards a contradiction, that $p^*_\rho(\cdot \mid s) \neq p$. Since Σ_s is open, the Bayesian posteriors $p^*_{\sigma'}(\cdot \mid s)$, as σ'_s ranges over Σ_s, form an open subset of $\Delta(\Omega)$, whose dimension is $|\Omega| - 1 \geq 3$. We may therefore choose $\sigma' \in \Sigma$ such that p, $p^*_\rho(\cdot \mid s)$, and $p^*_{\sigma'}(\cdot \mid s)$ are not collinear, and, again by openness, $\sigma \in \Sigma$ with $\sigma_s = \alpha\sigma'_s$ for some $\alpha \in (0,1)$.

By (4), $p_{\sigma'}(\cdot \mid s)$ and $p_\sigma(\cdot \mid s)$ lie on the segment joining $p^*_{\sigma'}(\cdot \mid s)$ and $p^*_\rho(\cdot \mid s)$, and they are distinct because $\alpha \neq 1$ and $\tau < 1$. The segment $\mathrm{co}\{p,\, p_{\sigma'}(\cdot \mid s)\}$ thus contains two distinct points of the line through $p^*_{\sigma'}(\cdot \mid s)$ and $p^*_\rho(\cdot \mid s)$, so p lies on this line as well—a contradiction.

Hence $p^*_\rho(\cdot \mid s) = p$, so that ρ_s is proportional to 1_Ω. As s was arbitrary, ρ is uninformative. $\square$

E.2 Proof of Proposition 2

Proof. Suppose Axiom 6 holds. If $\tau = 1$ then $p_\sigma(\cdot|s) = p^*_\sigma(\cdot|s)$ for all (σ, s) and conservatism is immediate, so assume $\tau < 1$. By Proposition 1, ρ is uninformative. Given

$$p_\sigma(\omega|s) = \tau(\sigma|s)p^*_\sigma(\omega|s) + (1-\tau(\sigma|s))p^*_\rho(\omega|s), \qquad \tau(\sigma|s) = \frac{\tau\sigma_p(s)}{\tau\sigma_p(s) + (1-\tau)\rho_p(s)},$$

we see that $p^*_\rho(\cdot|s) = p(\cdot)$ whenever $\rho_p(s) > 0$, while $\tau(\sigma|s) = 1$ whenever $\rho_p(s) = 0$. In either case $p_\sigma(\cdot|s)$ is a convex combination of $p^*_\sigma(\cdot|s)$ and $p(\cdot)$, and it follows that p_σ underreacts relative to p^*_σ for all s.

Conversely, suppose that the agent's beliefs exhibit conservatism. If $\tau = 1$ then $p_\sigma(\cdot|s) = p^*_\sigma(\cdot|s)$ for all (σ, s), so Axiom 6 holds trivially; assume henceforth $\tau < 1$. Let $\underline{\sigma}_s \in \Sigma_s$ be a constant likelihood vector, with corresponding experiment $\underline{\sigma} \in \Sigma$. Since $p^*_{\underline{\sigma}}(\cdot|s) = p(\cdot)$, conservatism implies that for all s,

$$p_{\underline{\sigma}}(\omega|s) = p(\omega).$$

Then

$$\begin{aligned} p_{\underline{\sigma}}(\omega|s) &= \tau(\underline{\sigma}|s)p^*_{\underline{\sigma}}(\omega|s) + (1-\tau(\underline{\sigma}|s))p^*_\rho(\omega|s) \\ \implies p(\omega) &= \tau(\underline{\sigma}|s)p(\omega) + (1-\tau(\underline{\sigma}|s))p^*_\rho(\omega|s) \\ \implies (1-\tau(\underline{\sigma}|s))p^*_\rho(\omega|s) &= (1-\tau(\underline{\sigma}|s))p(\omega). \end{aligned}$$

For any s, if $\tau(\underline{\sigma}|s) \neq 1$ then $p^*_\rho(\omega|s) = p(\omega)$ for all ω, and therefore ρ_s is a constant vector. If $\tau(\underline{\sigma}|s) = 1$ then $(1-\tau)\rho_p(s) = 0$, and since $\tau < 1$ this yields $\rho_p(s) = 0$; as p has full support and $\rho \geq 0$, it follows that $\rho(s|\omega) = 0$ for every ω, so ρ_s is again a constant vector. Hence ρ is uninformative, and Axiom 6 holds by Proposition 1. □

E.3 Proof for Proposition 3

By the separating hyperplane argument in the proof of Proposition 1, the agent overreacts to surprises at σ if and only if, for all $s \in S$ and all $\alpha > 1$ with $\alpha\sigma_s \in \Sigma_s$, $p_{\sigma'}(\cdot \mid s) \in \mathrm{co}\{p, p_\sigma(\cdot \mid s)\}$, where $\sigma'_s = \alpha\sigma_s$. Cancelling p componentwise, this holds if and only if there exist $a, b \geq 0$ with

$$\tau\alpha\sigma_s + (1-\tau)\rho_s = a1_\Omega + b\big[\tau\sigma_s + (1-\tau)\rho_s\big]. \tag{21}$$

We focus on the only if direction. Fix $s \in S$. Suppose first that σ_s is a constant vector. Then $p^*_\sigma(\omega \mid s) = p(\omega)$ for all ω, so no backfire at σ requires $p_\sigma(\omega \mid s) = p(\omega)$ for all ω, hence $p_\sigma(\cdot \mid s) = p$. Thus $\tau\sigma_s + (1-\tau)\rho_s$ is constant, so ρ_s is constant and admits the stated form.

Suppose now that σ_s is not constant. If 1_Ω, σ_s, and ρ_s were linearly independent, matching coefficients in (21) would give $b = \alpha$ from σ_s and $b = 1$ from ρ_s, contradicting $\alpha > 1$. Hence $\rho_s = \alpha_s\sigma_s - \beta_s 1_\Omega$ for some scalars α_s, β_s, so that

$$\tau\sigma_s + (1-\tau)\rho_s = \big[\tau + (1-\tau)\alpha_s\big]\sigma_s - (1-\tau)\beta_s 1_\Omega. \tag{22}$$

We claim $\tau + (1-\tau)\alpha_s > 0$. Since σ_s is not constant, there is ω with $\sigma(s \mid \omega) < \sigma_p(s)$, so that $p^*_\sigma(\omega \mid s) < p(\omega)$. If $\tau + (1-\tau)\alpha_s \leq 0$, then by (22), for $\pi := \tau\sigma + (1-\tau)\rho$,

$$\pi(s \mid \omega) \geq \pi_p(s),$$

which implies $p_\sigma(\omega \mid s) \geq p(\omega)$, violating no backfire at σ.

Since σ_s and 1_Ω are linearly independent, substituting (22) into (21) and equating the coefficients of σ_s and of 1_Ω gives

$$b = \frac{\tau\alpha + (1-\tau)\alpha_s}{\tau + (1-\tau)\alpha_s} > 1, \qquad a = (b-1)(1-\tau)\beta_s,$$

so $a \geq 0$ forces $\beta_s \geq 0$. Finally, if $\alpha_s < 0$, then $\rho_s = \alpha_s \sigma_s - \beta_s 1_\Omega$ has a negative component, contradicting $\rho_s \geq 0$; hence $\alpha_s \geq 0$.

For the second claim, suppose the agent overreacts at every $\sigma \in \Sigma$, and fix $s \in S$. Since Σ_s is open and $|\Omega| \geq 4$, choose $\sigma, \tilde{\sigma} \in \Sigma$ with 1_Ω, σ_s, $\tilde{\sigma}_s$ linearly independent. By the linear independence step above, ρ_s lies in the span of $\{\sigma_s, 1_\Omega\}$ and in the span of $\{\tilde{\sigma}_s, 1_\Omega\}$, whose intersection is the span of 1_Ω; nonnegativity gives $\rho_s = \lambda 1_\Omega$ with $\lambda \geq 0$. Matching coefficients in (21) at σ then gives $b = \alpha$ and $a = (1-\alpha)(1-\tau)\lambda$, so $a \geq 0$ and $\alpha > 1$ force $\lambda = 0$. Thus $\rho_s = 0$ for every $s \in S$, contradicting $\rho \neq 0$. $\square$

E.4 Proof for Proposition 4

By Proposition 3, $\rho_s = \alpha_s \sigma_s - \beta_s 1_\Omega$ with $\alpha_s \geq 0$, $\beta_s \geq 0$, so that

$$\tau\sigma_s + (1-\tau)\rho_s = \big[\tau + (1-\tau)\alpha_s\big]\big[\sigma_s - c_s 1_\Omega\big], \qquad c_s := \frac{(1-\tau)\beta_s}{\tau + (1-\tau)\alpha_s} \geq 0.$$

Hence $p_\sigma(\cdot \mid s)$ is the Bayesian posterior with respect to $\sigma_s - c_s 1_\Omega$, that is,

$$p_\sigma(\omega \mid s) = \frac{p(\omega)\big[\sigma(s \mid \omega) - c_s\big]}{\sigma_p(s) - c_s}, \qquad p^*_\sigma(\omega \mid s) = \frac{p(\omega)\sigma(s \mid \omega)}{\sigma_p(s)}.$$

Subtracting the two expressions,

$$p_\sigma(\omega \mid s) - p^*_\sigma(\omega \mid s) = \frac{p(\omega)\, c_s\big[\sigma(s \mid \omega) - \sigma_p(s)\big]}{\sigma_p(s)\big[\sigma_p(s) - c_s\big]},$$

and

$$p^*_\sigma(\omega \mid s) - p(\omega) = \frac{p(\omega)\big[\sigma(s \mid \omega) - \sigma_p(s)\big]}{\sigma_p(s)}.$$

Since $c_s \geq 0$ and $\sigma_p(s) - c_s > 0$, the two differences have the same sign. Therefore $p^*_\sigma(\omega \mid s) \geq p(\omega)$ implies $p_\sigma(\omega \mid s) \geq p^*_\sigma(\omega \mid s)$, and $p^*_\sigma(\omega \mid s) \leq p(\omega)$ implies $p_\sigma(\omega \mid s) \leq p^*_\sigma(\omega \mid s)$, which is the definition of overreaction at σ. $\square$

E.5 Proof of Proposition 5

Proof. Let $\pi = \tau\sigma + (1-\tau)\rho$. For an experiment e and a signal s, write $e_p(s) = \sum_\omega p(\omega) e(s \mid \omega)$ for the ex ante probability of s under e. Since signals are drawn from σ, the agent's average posterior on ω is

$$\bar{q}_p(\omega) = \sum_s p^*_\pi(\omega \mid s)\, \sigma_p(s).$$

Bayesian plausibility holds with respect to π, so $p(\omega) = \sum_s p^*_\pi(\omega \mid s)\, \pi_p(s)$, and hence

$$\bar{q}_p(\omega) - p(\omega) = \sum_s p^*_\pi(\omega \mid s)\big[\sigma_p(s) - \pi_p(s)\big] = (1-\tau) \sum_s p^*_\pi(\omega \mid s)\big[\sigma_p(s) - \rho_p(s)\big].$$

Since σ_p and ρ_p are both distributions over $S = \{s_1, s_2\}$, we have $\sigma_p(s_1) - \rho_p(s_1) = -[\sigma_p(s_2) - \rho_p(s_2)]$, so that

$$\bar{q}_p(\omega) - p(\omega) = (1-\tau)\big[\sigma_p(s_1) - \rho_p(s_1)\big]\big[p^*_\pi(\omega \mid s_1) - p^*_\pi(\omega \mid s_2)\big]. \tag{23}$$

Because there is no backfire at σ, the last bracket is strictly positive, and $\tau < 1$ makes the first factor strictly positive. Therefore

$$\bar{q}_p(\omega) > p(\omega) \iff \Phi(p) := \sigma_p(s_1) - \rho_p(s_1) > 0. \tag{24}$$

Write $a := \sigma(s_1 \mid \omega_1) - \rho(s_1 \mid \omega_1)$ and $b := \sigma(s_1 \mid \omega_2) - \rho(s_1 \mid \omega_2)$. Then

$$\Phi(p) = p(\omega_1)\, a + \big(1 - p(\omega_1)\big)\, b,$$

which is affine in $p(\omega_1)$, with $\Phi \to b$ as $p(\omega_1) \to 0$ and $\Phi \to a$ as $p(\omega_1) \to 1$. All three claims follow from the sign pattern of (a, b), using (24) and the fact that $p(\omega_1)$ ranges over the open interval $(0, 1)$.

ω_1-biased updating. $\Phi(p) > 0$ for every $p(\omega_1) \in (0, 1)$ if and only if $a \geq 0$ and $b \geq 0$ with at least one inequality strict. If both are nonnegative and one is strict, then Φ is a convex combination of a and b with strictly positive weights, hence strictly positive on $(0, 1)$. Conversely, if $a < 0$ then $\Phi(p) < 0$ for $p(\omega_1)$ close to 1, and if $b < 0$ then $\Phi(p) < 0$ for $p(\omega_1)$ close to 0; and if $a = b = 0$ then $\Phi \equiv 0$. This gives (11). The same argument applied to ω_2 in place of ω_1 characterizes ω_2-biased updating by the reverse inequalities $a \leq 0$ and $b \leq 0$, with at least one strict.

Confirmatory updating. Φ is affine and $\Phi(p) > 0$ if and only if $p(\omega_1)$ exceeds some $\theta \in (0, 1)$ precisely when Φ is strictly increasing in $p(\omega_1)$ and changes sign inside $(0, 1)$, that is, when $b < 0 < a$; the threshold is then $\theta = b/(b-a) \in (0, 1)$. Now $b < 0$ reads $\sigma(s_1 \mid \omega_2) < \rho(s_1 \mid \omega_2)$, which by $\sigma(s_1 \mid \omega_2) = 1 - \sigma(s_2 \mid \omega_2)$ and $\rho(s_1 \mid \omega_2) = 1 - \rho(s_2 \mid \omega_2)$ is equivalent to $\sigma(s_2 \mid \omega_2) > \rho(s_2 \mid \omega_2)$. Together with $a > 0$ this gives (12).

Average base-rate neglect. Symmetrically, $\Phi(p) > 0$ if and only if $p(\omega_1)$ falls below some $\theta \in (0, 1)$ precisely when $a < 0 < b$, again with $\theta = b/(b - a)$. Here $a < 0$ reads $\sigma(s_1 \mid \omega_1) < \rho(s_1 \mid \omega_1)$ and $b > 0$ is equivalent to $\sigma(s_2 \mid \omega_2) < \rho(s_2 \mid \omega_2)$, which gives (13).

Finally, the four sign patterns just described exhaust the possibilities apart from $a = b = 0$, in which case $\Phi \equiv 0$ and the martingale property holds at every prior. □

E.6 Proof of Proposition 6

Since $p(\cdot|s) \neq p_\sigma(\cdot|s)$ for some s, it must be that $1 - \tau > 0$. Fix ω and $p(\omega) \in (\theta, 1)$. Since p is fixed in what follows, we suppress the dependence of ρ on the prior p in the notation.

In our 2×2 context, since s_1 is diagnostic of ω_1 for any $\sigma \in \Sigma^1$, there is no backfire effect at α iff the agent exhibits $p_\sigma(\omega_1|s_1) > p_\sigma(\omega_2|s_1)$ at s_1 iff the agent exhibits $p_\sigma(\omega_1|s_2) < p_\sigma(\omega_2|s_2)$ at s_2.

Lemma 5. *(i) The following are equivalent:*
a) $p_\sigma(\omega_1|s_1) > p_\sigma(\omega_2|s_1)$
b) $p_\sigma(\omega_1|s_2) < p_\sigma(\omega_2|s_2)$
c) $\tau\sigma(s_1|\omega_1) + (1-\tau)\rho(s_1|\omega_1) > \tau\sigma(s_1|\omega_2) + (1-\tau)\rho(s_1|\omega_2)$
d) $\frac{\tau}{1-\tau}(\sigma(s_1|\omega_1) - \sigma(s_1|\omega_2)) > \rho(s_1|\omega_2) - \rho(s_1|\omega_1)$
(ii) There is no backfire effect at any $\sigma \in \Sigma_1$ *iff* $\rho(s_1|\omega_2) - \rho(s_1|\omega_1) \leq 0$.

Proof. (i) By the representation,

$$p_\sigma(\omega_1|s_1) > p_\sigma(\omega_2|s_1)$$

$$\iff \tau\sigma(s_1|\omega_1) + (1-\tau)\rho(s_1|\omega_1) > \tau\sigma(s_1|\omega_2) + (1-\tau)\rho(s_1|\omega_2)$$
$$\iff \frac{\tau}{1-\tau}(\sigma(s_1|\omega_1) - \sigma(s_1|\omega_2)) > \rho(s_1|\omega_2) - \rho(s_1|\omega_1)$$

Also,

$$p_\sigma(\omega_1|s_2) > p_\sigma(\omega_2|s_2)$$

$$\iff \tau\sigma(s_1|\omega_1)+(1-\tau)\rho(s_1|\omega_1) > \tau\sigma(s_1|\omega_2)+(1-\tau)\rho(s_1|\omega_2) \iff p_\sigma(\omega_1|s_1) > p_\sigma(\omega_2|s_1).$$

(ii) The hypothesis is equivalent to $\frac{\tau}{1-\tau}(\sigma(s_1|\omega_1) - \sigma(s_1|\omega_2)) > \rho(s_1|\omega_2) - \rho(s_1|\omega_1)$ for all $\sigma \in \Sigma_1$. Taking a sequence of experiments where $\sigma(s_1|\omega_1), \sigma(s_1|\omega_2) \to \frac{1}{2}$ implies $\rho(s_1|\omega_2) - \rho(s_1|\omega_1) \leq 0$. □

Lemma 6. *At* p, σ*, the agent exhibits* ω_i*-confirmatory reading* $p_\sigma(\omega_i|s_1, s_2) > p(\omega_1)$ *iff*

$$|\tau\sigma(s_1|\omega_i) + (1-\tau)\rho(s_1|\omega_i) - \frac{1}{2}| < |\tau\sigma(s_1|\omega_{-i}) + (1-\tau)\rho(s_1|\omega_{-i}) - \frac{1}{2}|.$$

Proof. Since $\frac{p_\sigma(\omega_1|s_1,s_2)}{p_\sigma(\omega_2|s_1,s_2)}$ is given by the expression

$$\frac{(\tau\sigma(s_2|\omega_1) + (1-\tau)\rho(s_2|\omega_1))\,(\tau\sigma(s_1|\omega_1) + (1-\tau)\rho(s_1|\omega_1))}{(\tau\sigma(s_2|\omega_2) + (1-\tau)\rho(s_2|\omega_2))\,(\tau\sigma(s_1|\omega_2) + (1-\tau)\rho(s_1|\omega_2))}\frac{p(\omega_1)}{p(\omega_2)}$$

we compute that ω_1-confirmatory reading is equivalent to

$$\frac{p_\sigma(\omega_1|s_1,s_2)}{p_\sigma(\omega_2|s_1,s_2)} > \frac{p(\omega_1)}{p(\omega_2)}$$

$$\iff (\tau\sigma(s_2|\omega_1) + (1-\tau)\rho(s_2|\omega_1))\,(\tau\sigma(s_1|\omega_1) + (1-\tau)\rho(s_1|\omega_1))$$
$$> (\tau\sigma(s_2|\omega_2) + (1-\tau)\rho(s_2|\omega_2))\,(\tau\sigma(s_1|\omega_2) + (1-\tau)\rho(s_1|\omega_2)).$$

Since the function $r \mapsto (1-r)r$ defined on the unit interval is hump-shaped and symmetric around its maximum at $r = \frac{1}{2}$, the above condition is equivalent to

$$|\tau\sigma(s_1|\omega_1) + (1-\tau)\rho(s_1|\omega_1) - \frac{1}{2}| < |\tau\sigma(s_1|\omega_2) + (1-\tau)\rho(s_1|\omega_2) - \frac{1}{2}|.$$

An analogous result holds for ω_2-confirmatory reading. □

Lemma 7. *Suppose there is no backfire effect at any $\sigma \in \Sigma^1$.At p, the agent exhibits ω_1-confirmatory reading (resp ω_2-confirmatory reading, no confirmatory reading) for all symmetric $\sigma \in \Sigma^1$ iff*

$$\rho(s_1|\omega_1) + \rho(s_1|\omega_2) < (\textit{resp.} \quad >, =) 1.$$

Proof. We begin by establishing sufficiency. Given the characterization above, ω_i-confirmatory reading at all σ implies that, after taking the limit of a sequence of experiments where $\sigma(s_1|\omega_1), \sigma(s_1|\omega_2) \to \frac{1}{2}$,

$$|\tau\frac{1}{2} + (1-\tau)\rho(s_1|\omega_i) - \frac{1}{2}| \leq |\tau\frac{1}{2} + (1-\tau)\rho(s_1|\omega_{-i}) - \frac{1}{2}|$$

$$\implies |\rho(s_1|\omega_i) - \frac{1}{2}| \leq |\rho(s_1|\omega_{-i}) - \frac{1}{2}|.$$

Moreover, if there is no backfire at any σ then

$$\rho(s_1|\omega_1) \geq \rho(s_1|\omega_2).$$

Step 1: Show that ω_1-confirmatory reading for all symmetric $\sigma \in \Sigma^1$ implies

$$\rho(s_1|\omega_1) + \rho(s_1|\omega_2) < 1.$$

By the preceding $\rho(s_1|\omega_1) \geq \rho(s_1|\omega_2)$ and $|\rho(s_1|\omega_1) - \frac{1}{2}| \leq |\rho(s_1|\omega_2) - \frac{1}{2}|$. Consider the following cases:

a) ω_1-confirmatory reading for all symmetric $\sigma \in \Sigma^1$ and $\rho(s_1|\omega_1) \geq \frac{1}{2}$.

Then it must be that $\rho(s_1|\omega_1) \geq \frac{1}{2} \geq \rho(s_1|\omega_2)$ and $\rho(s_1|\omega_1) - \frac{1}{2} \leq \frac{1}{2} - \rho(s_1|\omega_2)$ and in particular $\rho(s_1|\omega_1) + \rho(s_1|\omega_2) \leq 1$. Suppose by way of contradiction that $\rho(s_1|\omega_1) + \rho(s_1|\omega_2) = 1$. Then, as $\rho(s_1|\omega_1) \geq \frac{1}{2} \geq \rho(s_1|\omega_2)$, this is possible only if $\rho(s_1|\omega_1) = \rho(s_1|\omega_2) = \frac{1}{2}$. Moreover, it must be that $\tau\sigma(s_1|\omega_1) + (1-\tau)\rho(s_1|\omega_1) > \frac{1}{2} > \tau\sigma(s_1|\omega_2) + (1-\tau)\rho(s_1|\omega_2)$. So

$$|\tau\sigma(s_1|\omega_1) + (1-\tau)\rho(s_1|\omega_1) - \frac{1}{2}| < |\tau\sigma(s_1|\omega_2) + (1-\tau)\rho(s_1|\omega_2) - \frac{1}{2}|$$

$$\iff \tau\sigma(s_1|\omega_1) + (1-\tau)\rho(s_1|\omega_1) - \frac{1}{2} < \frac{1}{2} - (\tau(1-\sigma(s_1|\omega_1)) + (1-\tau)\rho(s_1|\omega_2))$$

$$\iff \tau\sigma(s_1|\omega_1) + (1-\tau)\rho(s_1|\omega_1) < \tau\sigma(s_1|\omega_1) + (1-\tau)(1-\rho(s_1|\omega_2))$$

$$\iff \rho(s_1|\omega_1) + \rho(s_1|\omega_2) < 1,$$

a contradiction.

b) ω_1-confirmatory reading for all symmetric $\sigma \in \Sigma^1$ and $\rho(s_1|\omega_1) < \frac{1}{2}$.

Then $\frac{1}{2} > \rho(s_1|\omega_1) \geq \rho(s_1|\omega_2)$ and in particular, $\rho(s_1|\omega_1) + \rho(s_1|\omega_2) < 1$.

Step 2: Show that ω_2-confirmatory reading for all symmetric $\sigma \in \Sigma^1$ implies

$$\rho(s_1|\omega_1) + \rho(s_1|\omega_2) > 1.$$

Recall that $\rho(s_1|\omega_1) \geq \rho(s_1|\omega_2)$ and $|\rho(s_1|\omega_2) - \frac{1}{2}| \leq |\rho(s_1|\omega_1) - \frac{1}{2}|$. Consider the following cases:

a) ω_2-confirmatory reading for all symmetric $\sigma \in \Sigma^1$ and $\rho(s_1|\omega_1) \geq \rho(s_1|\omega_2) \geq \frac{1}{2}$.

Then $\rho(s_1|\omega_1)+\rho(s_1|\omega_2) \geq 1$. Suppose by way of contradiction that $\rho(s_1|\omega_1)+\rho(s_1|\omega_2) = 1$, which is equivalent to $\rho(s_1|\omega_1) = \rho(s_1|\omega_2) = \frac{1}{2}$ in this case. Then $\tau\sigma(s_1|\omega_1)+(1-\tau)\rho(s_1|\omega_1) > \frac{1}{2} > \tau\sigma(s_1|\omega_2)+(1-\tau)\rho(s_1|\omega_2)$ for any symmetric $\sigma \in \Sigma^1$. But then

$$|\tau\sigma(s_1|\omega_2) + (1-\tau)\rho(s_1|\omega_2) - \frac{1}{2}| < |\tau\sigma(s_1|\omega_1) + (1-\tau)\rho(s_1|\omega_1) - \frac{1}{2}|$$

$$\iff \frac{1}{2} - (\tau\sigma(s_1|\omega_2) + (1-\tau)\rho(s_1|\omega_2)) < \tau\sigma(s_1|\omega_1) + (1-\tau)\rho(s_1|\omega_1) - \frac{1}{2}$$

$$\iff 1 - \rho(s_1|\omega_2) < \rho(s_1|\omega_1) \iff \rho(s_1|\omega_1) + \rho(s_1|\omega_2) > 1,$$

a contradiction.

b) ω_2-confirmatory reading for all symmetric $\sigma \in \Sigma^1$ and $\rho(s_1|\omega_1) \geq \frac{1}{2} > \rho(s_1|\omega_2)$.

We show that this case is not possible. In this case it must be that $\rho(s_1|\omega_2) - \frac{1}{2} \leq \frac{1}{2} - \rho(s_1|\omega_1)$, which implies $\rho(s_1|\omega_1) + \rho(s_1|\omega_2) \leq 1$. Moreover, in this case we have $\tau\sigma(s_1|\omega_1) + (1-\tau)\rho(s_1|\omega_1) > \frac{1}{2} > \tau\sigma(s_1|\omega_2) + (1-\tau)\rho(s_1|\omega_2)$ for any symmetric σ. But then as we saw in case (a) it must be that

$$|\tau\sigma(s_1|\omega_2) + (1-\tau)\rho(s_1|\omega_2) - \frac{1}{2}| < |\tau\sigma(s_1|\omega_1) + (1-\tau)\rho(s_1|\omega_1) - \frac{1}{2}|$$

$$\iff \rho(s_1|\omega_1) + \rho(s_1|\omega_2) > 1,$$

a contradiction.

c) ω_2-confirmatory reading for all symmetric $\sigma \in \Sigma^1$ and $\frac{1}{2} > \rho(s_1|\omega_1) \geq \rho(s_1|\omega_2)$.

We show that this case is not possible. In this case $\rho(s_1|\omega_1) + \rho(s_1|\omega_2) < 1$. Also there exists σ with $\sigma(s_1|\omega_1)$ sufficiently high that $\tau\sigma(s_1|\omega_1) + (1-\tau)\rho(s_1|\omega_1) > \frac{1}{2} > \tau\sigma(s_1|\omega_2) + (1-\tau)\rho(s_1|\omega_2)$, and so again as in case (a), it must be that

$$|\tau\sigma(s_1|\omega_2) + (1-\tau)\rho(s_1|\omega_2) - \frac{1}{2}| < |\tau\sigma(s_1|\omega_1) + (1-\tau)\rho(s_1|\omega_1) - \frac{1}{2}|$$

$$\iff \rho(s_1|\omega_1) + \rho(s_1|\omega_2) > 1,$$

a contradiction.

Step 3: Show that no confirmatory reading for all symmetric $\sigma \in \Sigma^1$ implies

$$\rho(s_1|\omega_1) + \rho(s_1|\omega_2) = 1.$$

If there is no confirmatory reading for all σ then $|\rho(s_1|\omega_1) - \frac{1}{2}| = |\rho(s_1|\omega_2) - \frac{1}{2}|$. Given $\rho(s_1|\omega_1) \geq \rho(s_1|\omega_2)$, this possible only if $\rho(s_1|\omega_1) \geq \frac{1}{2} \geq \rho(s_1|\omega_2)$ and in turn $\rho(s_1|\omega_1) + \rho(s_1|\omega_2) = 1$. □

References

[1] Allcott, H. and M. Gentzkow (2017): “Social Media and Fake News in the 2016 Election”, *Journal of Economic Perspectives* 31(2), pp. 211–236.

[2] Ba, C. (2026): ”Robust Misspecified Models”, *American Economic Review* 116(4), pp. 1340–1379.

[3] Benjamin, D. (2019): “Errors in Probabilistic Reasoning and Judgment Biases.” Chapter for the *Handbook of Behavioral Economics* (edited by Doug Bernheim, Stefano DellaVigna, and David Laibson). Elsevier Press.

[4] Bohren, J. A., and D. N. Hauser (2023): “The Behavioral Foundations of Model Misspecification: A Decomposition”, Working Paper, University of Pennsylvania.

[5] Bohren, J. A., and D. N. Hauser (2025): "Misspecified Models in Learning and Games", *Annual Review of Economics* 17, pp. 427–451.

[6] Camerer, C. F., A.Dreber, E. Forsell, T. H. Ho, J. Huber, M. Johannesson, M. Kirchler, J. Almenberg, A. Altmejd, T. Chan, E. Heikensten, F. Holzmeister, T. Imai, S. Isaksson, G. Nave, T. Pfeiffer, M. Razen, H. Wu (2016): “Evaluating Replicability of Laboratory Experiments in Economics”, *Science* 351, pp. 1433-6. doi: 10.1126/science.aaf0918.

[7] Chambers, C. P., Y. Masatlioglu, and C. Raymond (2024): “Coherent Distorted Beliefs”, Working Paper, arXiv:2310.09879.

[8] Chan, K. (2026): “An Axiomatic Model and Test of Grether (1980) and Bayes' Rule”, working paper, SSRN 5084221.

[9] Cheng, I. and A. Hsiaw (2022): “Distrust in Experts and the Origins of Disagreement”, *Journal of Economic Theory* 200, 105401, ISSN 0022-0531, https://doi.org/10.1016/j.jet.2021.105401.

[10] Cripps, M. W. (2018): “Divisible Updating”, Working Paper, University College London.

[11] DeGroot, M. H. (1974): “Reaching a Consensus”, *Journal of the American Statistical Association* 69(345), pp. 118–121.

[12] Dominiak, A., M. Kovach, and G. Tserenjigmid (2025): “Inertial Updating with General Information”, Working Paper, arXiv:2502.00958.

[13] Eil, D. and J. Rao (2011): “The Good News-Bad News Effect: Asymmetric Processing of Objective Information about Yourself”, American Economic Journal: Microeconomics 3(2), pp. 114–138.

[14] Epstein, L. , Noor, J. and A. Sandroni (2008): “Non-Bayesian Updating: A Theoretical Framework”, *Theoretical Economics* 3(2), pp. 193-229.

[15] Farina, A. and C. Herman (2026): "Communicating with Data Generating Processes: An Experimental Analysis", mimeo.

[16] Faure, C. A. (2002): "An Elementary Proof of the Fundamental Theorem of Projective Geometry", *Geometriae Dedicata* 90, pp. 145-151.

[17] Faure, C.-A. and A. Froelicher (1994): "Morphisms of Projective Geometries and Semilinear Maps", *Geometriae Dedicata* 53(3), pp. 237–262.

[18] Fudenberg, D. and G. Lanzani (2026): "Forecasts and Conditionally IID Models", mimeo.

[19] Gentzkow, M. and J. Shapiro (2006): "Media Bias and Reputation", *Journal of Political Economy* 114(2), pp. 280-316.

[20] Ghirardato, P. (2002): "Revisiting Savage in a Conditional World", *Economic Theory* 20(1), pp. 83–92.

[21] Grether, D. M. (1980): "Bayes Rule as a Descriptive Model: The Representativeness Heuristic", *Quarterly Journal of Economics* 95(3), pp. 537–557.

[22] Haglin K. (2017): "The Limitations of the Backfire Effect", *Research and Politics* 4, pp. 31–35.

[23] Havlicek, H. (1994): "A Generalization of Brauner's Theorem on Linear Mappings", *Mitteilungen aus dem Mathematischen Seminar Giessen* 215, pp. 27–41.

[24] Hu, E. (2026): "Confidence in Inference", working paper.

[25] Jakobsen, A. M. (2026): "Coarse Bayesian Updating", *Review of Economic Studies*, forthcoming.

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

[27] Ke, S., B. Wu, and C. Zhao (2024): "Learning from a Black Box", *Journal of Economic Theory* 221, 105886.

[28] Kovach, M. (2021): "Conservative Updating", Working Paper, arXiv:2102.00152.

[29] Lenz, H. (1958): "Einige Anwendungen der projektiven Geometrie auf Fragen der Flaechentheorie", *Mathematische Nachrichten* 18(1-6), pp. 346–359.

[30] Liu, B. and A. Moss (2024): "The Role of Accounting Information in an Era of Fake News", *Journal of Accounting and Economics* 79, article 101764.

[31] Lord, C. G., L. Ross, and M. R. Lepper (1979): "Biased Assimilation and Attitude Polarization: The Effects of Prior Theories on Subsequently Considered Evidence", *Journal of Personality and Social Psychology* 37, pp. 2098-2109.

[32] Lowen, R. (1982): "A local "Fundamental Theorem" for Classical Topological Projective Spaces", Archiv der Mathematik 38, pp. 286–288. https://doi.org/10.1007/BF01304789

[33] Ma, J., and C. Zhao (2024): "Logic-Based Updating", *Journal of Economic Theory* 221, 105901.

[34] Molavi, P. (2024): "The Empirical Content of Bayesianism," working paper.

[35] Noor, J. and F. Payró (2026): "Belief in Mean Reversion and the Law of Small Numbers", *Theoretical Economics*, forthcoming.

[36] Nyhan, B. and J. Reifler, (2010): "When Corrections Fail: the Persistence of Political Misperceptions", *Political Behavior* 32, pp. 303–330.

[37] Nyhan, B. and J. Reifler, (2015): "Does Correcting Myths about the Flu Vaccine Work? An Experimental Evaluation of the Effects of Corrective Information" *Vaccine* 33, pp. 459–464.

[38] Open Science Collaboration (2015): "Estimating the Reproducibility of Psychological Science", *Science* 349(6251), aac4716. doi: 10.1126/science.aac4716.

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

[40] Ortoleva, P. (2024): "Alternatives to Bayesian Updating", *Annual Review of Economics* 16, pp. 545–570.

[41] Sancho de Salas, J. B. (2026): "The Fundamental Theorem for Locally Projective Geometries", *Geometriae Dedicata* 220(5), pp. 145-151. https://doi.org/10.1007/s10711-025-01066-7

[42] Shmaya, E. and L. Yariv (2016): "Experiments on Decisions under Uncertainty: A Theoretical Framework", American Economic Review 106(7), pp. 1775–1801.

[43] Whitmeyer, M. (2026): "Blackwell-Monotone Updating Rules", *Journal of Political Economy*, forthcoming.

[44] Yang, Y. (2025): "Overprecise Updating", mimeo.

[45] Y. Yang and C. Zhao (2026): "Representativeness and Belief Updating", working paper.

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